Robust SAC-Enabled UAV-RIS Assisted Secure MISO Systems With Untrusted EH Receivers

Hamid Reza Hashempour, Le-Nam Tran, *Senior Member, IEEE*, Duy H. N. Nguyen, *Senior Member, IEEE* and Hien Quoc Ngo, *Fellow, IEEE*

arXiv:2602.14191v3 [eess.SP] 19 Nov 2026

***Abstract*—Secure downlink transmission in UAV-mounted reconfigurable intelligent surface (RIS)-assisted multi-user multiple-input single-output (MISO) networks is challenged by imperfect channel state information (CSI), internal eavesdropping by untrusted energy-harvesting receivers (UEHRs), discrete RIS phase shifts, and nonlinear energy harvesting (EH). This paper jointly optimizes the UAV hovering position, base station (BS) transmit power allocation, and discrete RIS phase shifts to maximize the worst-case secrecy energy efficiency (WCSEE) under bounded CSI uncertainty and nonlinear EH constraints. The resulting problem is a highly non-convex mixed-integer fractional optimization problem with tightly coupled spatial, power, and RIS variables. To address these challenges, we develop two complementary solution frameworks. First, for perfect CSI and continuous RIS phase shifts, we derive an analytical benchmark based on block coordinate descent (BCD) and successive convex approximation (SCA). Second, for the general robust setting with discrete phase quantization and CSI uncertainty, we propose a customized soft actor–critic (SAC) framework that incorporates continuous-to-discrete action projection, twin critics to mitigate value overestimation, and a feasibility-aware reward mechanism. Numerical results demonstrate that the proposed SAC policy achieves up to 28%, 16%, and 6% higher WCSEE than the SCA benchmark, DDPG, and TD3, respectively, while exhibiting faster convergence and maintaining stable performance across varying user densities, UEHR counts, and CSI uncertainty levels.**

***Index Terms*—UAV-mounted RIS, physical-layer security, SWIPT, worst-case secrecy energy efficiency, successive convex approximation, soft actor–critic, deep reinforcement learning.**

I. Introduction

RECONFIGURABLE intelligent surfaces (RISs) and unmanned aerial vehicles (UAVs) have emerged as key technologies for enhancing coverage, spectral efficiency, and energy efficiency in beyond-5G and 6G wireless systems [1]–[4]. By employing low-cost reflecting elements with programmable phase shifts, RISs can reconfigure the wireless propagation environment [5]. However, terrestrial RISs are constrained by fixed deployment locations and may suffer from blockages, shadowing, and unfavorable propagation conditions. Mounting an RIS on a hovering UAV provides flexible three-dimensional (3D) deployment and facilitates dominant line-of-sight (LoS) links with terrestrial base stations (BSs) and ground users [6], [7].

Meanwhile, simultaneous wireless information and power transfer (SWIPT) offers an effective approach for powering energy-constrained Internet of Things (IoT) devices and battery-free sensors [8], [9]. RIS-assisted SWIPT designs have been widely studied to enhance energy harvesting performance [10], [11]. In practical multi-user SWIPT systems, a separated-receiver architecture is commonly considered [12], where the BS simultaneously transmits dual-purpose radio-frequency (RF) signals to legitimate information receivers (IRs) and energy-harvesting receivers (EHRs). While this architecture improves energy sustainability, the broadcast nature of RF transmission creates inherent physical-layer security (PLS) vulnerabilities [13], [14].

In particular, practical nonlinear energy-harvesting (EH) rectifiers require sufficient incident RF power to overcome diode activation thresholds [15]. Consequently, EHRs may be located close to the transmitter or receive high-power signal beams, making them potential internal eavesdroppers when they are not authorized to access the transmitted information [16], [17]. Such untrusted EHRs (UEHRs) create a fundamental security–energy tradeoff: the network must deliver sufficient RF power to satisfy their EH requirements while limiting the information leakage they may obtain.

This tradeoff becomes more challenging in UAV-mounted RIS systems. First, since UEHRs are untrusted, their channels cannot be assumed to be accurately acquired through cooperative training; therefore, imperfect channel state information (CSI) with bounded uncertainty is required for robust security design [18], [19]. Second, practical RIS hardware supports only finite-resolution phase shifts, resulting in discrete reflection coefficients [20]. Third, the aerial channel characteristics depend nonlinearly on the UAV's 3D position [21]. The joint design of secure information transmission, nonlinear EH, discrete RIS reflection, and UAV positioning under worst-case CSI uncertainty therefore leads to a highly non-convex, mixed-integer, and high-dimensional optimization problem. These coupled challenges motivate the development of efficient robust optimization and learning-based approaches for worst-case secrecy energy efficiency (WCSEE) maximization.

A. Related Work

The literature most relevant to this work can be broadly categorized into four areas: aerial RIS-assisted communications, PLS for SWIPT networks, robust optimization under

Hamid Reza Hashempour and Hien Quoc Ngo are with the Center for Wireless Innovation (CWI), Queen's University Belfast, BT3 9DT Belfast, U.K., (Email:{h.hashempoor, hien.ngo}@qub.ac.uk).

Le-Nam Tran is with the School of Electrical and Electronic Engineering, University College Dublin, Belfield, Dublin 4, D04 V1W8, Ireland (Email: nam.tran@ucd.ie)

Duy H. N. Nguyen is with the department of Electrical and Computer Engineering, San Diego State University, San Diego, CA 92182, USA. (Email: duy.nguyen@sdsu.edu).

imperfect CSI, and deep reinforcement learning (DRL) for wireless resource management.

1) Aerial RIS-Assisted Networks: Integrating RISs with UAV platforms provides flexible 3D deployment and favorable propagation conditions, enabling joint optimization of aerial positioning and passive beamforming [6], [7], [21]–[23]. Misbah *et al.* [6] jointly optimized RIS phase shifts and UAV 3D positioning for downlink sum-rate maximization using alternating optimization (AO). The authors in [7] investigated RIS-assisted cognitive UAV networks by jointly optimizing the sensing duration, passive beamforming, and the 3D position of the leading UAV to maximize network throughput via AO and successive convex approximation (SCA). UAV-assisted simultaneously transmitting and reflecting RIS (STAR-RIS) architectures were explored in [22] to achieve full-space coverage by jointly optimizing 3D UAV positioning, power allocation, and transmission/reflection beamforming for sum-rate maximization, whereas [21] investigated robust aerial RIS beamforming under imperfect CSI. Despite these advances, existing aerial RIS studies primarily focus on information transmission and do not jointly consider EH requirements and internal eavesdropping in SWIPT networks.

2) PLS in SWIPT Systems: While SWIPT has been extensively investigated under both colocated and separated receiver architectures [8], [10]–[12], incorporating PLS into SWIPT networks introduces unique design challenges [14]–[17]. Specifically, in [17], robust active and passive beamforming was developed for secure STAR-RIS-assisted SWIPT with nonlinear EH under imperfect CSI. The work in [15] optimized global energy efficiency for secure MISO SWIPT with a nonlinear power-splitting EH model. For untrusted EH receivers, [14], [16] investigated secure beamforming designs that suppress information leakage while satisfying EH requirements through artificial noise and passive beamforming. However, these studies mainly consider terrestrial systems with fixed transceiver locations and do not jointly address UAV positioning, discrete RIS phase shifts, nonlinear EH, and bounded CSI uncertainty. The resulting WCSEE problem in UAV-mounted RIS-assisted SWIPT therefore remains largely unexplored.

3) Robust Optimization Under CSI Uncertainty: Imperfect channel estimation and feedback errors can significantly degrade PLS, motivating robust transmission designs based on bounded CSI uncertainty [18], [19]. Norm-bounded worst-case uncertainty models are widely adopted to provide deterministic security guarantees [21], [24], [25]. Conventional robust designs commonly employ the S-procedure, block coordinate descent (BCD), SCA, and related fractional-programming techniques to obtain tractable approximations [26]–[28]. Although these model-based approaches provide useful structural insights and performance benchmarks, their computational burden increases substantially with the system dimension. Moreover, discrete RIS phase shifts introduce non-differentiability, typically requiring continuous relaxation followed by quantization, which can cause performance loss and complicate feasibility preservation. These limitations become particularly pronounced when UAV positioning, transmit power, RIS configuration, nonlinear EH, and worst-case secrecy constraints must be optimized jointly.

4) DRL for Wireless Resource Management: DRL has been increasingly adopted for complex wireless resource-allocation problems, including RIS configuration, beamforming, UAV control, and secrecy optimization [29]–[32]. In [29], DRL was employed for ecrecy energy efficiency (SEE) maximization in RIS-assisted wireless networks. Deterministic policy-gradient methods, including deep deterministic policy gradient (DDPG) and twin delayed DDPG (TD3), have also been applied to optimize RIS configurations, UAV trajectories, and beamforming variables [30], [31]. Soft actor–critic (SAC) [33], [34], based on maximum-entropy reinforcement learning, has been investigated as an alternative that improves exploration and training stability through entropy regularization. However, existing DRL-based approaches have not jointly considered discrete RIS phase shifts, UAV hovering optimization, nonlinear EH, and worst-case CSI uncertainty in secure SWIPT systems.

B. Motivations and Contributions

As summarized in Table I, existing studies have investigated UAV communications, RIS beamforming, SWIPT, and PLS largely in isolation or under simplified assumptions. A unified framework that jointly captures UAV positioning, discrete RIS phase configuration, active power control, nonlinear EH, and imperfect CSI for secure SWIPT remains lacking. This motivates the development of a robust and computationally efficient design for UAV-mounted RIS-assisted secure networks.

To this end, we consider a secure downlink multi-user MISO SWIPT system in which a hovering UAV-mounted RIS assists a multi-antenna BS in serving multiple legitimate IRs in the presence of distributed UEHRs. Zero-forcing (ZF) precoding is adopted at the BS to suppress inter-user interference among the legitimate IRs. The UAV hovering position, discrete RIS phase shifts, and BS transmit power allocation are jointly optimized to maximize the WCSEE under nonlinear EH requirements and bounded CSI uncertainty. The main contributions are summarized as follows:

- We formulate a robust WCSEE maximization problem for a UAV-mounted RIS-assisted MISO SWIPT system with nonlinear EH and norm-bounded CSI uncertainty at untrusted receivers. The formulation jointly accounts for discrete RIS phase shifts, UAV hovering constraints, BS transmit-power limits, and minimum harvested-energy requirements.
- For the continuous-phase and perfect-CSI setting, we develop a model-based BCD algorithm incorporating SCA and the Charnes–Cooper transformation. This framework provides an analytical benchmark for jointly optimizing BS power allocation, RIS reflection phases, and UAV positioning, with convergence to a stationary solution.
- For the general robust problem with discrete RIS phases and bounded CSI uncertainty, we develop a customized SAC framework. The proposed design employs continuous-to-discrete action projection for phase quantization, twin critics to alleviate value overestimation, and a feasibility-aware reward mechanism to improve constraint satisfaction without extensive manual penalty tuning.

TABLE I
COMPARISON OF THE PROPOSED FRAMEWORK WITH EXISTING STUDIES

Feature / Attribute	**Ours**	[6], [7], [21]	[22]	[10], [11]	[17]	[15]	[14]	[16]	[30]	[29]	[31]
UAV Location Optimization	✓	✓	✓								
RIS / STAR-RIS Phase Optimization	✓	✓	✓	✓	✓			✓	✓	✓	✓
SWIPT / EH	✓			✓	✓	✓	✓	✓			
Non-linear EH Model	✓			✓	✓	✓	✓				
Untrusted Receivers (Eavesdroppers)	✓				✓	✓	✓	✓			
Worst-Case / Bounded CSI Uncertainty	✓	✓			✓			✓			
Discrete RIS Phase Quantization	✓									✓	✓
Model-Based Optimization (BCD / SCA)	✓	✓	✓	✓	✓	✓	✓	✓	✓		
DRL	✓								✓	✓	✓
SEE / WCSEE	✓					✓	✓			✓	

- Extensive simulations demonstrate that the proposed SAC framework achieves up to 28% and 16% higher WC-SEE than the SCA benchmark and DDPG, respectively, while consistently outperforming TD3. Although SCA becomes comparable to DDPG at high transmit powers or with large RIS sizes, SAC maintains the best performance across all considered scenarios. The results further demonstrate the superior sample efficiency and robustness of SAC under different mini-batch sizes, user densities, UEHR numbers, and CSI uncertainty levels.

Organization: Section II presents the system model, channel representations, and the WCSEE problem formulation. Section III details the BCD–SCA optimization benchmark. Section IV develops the proposed SAC-based framework. Section V discusses numerical simulation results, and Section VI concludes the paper.

Notation: Bold lowercase and uppercase letters denote vectors and matrices, respectively. The operators $(\cdot)^T$, $(\cdot)^H$, and $\mathrm{Tr}(\cdot)$ denote the transpose, conjugate transpose, and trace of a matrix, respectively. $\|\cdot\|_2$ and $|\cdot|$ denote the Euclidean vector norm and complex modulus, respectively. $\mathfrak{Re}\{\cdot\}$ and $\mathfrak{Im}\{\cdot\}$ extract the real and imaginary components. $\mathbb{E}[\cdot]$ denotes statistical expectation. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ represents a circularly symmetric complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. $\mathbf{I}_N$ denotes the $N \times N$ identity matrix, $[\mathbf{x}]_m$ denotes the m-th element of vector $\mathbf{x}$, and $\mathrm{diag}\{\mathbf{x}\}$ forms a diagonal matrix from vector $\mathbf{x}$. $\mathbb{R}^{M\times N}$ and $\mathbb{C}^{M\times N}$ denote the spaces of $M \times N$ real and complex matrices, respectively.

II. SYSTEM MODEL AND PROBLEM FORMULATION

In the considered network, a multi-antenna BS simultaneously transmits confidential information to legitimate IRs and delivers RF energy to UEHRs through downlink transmission. The IRs are trusted users whose primary objective is to reliably decode their intended confidential information streams. In contrast, the UEHRs are authorized network nodes that harvest RF energy to sustain their operations but do not have the security clearance to decode the confidential information intended for the IRs. Therefore, the UEHRs may act as potential non-colluding internal eavesdroppers, resulting in an inherent trade-off between wireless power transfer and PLS. To mitigate severe terrestrial blockages that prevent direct LoS communication between the BS and the ground receivers,

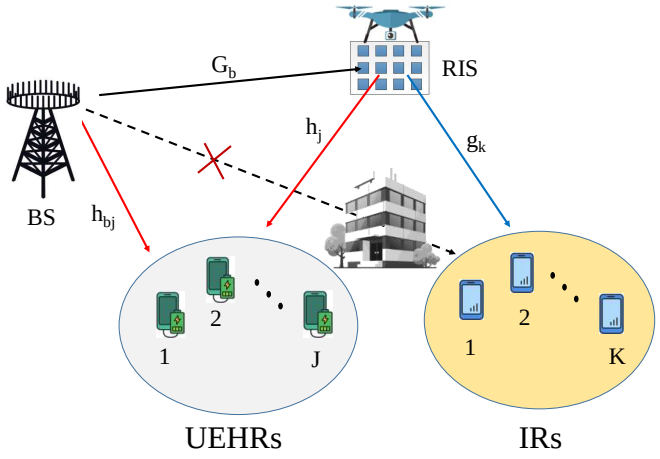


Fig. 1. System model of the UAV-assisted RIS-aided secure downlink network.

a UAV equipped with a RIS is deployed as an aerial passive relay. By jointly optimizing the RIS phase shifts, the UAV horizontal position, and the BS transmit power, the proposed system aims to enhance the reception of confidential signals at the IRs while satisfying the required EH demands of the UEHRs and limiting information leakage to them.

Consider a downlink wireless communication system consisting of one UAV-assisted RIS, J non-colluding UEHRs, K legitimate users, and one BS. The set of K IRs and J UEHRs are denoted by $\mathcal{K} = \{1, \ldots, K\}$ and $\mathcal{J} = \{1, \ldots, J\}$, respectively. Due to the presence of high obstacles, there exists no LoS channel between the BS and the IRs. The UAV, equipped with a RIS, serves as a passive relay to assist communication between the BS and the users. The RIS is composed of M reflecting elements, and the phase shift of each element can be controlled by the UAV. The BS is equipped with N_t antennas.

All UEHRs, IRs, and the BS are located on the ground. The horizontal coordinates of user $k \in \mathcal{K}$, UEHR $j \in \mathcal{J}$, and the BS are denoted by $\mathbf{w}_k^I = [x_k^I, y_k^I]^T$, $\mathbf{w}_j^U = [x_j^U, y_j^U]^T$, and $\mathbf{w}_b = [x_b, y_b]^T$, respectively. The UAV-mounted RIS flies at a fixed altitude H and hovers to enhance communication. The system model is illustrated in Fig. 1.

The UAV (RIS) horizontal coordinate is denoted by $\mathbf{q} = [x_r, y_r]^T$. To capture realistic flight restrictions, the UAV is

constrained to hover inside a given rectangular region

$$\mathcal{L} \triangleq \left\{ \mathbf{q} = [x_r, y_r]^T \,\middle|\, x^r_{\min} \le x_r \le x^r_{\max},\, y^r_{\min} \le y_r \le y^r_{\max} \right\}. \tag{1}$$

Thus, the UAV location must satisfy $\mathbf{q} \in \mathcal{L}$.

A. Channel Model

Let d_b denote the distance between the BS and the UAV, and d_k and d_j denote the distances between the UAV and user k and UEHR j, respectively, i.e.,

$$d_b = \sqrt{\|\mathbf{q} - \mathbf{w}_b\|^2 + H^2}, \tag{2}$$

$$d_k = \sqrt{\|\mathbf{q} - \mathbf{w}^I_k\|^2 + H^2}, \quad k \in \mathcal{K}, \tag{3}$$

$$d_j = \sqrt{\|\mathbf{q} - \mathbf{w}^U_j\|^2 + H^2}, \quad j \in \mathcal{J}. \tag{4}$$

We adopt a distance-dependent path-loss model with exponent $\alpha \ge 2$ and a Rician small-scale fading model for all air-to-ground links. Specifically, the channel between the UAV-mounted RIS and user k is given by $\mathbf{g}_k = \sqrt{\rho_0\, d_k^{-\alpha_k}}\, \tilde{\mathbf{g}}_k$, and

$$\tilde{\mathbf{g}}_k = \sqrt{\frac{K_k}{K_k+1}}\, \mathbf{g}_k^{\mathrm{LoS}} + \sqrt{\frac{1}{K_k+1}}\, \hat{\mathbf{g}}_k, \tag{5}$$

where ρ_0 is the path loss at the reference distance of 1 meter, $\hat{\mathbf{g}}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$, and K_k is the Rician K-factor. Similarly, the UAV–UEHR j channel is $\mathbf{h}_j = \sqrt{\rho_0\, d_j^{-\alpha_j}}\, \tilde{\mathbf{h}}_j$, and

$$\tilde{\mathbf{h}}_j = \sqrt{\frac{K_j}{K_j+1}}\, \mathbf{h}_j^{\mathrm{LoS}} + \sqrt{\frac{1}{K_j+1}}\, \hat{\mathbf{h}}_j, \tag{6}$$

where $\hat{\mathbf{h}}_j \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ and K_j is the Rician factor. The LoS components $\mathbf{g}_k^{\mathrm{LoS}}$ and $\mathbf{h}_j^{\mathrm{LoS}}$ are modeled via the RIS array steering vector. For simplicity, we adopt a uniform linear array (ULA) model for the RIS. This assumption does not affect the generality of the proposed framework, as it can be readily extended to a uniform planar array (UPA) by incorporating the corresponding three-dimensional (3D) array response. Specifically, assuming a ULA with inter-element spacing $\delta = \lambda/2$, where λ is the carrier wavelength, the steering vector of the RIS towards angle φ is

$$\mathbf{a}_{\mathrm{RIS}}(\varphi) = \left[1, e^{-j\frac{2\pi}{\lambda}\delta\sin\varphi}, \ldots, e^{-j\frac{2\pi}{\lambda}\delta(M-1)\sin\varphi}\right]^T. \tag{7}$$

The array is assumed parallel to the ground, so that the angular dependence of the array response is captured through the horizontal azimuth angle φ, while the UAV altitude H only affects the large-scale path loss via the 3D distance. Hence, $\mathbf{g}_k^{\mathrm{LoS}} = \mathbf{a}_{\mathrm{RIS}}(\varphi_k)$ and $\mathbf{h}_j^{\mathrm{LoS}} = \mathbf{a}_{\mathrm{RIS}}(\varphi_j)$, where φ_k and φ_j are the angles of departure from the RIS towards user k and UEHR j, respectively, determined by $\varphi_k = \arctan\!\left(\frac{y_k - y_r}{x_k - x_r}\right)$ and $\varphi_j = \arctan\!\left(\frac{y_j - y_r}{x_j - x_r}\right)$.

We define the RIS phase shift vector as $\boldsymbol{s} = [s_1, \ldots, s_M]^T$, where each element is given by $s_m = e^{j\theta_m}$ for all $m \in \{1, \ldots, M\}$. The RIS reflection-coefficient matrix is

$$\boldsymbol{\Theta} = \mathrm{diag}(s_1, \ldots, s_M) \in \mathbb{C}^{M \times M}, \tag{8}$$

In the ideal continuous-phase model, the phase shift satisfies $\theta_m \in [0, 2\pi)$. In practical implementations with finite-resolution phase shifters, θ_m is restricted to a discrete set

$$\theta_m \in \mathcal{F} \triangleq \left\{ 0,\, \frac{2\pi}{2^L},\, \frac{2\pi \cdot 2}{2^L},\, \ldots,\, \frac{2\pi(2^L-1)}{2^L} \right\}, \tag{9}$$

where L denotes the number of quantization bits of each RIS element. The channel matrix between the BS and the RIS is denoted by $\mathbf{G}_b \in \mathbb{C}^{M \times N_t}$, where $[\mathbf{G}_b]_{m,n}$ represents the channel coefficient between the nth transmit antenna at the BS and the mth reflecting element of the RIS. It is modeled as $\mathbf{G}_b = \sqrt{\rho_0\, d_b^{-\alpha_b}}\, \tilde{\mathbf{G}}_b$, and

$$\tilde{\mathbf{G}}_b = \left(\sqrt{\frac{K_b}{K_b+1}}\, \mathbf{G}_b^{\mathrm{LoS}} + \sqrt{\frac{1}{K_b+1}}\, \hat{\mathbf{G}}_b \right) \tag{10}$$

where $\hat{\mathbf{G}}_b \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{M \times N_t})$ and K_b is the Rician K-factor. The LoS component is given by

$$\mathbf{G}_b^{\mathrm{LoS}} = \mathbf{a}_{\mathrm{RIS}}(\varphi_{BR})\, \mathbf{a}_{\mathrm{BS}}^H(\varphi_{BR}), \tag{11}$$

with $\mathbf{a}_{\mathrm{BS}}(\cdot)$ denoting the BS array response, and φ_{BR} the angle of the BS–RIS link.

As a worst-case assumption for secrecy analysis, we also consider a direct link between the BS and each UEHR. The corresponding channel vector between the BS and the j-th UEHR is denoted by $\mathbf{h}_{bj} \in \mathbb{C}^{N_t \times 1}$ for all $j \in \mathcal{J}$, where the small-scale fading follows a Rayleigh model with

$$\mathbf{h}_{bj} \sim \mathcal{CN}\big(\mathbf{0}, \rho_0 d_{bj}^{-\alpha_{bj}} \mathbf{I}_{N_t}\big), \tag{12}$$

and $d_{bj} = \|\mathbf{w}_b - \mathbf{w}^U_j\|$ is the BS–UEHR distance.

B. Signal Model and ZF Beamforming

The BS employs linear ZF precoding to simultaneously serve K IRs. Let

$$\mathbf{H}_{\mathrm{c}} = \left[\, \mathbf{h}_{\mathrm{c},1}, \ldots, \mathbf{h}_{\mathrm{c},K} \,\right] \in \mathbb{C}^{N_t \times K}$$

denote the combined channel matrix of BS–RIS–IR, where

$$\mathbf{h}_{\mathrm{c},k} = \mathbf{G}_b^H \boldsymbol{\Theta}^H \mathbf{g}_k, \qquad k \in \mathcal{K}. \tag{13}$$

The unnormalized ZF precoding matrix is obtained from the Moore–Penrose pseudo-inverse as

$$\bar{\mathbf{P}} = \mathbf{H}_{\mathrm{c}} \left(\mathbf{H}_{\mathrm{c}}^H \mathbf{H}_{\mathrm{c}} \right)^{-1}. \tag{14}$$

The k-th column $\bar{\mathbf{p}}_k$ of $\bar{\mathbf{P}}$ satisfies $\mathbf{h}_{\mathrm{c},i}^H \bar{\mathbf{p}}_k = 0$ for all $i \ne k$. To incorporate transmit power allocation, each ZF precoder is normalized and scaled as

$$\mathbf{p}_k = \sqrt{\tilde{p}_k}\, \hat{\mathbf{p}}_k, \qquad k = 1, \ldots, K, \tag{15}$$

where $\tilde{p}_k \ge 0$ is the transmit power allocated to user k and $\hat{\mathbf{p}}_k$ is the corresponding normalized ZF beam direction defined as

$$\hat{\mathbf{p}}_k \triangleq \frac{\bar{\mathbf{p}}_k}{\|\bar{\mathbf{p}}_k\|}, \qquad k = 1, \ldots, K. \tag{16}$$

The transmitted signal from the BS is therefore

$$\mathbf{x} = \sum_{k=1}^{K} \mathbf{p}_k\, \tilde{s}_k, \tag{17}$$

where $\tilde{s}_k$ is a unit-power information symbol intended for IR k. Using (17), the received baseband signal at IR k is

$$y_k = \mathbf{h}_{\mathrm{c},k}^H \mathbf{x} + n_k, \tag{18}$$

where $n_k \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise (AWGN). Following the ZF property that $\mathbf{h}_{\mathrm{c},k}^H \hat{\mathbf{p}}_l = 0$ for all $\ell \neq k$, the signal-to-interference-plus-noise ratio (SINR) at the legitimate user k, obtained as

$$\gamma_k = \frac{\left|\mathbf{h}_{\mathrm{c},k}^H \mathbf{p}_k\right|^2}{\sum\limits_{\ell \neq k} \left|\mathbf{h}_{\mathrm{c},k}^H \mathbf{p}_\ell\right|^2 + \sigma^2} = \frac{\tilde{p}_k a_k}{\sigma^2}. \tag{19}$$

where $a_k \triangleq \left|\mathbf{h}_{\mathrm{c},k}^H \hat{\mathbf{p}}_k\right|^2$. The received signal at UEHR j is

$$y_j = \left(\mathbf{h}_{bj}^H + \mathbf{h}_j^H \mathbf{\Theta} \mathbf{G}_b\right)\mathbf{x} + n_j. \tag{20}$$

For simplicity of notations we define $\mathbf{u}_j \triangleq \mathbf{h}_{bj} + \mathbf{G}_b^H \mathbf{\Theta}^H \mathbf{h}_j$. Then the eavesdropping SINR for decoding user k is

$$\gamma_{j,k} = \frac{\left|\mathbf{u}_j^H \mathbf{p}_k\right|^2}{\sum_{\ell \neq k} \left|\mathbf{u}_j^H \mathbf{p}_\ell\right|^2 + \sigma^2} = \frac{\tilde{p}_k b_{j,k}}{\sum_{\ell \neq k} \tilde{p}_\ell b_{j,\ell} + \sigma^2}, \tag{21}$$

where $b_{j,k} \triangleq |\mathbf{u}_j^H \hat{\mathbf{p}}_k|^2$.

C. Worst-Case Secrecy Rate Under CSI Uncertainty

The BS is assumed to possess perfect CSI (PCSI) for the legitimate cascaded links $\mathbf{h}_{\mathrm{c},k}$, which can be reliably acquired in time-division duplexing (TDD) systems via uplink pilot training and standard cascaded channel estimation protocols (e.g., [35], [36]). Due to high-elevation line-of-sight propagation, these aerial links exhibit high spatial correlation and change slowly with time, significantly reducing pilot overhead [37]. In contrast, UEHRs are untrusted nodes that do not engage in reliable channel reporting. Thus, only imperfect estimates $\hat{\mathbf{u}}_j$ are available at the BS [16].

To capture this inaccuracy, we adopt the widely used norm-bounded CSI uncertainty model [16], [18], [19], [25]. Let $\hat{\mathbf{u}}_j \in \mathbb{C}^{N_t \times 1}$ denote the estimated cascaded BS–RIS–UEHR channel for UEHR j, and let $\mathbf{u}_j$ be the true (unknown) channel. The imperfect CSI (ICSI) model is

$$\mathbf{u}_j = \hat{\mathbf{u}}_j + \Delta\mathbf{u}_j, \qquad \Delta\mathbf{u}_j \in \mathcal{U}, \tag{22}$$

where the uncertainty set is

$$\mathcal{U} \triangleq \left\{\Delta\mathbf{u}_j \in \mathbb{C}^{N_t \times 1} :\ \|\Delta\mathbf{u}_j\|_2 \leq \nu\right\}. \tag{23}$$

Here, $\nu > 0$ defines the radius of the CSI error region and reflects the UEHR's infrequent or unreliable feedback. When $\nu = 0$, the BS has perfect knowledge of the UEHR channel. This norm-bounded uncertainty model enables a worst-case secrecy formulation, ensuring robustness against the strongest possible UEHR eavesdropping behavior within the uncertainty region.

Since $\mathbf{u}_j$ is imperfectly known, the BS must ensure robustness against the *worst-case eavesdropping capability* within the uncertainty region. To account for channel uncertainty, we adopt standard norm-bounded channel uncertainty results. For any channel model $\mathbf{h} = \hat{\mathbf{h}} + \Delta\mathbf{h}$ with $\|\Delta\mathbf{h}\|_2 \leq \nu$ and for any vector $\mathbf{p} \in \mathbb{C}^n$, the following bounds hold (see [16], [24])

$$\max_{\|\Delta\mathbf{h}\| \leq \nu} |\mathbf{h}^H \mathbf{p}| = |\hat{\mathbf{h}}^H \mathbf{p}| + \nu\|\mathbf{p}\|_2, \tag{24a}$$

$$\min_{\|\Delta\mathbf{h}\| \leq \nu} |\mathbf{h}^H \mathbf{p}| = \left[|\hat{\mathbf{h}}^H \mathbf{p}| - \nu\|\mathbf{p}\|_2\right]^+, \tag{24b}$$

where $[x]^+ = \max(x, 0)$. Applying (24) to the numerator and denominator terms of the eavesdropping SINR in (21) yields

$$\max_{\|\Delta\mathbf{u}_j\| \leq \nu} \left|\mathbf{u}_j^H \mathbf{p}_k\right|^2 = \left(\left|\hat{\mathbf{u}}_j^H \mathbf{p}_k\right| + \nu\|\mathbf{p}_k\|\right)^2, \tag{25}$$

$$\min_{\|\Delta\mathbf{u}_j\| \leq \nu} \left|\mathbf{u}_j^H \mathbf{p}_\ell\right|^2 = \left(\left[|\hat{\mathbf{u}}_j^H \mathbf{p}_\ell| - \nu\|\mathbf{p}_\ell\|\right]^+\right)^2, \quad \ell \neq k. \tag{26}$$

Accordingly, a conservative upper bound on the worst-case eavesdropping SINR for user k at UEHR j is obtained as [1]

$$\begin{aligned}
\gamma_{E,k} &= \max_{j \in \mathcal{J}} \left\{ \max_{\|\Delta\mathbf{u}_j\| \leq \nu} \hat{\gamma}_{j,k}(\Delta\mathbf{u}_j) \right\} \\
&\leq \max_{j \in \mathcal{J}} \frac{\left(\left|\hat{\mathbf{u}}_j^H \mathbf{p}_k\right| + \nu\|\mathbf{p}_k\|\right)^2}{\sum_{\ell \neq k} \left(\left[|\hat{\mathbf{u}}_j^H \mathbf{p}_\ell| - \nu\|\mathbf{p}_\ell\|\right]^+\right)^2 + \sigma^2} \\
&= \max_{j \in \mathcal{J}} \frac{\tilde{p}_k \left(\sqrt{\hat{b}_{j,k}} + \nu\right)^2}{\sum_{\ell \neq k} \tilde{p}_\ell \left(\left[\sqrt{\hat{b}_{j,\ell}} - \nu\right]^+\right)^2 + \sigma^2},
\end{aligned} \tag{27}$$

where $\hat{b}_{j,k} \triangleq \left|\hat{\mathbf{u}}_j^H \hat{\mathbf{p}}_k\right|^2$, $\forall j \in \mathcal{J}$, $k \in \mathcal{K}$. Given the legitimate-user SINR γ_k and the upper bound on the worst-case eavesdropping SINR $\gamma_{E,k}$ in (27), the worst-case secrecy rate of IR k is expressed as

$$R_k^{\mathrm{sec}} = \left[\log_2(1 + \gamma_k) - \log_2(1 + \gamma_{E,k})\right]^+. \tag{28}$$

This formulation guarantees secrecy robustness against all UEHR channel realizations satisfying $\|\Delta\mathbf{u}_j\| \leq \nu$.

The total received RF power at UEHR j is

$$P_j^{\mathrm{EH}} = \sum_{k=1}^{K} \left|\mathbf{u}_j^H \mathbf{p}_k\right|^2, \tag{29}$$

and the harvested direct current (DC) power follows the logistic nonlinear model [38]

$$\Omega(P_j^{\mathrm{EH}}) = \frac{b_2}{k_1'\left(1 + \exp(-b_0(P_j^{\mathrm{EH}} - b_1))\right)} - k_2'. \tag{30}$$

where $\Omega(\cdot)$ is a logistic function representing the non-linear energy harvesting model, and k_1', k_2', b_0, b_1, b_2 are positive constants. The required EH power must satisfy the following robust constraint under channel uncertainty

$$\sum_{j \in \mathcal{J}} \min_{\|\Delta\mathbf{u}_j\| \leq \nu} P_j^{\mathrm{EH}} \geq \Omega^{-1}(E_h), \tag{31}$$

where E_h denotes the minimum harvested energy requirement at the set of UEHRs. The inverse function $\Omega^{-1}(x)$ is defined as

$$\Omega^{-1}(x) = b_1 - \frac{1}{b_0} \ln\left(\frac{b_2}{k_1'(x + k_2')} - 1\right). \tag{32}$$

[1]This conservative bound ensures robustness under channel uncertainty at the cost of performance loss.

To obtain a tractable deterministic formulation, (31) is enforced via the following sufficient lower-bound constraint

$$\sum_{j\in\mathcal{J}} \underline{P}_j^{\mathrm{EH}} \;\geq\; \Omega^{-1}(E_h), \tag{33}$$

where

$$\underline{P}_j^{\mathrm{EH}} \triangleq \sum_{k=1}^{K} \Big(\big[\, \big|\hat{\mathbf{u}}_j^H \mathbf{p}_k\big| - \nu\|\mathbf{p}_k\|_2 \,\big]^+ \Big)^2, \qquad \forall j \in \mathcal{J}. \tag{34}$$

D. Problem Formulation

To guarantee confidentiality against UEHRs, the secrecy rate of each legitimate IR is defined under worst-case eavesdropping as in (28). To ensure fairness among users, we adopt the minimum secrecy rate

$$R^{\mathrm{sec}} = \min_{k\in\mathcal{K}} R_k^{\mathrm{sec}}. \tag{35}$$

We aim to maximize the WCSEE, defined as the ratio between the minimum secrecy rate and the total power consumption at the BS. The total power includes the radiated transmit power and the circuit power consumption. The optimization variables include the ZF power allocation, RIS phase shifts, and the UAV horizontal position. The resulting problem is formulated as

$$\max_{\tilde{\mathbf{p}},\boldsymbol{\theta},\mathbf{q}} \quad \frac{R^{\mathrm{sec}}}{\varrho\sum_{k=1}^{K}\tilde{p}_k + P_0} \tag{36a}$$

$$\text{s.t.} \quad \sum_{k=1}^{K}\tilde{p}_k \leq P_{\max}, \tag{36b}$$

$$\theta_m \in \mathcal{F}, \quad m = 1,\ldots,M, \tag{36c}$$

$$\mathbf{q} \in \mathcal{L}, \tag{36d}$$

$$(33). \tag{36e}$$

where ϱ is the reciprocal of the drain efficiency of the power amplifier at the BS [27], $\tilde{\mathbf{p}} = \{\tilde{p}_k\}_{k=1}^K$, $\boldsymbol{\theta} = \{\theta_m\}_{m=1}^M$ and P_0 represents the circuit power consumption.

Remark 1. *Problem* (36) *is highly nonconvex due to the fractional objective, coupled UAV–RIS channel geometry, worst-case secrecy constraints, and discrete RIS phase shifts. This motivates the use of advanced optimization and learning-based approaches to handle the resulting high-dimensional problem.*

III. SCA-Based Solution

To obtain a tractable benchmark, we consider an idealized setting with perfect CSI and continuous RIS phase shifts. Under these assumptions, problem (36) admits a simplified formulation that can be efficiently solved via a BCD framework combined with SCA, and thus reduces to

$$\max_{\tilde{\mathbf{p}},\boldsymbol{\theta},\mathbf{q}} \quad \frac{R^{\mathrm{sec}}}{\varrho\sum_{k=1}^{K}\tilde{p}_k + P_0} \tag{37a}$$

$$\text{s.t.} \quad \sum_{j\in\mathcal{J}} P_j^{\mathrm{EH}} \geq \Omega^{-1}(E_h), \tag{37b}$$

$$\theta_m \in [0,2\pi], \quad \forall m \in \{1,\ldots,M\}, \tag{37c}$$

$$(36b), (36d). \tag{37d}$$

Problem (37) is addressed using a BCD framework. Specifically, the power allocation, RIS reflection coefficients, and UAV location are optimized in an alternating manner, while the remaining variables are fixed in each subproblem.

A. Power Allocation

The RIS configuration $\boldsymbol{\Theta}$ and the UAV location $\mathbf{q}$ are assumed fixed. Accordingly, the BS computes the ZF beamforming directions from the effective channel and only optimizes the power loading. Given fixed $\boldsymbol{\Theta}$ and $\mathbf{q}$, the power allocation benchmark problem is formulated as

$$\max_{\tilde{\mathbf{p}}\succeq\mathbf{0}} \quad \frac{R^{\mathrm{sec}}(\tilde{\mathbf{p}})}{\varrho\sum_{k=1}^{K}\tilde{p}_k + P_0} \tag{38a}$$

$$\text{s.t.} \quad \sum_{j\in\mathcal{J}}\sum_{k=1}^{K}\tilde{p}_k\, b_{j,k} \geq \Omega^{-1}(E_h), \tag{38b}$$

$$(36b). \tag{38c}$$

In (38b), we have used (15), under which the harvested RF power at UEHR j is given by

$$P_j^{\mathrm{EH}}(\tilde{\mathbf{p}}) = \sum_{k=1}^{K}\tilde{p}_k\, b_{j,k}. \tag{39}$$

As a result, the EH constraint in (38b) is affine with respect to the power allocation vector $\tilde{\mathbf{p}}$.

Problem (38) is a nonlinear fractional program due to the fractional objective and the nonconvex secrecy-rate term. Introducing a slack variable ζ, the secrecy constraint can be equivalently enforced for all UEHRs as

$$\zeta \leq \log_2\Big(\sigma^2 + \tilde{p}_k a_k\Big) - \log_2(\sigma^2) + \log_2\Big(\sigma^2 + \sum_{\ell\neq k}\tilde{p}_\ell b_{j,\ell}\Big) - \log_2\Big(\sigma^2 + \sum_{\ell=1}^{K}\tilde{p}_\ell b_{j,\ell}\Big), \quad \forall k\in\mathcal{K}, \forall j\in\mathcal{J}. \tag{40}$$

The last term in (40) is upper bounded by its first-order Taylor expansion at $\tilde{\mathbf{p}}^{(t)}$ as

$$\log_2\Big(\sigma^2 + \sum_{\ell=1}^{K}\tilde{p}_\ell b_{j,\ell}\Big) \leq \log_2\big(S_j^{(t)}\big) + \sum_{\ell=1}^{K}\eta_{j,\ell}^{(t)}\big(\tilde{p}_\ell - \tilde{p}_\ell^{(t)}\big), \tag{41}$$

where $S_j^{(t)} = \sigma^2 + \sum_{\ell=1}^{K}\tilde{p}_\ell^{(t)} b_{j,\ell}$ and $\eta_{j,\ell}^{(t)} = b_{j,\ell}/(\ln(2) S_j^{(t)})$. Substituting (41) into (40) gives the concave lower bound

$$\zeta \leq \widehat{R}_{k,j}^{(t)}(\tilde{\mathbf{p}}), \quad \forall k\in\mathcal{K}, \forall j\in\mathcal{J}, \tag{42}$$

where

$$\widehat{R}_{k,j}^{(t)}(\tilde{\mathbf{p}}) \triangleq \log_2\Big(\sigma^2 + \tilde{p}_k a_k\Big) - \log_2(\sigma^2) - \log_2\big(S_j^{(t)}\big) + \log_2\Big(\sigma^2 + \sum_{\ell\neq k}\tilde{p}_\ell b_{j,\ell}\Big) - \sum_{\ell=1}^{K}\eta_{j,\ell}^{(t)}\big(\tilde{p}_\ell - \tilde{p}_\ell^{(t)}\big). \tag{43}$$

To handle the fractional objective, we apply the Charnes–Cooper transformation

$$\chi = \frac{1}{\varrho\sum_{k=1}^{K}\tilde{p}_k + P_0}, \qquad \bar{p}_k = \chi\tilde{p}_k, \qquad \bar{\zeta} = \chi\zeta. \tag{44}$$

Algorithm 1: Charnes–Cooper-SCA Power Allocation

Input: $\tilde{\mathbf{p}}^{(0)}$, tolerance ε.
Output: $\tilde{\mathbf{p}}^{\star}$

Set $t=0$ and $\Delta=\infty$;
repeat
  Construct $\widehat{R}_{k,j}^{(t)}(\tilde{\mathbf{p}})$ using (43);
  Solve (45) and recover $\tilde{p}_k^{(t+1)}=\bar{p}_k^{\star}/\chi^{\star}$, $\forall k$;
  Set $\Delta=\|\tilde{\mathbf{p}}^{(t+1)}-\tilde{\mathbf{p}}^{(t)}\|_2$ and $t\leftarrow t+1$;
until $\Delta\le\varepsilon$;
return $\tilde{\mathbf{p}}^{(t)}$.

Thus, at SCA iteration t, the power-allocation subproblem (38) is transformed into the following convex problem

$$\max_{\bar{\mathbf{p}}\succeq\mathbf{0},\,\bar{\zeta},\,\chi>0}\quad \bar{\zeta} \tag{45a}$$

$$\text{s.t.}\quad \varrho\sum_{k=1}^{K}\bar{p}_k+P_0\chi=1, \tag{45b}$$

$$\sum_{k=1}^{K}\bar{p}_k\le P_{\max}\chi, \tag{45c}$$

$$\sum_{j\in\mathcal{J}}\sum_{k=1}^{K}\bar{p}_k b_{j,k}\ge\chi\Omega^{-1}(E_h), \tag{45d}$$

$$\bar{\zeta}\le\chi\widehat{R}_{k,j}^{(t)}\left(\frac{\bar{\mathbf{p}}}{\chi}\right),\quad \forall k\in\mathcal{K},\forall j\in\mathcal{J}. \tag{45e}$$

Problem (45) is convex since $\chi\widehat{R}_{k,j}^{(t)}(\bar{\mathbf{p}}/\chi)$ is the perspective of the concave lower bound in (42), and the remaining constraints are affine. After solving (45), the original power allocation is recovered as $\tilde{p}_k^{\star}=\bar{p}_k^{\star}/\chi^{\star}$. Unlike Dinkelbach's method [28], the Charnes–Cooper transformation avoids an additional fractional-programming loop and requires only one convex solve per SCA iteration. This leads to a more compact and computationally efficient power-allocation procedure, as summarized in Algorithm 1.

B. RIS Reflection Phase Optimization

In this subsection, we optimize the RIS reflection coefficients for fixed ZF beamforming directions, power allocation $\tilde{\mathbf{p}}$, and UAV location $\mathbf{q}$. Since the transmit power consumption at the BS is fixed under given $\tilde{\mathbf{p}}$, the WCSEE maximization reduces to a worst-case secrecy rate maximization problem. Accordingly, the RIS phase optimization subproblem can be formulated as

$$\max_{\mathbf{s},\,\zeta}\quad \zeta \tag{46a}$$

$$\text{s.t.}\quad R_k^{\mathrm{sec}}\ge\zeta,\quad \forall k\in\mathcal{K}, \tag{46b}$$

$$(37\text{b}),(37\text{c}). \tag{46c}$$

For any vectors $\mathbf{v}\in\mathbb{C}^M$ and $\mathbf{w}\in\mathbb{C}^{N_t}$, the cascaded channel term can be rewritten as $\mathbf{v}^H\boldsymbol{\Theta}\mathbf{G}_b\mathbf{w}=\mathbf{t}^H\mathbf{s}$, where $\mathbf{t}\triangleq(\mathrm{diag}(\mathbf{v}^H)\mathbf{G}_b\mathbf{w})^*$. Accordingly, we define

$$\mathbf{t}_{k,\ell}\triangleq\left(\mathrm{diag}(\mathbf{g}_k^H)\mathbf{G}_b\mathbf{p}_\ell\right)^*,\qquad \forall k\in\mathcal{K},\ \forall\ell\in\mathcal{K}, \tag{47}$$

$$\mathbf{t}_{j,\ell}\triangleq\left(\mathrm{diag}(\mathbf{h}_j^H)\mathbf{G}_b\mathbf{p}_\ell\right)^*,\qquad \forall j\in\mathcal{J},\ \forall\ell\in\mathcal{K}, \tag{48}$$

$$c_{j,\ell}\triangleq\mathbf{h}_{bj}^H\mathbf{p}_\ell,\qquad \forall j\in\mathcal{J},\ \forall\ell\in\mathcal{K}. \tag{49}$$

Then the useful and interference terms become

$$\mathbf{h}_{\mathrm{c},k}^H\mathbf{p}_\ell=\mathbf{g}_k^H\boldsymbol{\Theta}\mathbf{G}_b\mathbf{p}_\ell=\mathbf{t}_{k,\ell}^H\mathbf{s}, \tag{50}$$

$$\mathbf{u}_j^H\mathbf{p}_\ell=\mathbf{h}_{bj}^H\mathbf{p}_\ell+\mathbf{h}_j^H\boldsymbol{\Theta}\mathbf{G}_b\mathbf{p}_\ell=c_{j,\ell}+\mathbf{t}_{j,\ell}^H\mathbf{s}. \tag{51}$$

Using the definitions above, the received signal powers appearing in γ_k and $\gamma_{j,k}$ can be expressed as quadratic forms in $\mathbf{s}$. The secrecy constraint (46b) is nonconvex due to the difference of logarithmic functions. Following the SCA principle, we introduce first-order approximations of the nonconvex quadratic terms. Introduce auxiliary variables ρ_k and $\rho_{E,j,k}$ such that

$$\rho_k\le\gamma_k,\qquad \forall k\in\mathcal{K}, \tag{52a}$$

$$\rho_{E,j,k}\ge\gamma_{j,k},\qquad \forall j\in\mathcal{J},\ \forall k\in\mathcal{K}. \tag{52b}$$

Also introduce $f_{E,k}$ to handle the worst UEHR,

$$f_{E,k}\ge\log_2(1+\rho_{E,j,k}),\qquad \forall j\in\mathcal{J},\ \forall k\in\mathcal{K}, \tag{53}$$

Proposition 1. *An affine approximation of the secrecy-related constraints in* (46b)*, for all $k\in\mathcal{K}$ and $j\in\mathcal{J}$, is given by*

$$\sum_{\ell\neq k}\left|\mathbf{t}_{k\ell}^H\mathbf{s}\right|^2+\sigma^2-\Psi^{(t)}\big(\mathbf{s},\rho_k;\mathbf{t}_{kk}\big)\le 0,\quad \forall k\in\mathcal{K}, \tag{54a}$$

$$\left|c_{j,k}+\mathbf{t}_{j,k}^H\mathbf{s}\right|^2\le\Theta^{(t)}\big(\xi_{j,k},\rho_{E,j,k}\big),\quad \forall j\in\mathcal{J},\ \forall k\in\mathcal{K}, \tag{54b}$$

$$\xi_{j,k}\le\sum_{\ell\neq k}\vartheta^{(t)}\big(c_{j,\ell},\mathbf{t}_{j,\ell};\mathbf{s}\big)+\sigma^2,\quad \forall j\in\mathcal{J},\ \forall k\in\mathcal{K}, \tag{54c}$$

$$1-\Gamma^{(t)}(f_{E,k})+\rho_{E,j,k}\le 0,\quad \forall j\in\mathcal{J},\ \forall k\in\mathcal{K}, \tag{54d}$$

$$\zeta\le\log_2(1+\rho_k)-f_{E,k},\quad \forall k\in\mathcal{K}. \tag{54e}$$

where $\Psi^{(t)}(\mathbf{s},\rho_k;\mathbf{t}_{kk})$ is the first-order Taylor approximation of $\frac{\left|\mathbf{t}_{kk}^H\mathbf{s}\right|^2}{\rho_k}$, $\Gamma^{(t)}(x)\triangleq 2^{x^{(t-1)}}\big[1+\ln(2)\,(x-x^{(t-1)})\big]$ is the first-order Taylor approximation of 2^x, and

$$\Theta^{(t)}(x,y)\triangleq\tfrac{1}{2}\big(x^{(t-1)}+y^{(t-1)}\big)(x+y)-\tfrac{1}{4}\big(x^{(t-1)}+y^{(t-1)}\big)^2-\tfrac{1}{4}(x-y)^2, \tag{55}$$

is used to obtain an affine lower bound of the bilinear term xy. Moreover, $\vartheta^{(t)}(c;\mathbf{t};\mathbf{s})$ is the first-order Taylor approximation of $\left|c+\mathbf{t}^H\mathbf{s}\right|^2$ at iteration t:

$$\vartheta^{(t)}(c;\mathbf{t};\mathbf{s})=2\,\mathfrak{Re}\Big\{\big(c+\mathbf{t}^H\mathbf{s}^{(t-1)}\big)^H\mathbf{t}^H\mathbf{s}\Big\}-\left|c+\mathbf{t}^H\mathbf{s}^{(t-1)}\right|^2. \tag{56}$$

Proof. See Appendix A. ∎

Constraint (37b) is nonconvex due to the quadratic terms in $\mathbf{s}$, and the unit-modulus constraint $|s_m|=1$ is also nonconvex. In the following, we construct convex surrogate constraints using SCA. Under fixed $\{\mathbf{p}_k\}$ and $\mathbf{q}$, the harvested RF power at UEHR j can be written as

$$P_j^{\mathrm{EH}}=\sum_{k=1}^{K}\left|c_{j,k}+\mathbf{t}_{j,k}^H\mathbf{s}\right|^2. \tag{57}$$

Since $\left|c_{j,k}+\mathbf{t}_{j,k}^H\mathbf{s}\right|^2$ is convex in $\mathbf{s}$, its first-order Taylor expansion $\vartheta^{(t)}(c_{j,k};\mathbf{t}_{j,k};\mathbf{s})$ provides a global affine lower bound. Hence, an SCA-safe sufficient condition for (37b) is

$$\sum_{j\in\mathcal{J}}\sum_{k=1}^{K}\vartheta^{(t)}(c_{j,k};\mathbf{t}_{j,k};\mathbf{s})\ge\Omega^{-1}(E_h). \tag{58}$$

Algorithm 2: SCA-Based RIS Phase Optimization (Given $\tilde{\mathbf{p}}$ and $\mathbf{q}$)

Input: Initial $\mathbf{s}^{(0)}$, tolerance $\varepsilon > 0$, penalty weight $C > 0$.
Output: Optimized RIS phase vector $\mathbf{s}^\star$.
Set $t \leftarrow 1$.;
repeat
  Solve (59) to obtain $(\mathbf{s}^{(t)}, \zeta^{(t)})$.;
  $t \leftarrow t + 1$.;
until $|\zeta^{(t)} - \zeta^{(t-1)}| \le \varepsilon$;
return $\mathbf{s}^{(t-1)}$.;

In the ideal continuous-phase model, the RIS coefficients satisfy $|s_m| = 1$, $\forall m$. Following the penalty-based approach [26], we relax $|s_m| = 1$ to $|s_m| \le 1$ and add a penalty term to encourage $|s_m|$ approaches 1. Given $\mathbf{s}^{(t-1)}$, we solve the following convex problem

$$\max_{\boldsymbol{\Xi}} \quad \zeta + C\Big(2\,\mathrm{Re}\Big\{(\mathbf{s}^{(t-1)})^H\mathbf{s}\Big\} - \|\mathbf{s}^{(t-1)}\|^2\Big) \tag{59a}$$

$$\text{s.t.} \quad (54a) - (54e), (58), \tag{59b}$$

$$|s_m| \le 1, \quad \forall m = 1, \ldots, M, \tag{59c}$$

where the optimization variable set is defined as $\boldsymbol{\Xi} \triangleq \{\mathbf{s}, \zeta, \boldsymbol{\rho}, \boldsymbol{\rho}_E, \mathbf{f}_E, \boldsymbol{\xi}\}$ with $\boldsymbol{\rho} \triangleq \{\rho_k\}_{k\in\mathcal{K}}$, $\boldsymbol{\rho}_E \triangleq \{\rho_{E,j,k}\}_{j\in\mathcal{J},\,k\in\mathcal{K}}$, $\mathbf{f}_E \triangleq \{f_{E,k}\}_{k\in\mathcal{K}}$ and $\boldsymbol{\xi} \triangleq \{\xi_{j,k}\}_{j\in\mathcal{J},\,k\in\mathcal{K}}$. The RIS phase optimization procedure is summarized in Algorithm 2.

C. UAV Location Optimization

In this subsection, we optimize the UAV horizontal position $\mathbf{q}$ while keeping the BS precoders $\{\mathbf{p}_k\}_{k\in\mathcal{K}}$, and the RIS phase matrix $\boldsymbol{\Theta}$ fixed. Under fixed $\{\mathbf{p}_k\}$, the BS transmit power is constant; hence, the SEE maximization reduces to maximizing the minimum secrecy rate. To obtain a numerically stable and low-complexity SCA benchmark, we further consider the relaxed objective of maximizing the minimum rate of IRs instead of the secrecy rate.

Using the distance definitions in (2)–(4), the channels can be expressed as

$$\mathbf{g}_k(\mathbf{q}) = \sqrt{\rho_0}\, d_k(\mathbf{q})^{-\frac{\alpha}{2}}\, \tilde{\mathbf{g}}_k, \qquad \forall k \in \mathcal{K}, \tag{60}$$

$$\mathbf{h}_j(\mathbf{q}) = \sqrt{\rho_0}\, d_j(\mathbf{q})^{-\frac{\alpha}{2}}\, \tilde{\mathbf{h}}_j, \qquad \forall j \in \mathcal{J}, \tag{61}$$

$$\mathbf{G}_b(\mathbf{q}) = \sqrt{\rho_0}\, d_b(\mathbf{q})^{-\frac{\alpha}{2}}\, \tilde{\mathbf{G}}_b. \tag{62}$$

where $\alpha = \alpha_b = \alpha_k = \alpha_j$ for notational simplicity. Define the following fixed gains that incorporate the beamforming vectors $\{\mathbf{p}_k\}$:

$$\tilde{A}_{k,\ell} \triangleq \rho_0^2 \Big|\tilde{\mathbf{g}}_k^H \boldsymbol{\Theta} \tilde{\mathbf{G}}_b \mathbf{p}_\ell\Big|^2, \qquad \forall k, \ell \in \mathcal{K}, \tag{63}$$

$$\tilde{D}_{j,\ell} \triangleq \rho_0 \tilde{\mathbf{h}}_j^H \boldsymbol{\Theta} \tilde{\mathbf{G}}_b \mathbf{p}_\ell, \qquad \forall j \in \mathcal{J},\ \forall \ell \in \mathcal{K}. \tag{64}$$

Since both the desired signal and interference powers scale with the product $(d_k(\mathbf{q})d_b(\mathbf{q}))^{-\alpha}$, it is convenient to decouple the distance-dependent terms. To this end, we introduce slack variables $\{y_{kb}\}$ and $\{y_{jb}\}$ such that

$$d_k(\mathbf{q})\, d_b(\mathbf{q}) \le y_{kb}^{\frac{1}{\alpha}}, \qquad \forall k \in \mathcal{K}, \tag{65a}$$

$$d_j(\mathbf{q})\, d_b(\mathbf{q}) \le y_{jb}^{\frac{1}{\alpha}}, \qquad \forall j \in \mathcal{J}. \tag{65b}$$

Although the function $y^{1/\alpha}$ is concave for $\alpha \ge 2$, the constraints in (65) remain nonconvex due to the bilinear terms $d_k(\mathbf{q})d_b(\mathbf{q})$ and $d_j(\mathbf{q})d_b(\mathbf{q})$. To construct a convex surrogate, we employ the arithmetic–geometric mean (AGM) inequality. Specifically, at iteration t, for any nonnegative variables a and b, the product ab is upper bounded as

$$ab \le \widetilde{\Theta}^{(t)}(a,b) \triangleq \frac{1}{2}\left(\frac{b^{(t-1)}}{a^{(t-1)}}a^2 + \frac{a^{(t-1)}}{b^{(t-1)}}b^2\right), \tag{66}$$

where $(a^{(t-1)}, b^{(t-1)})$ denotes the point from the previous iteration. The surrogate $\widetilde{\Theta}^{(t)}(a,b)$ is convex and tight at $(a^{(t-1)}, b^{(t-1)})$. Consequently, the slack-variable constraints admit the following convex SCA approximation

$$\widetilde{\Theta}^{(t)}(d_k(\mathbf{q}), d_b(\mathbf{q})) \le y_{kb}^{\frac{1}{\alpha}}, \ \forall k, \tag{67}$$

$$\widetilde{\Theta}^{(t)}(d_j(\mathbf{q}), d_b(\mathbf{q})) \le y_{jb}^{\frac{1}{\alpha}}, \ \forall j. \tag{68}$$

By multiplying both the numerator and denominator of the SINR expression by $(d_k(\mathbf{q})d_b(\mathbf{q}))^\alpha$ and using the slack variable definition in (65), the legitimate-user SINR can be conservatively represented as

$$\gamma_k(\mathbf{q}) = \frac{\tilde{A}_{k,k}}{\sum_{\ell\neq k}\tilde{A}_{k,\ell} + \sigma^2\, y_{kb}}, \qquad \forall k \in \mathcal{K}. \tag{69}$$

Introduce auxiliary SINR variables $\{\tilde{\rho}_k\}$ and the rate variable ζ. Using (69), the rate constraints can be equivalently written as

$$\zeta \le \log_2(1 + \tilde{\rho}_k), \qquad \forall k \in \mathcal{K}, \tag{70a}$$

$$\tilde{\rho}_k\left(\sum_{\ell\neq k}\tilde{A}_{k,\ell} + \sigma^2 y_{kb}\right) \le \tilde{A}_{k,k}, \qquad \forall k \in \mathcal{K}. \tag{70b}$$

The bilinear term $\tilde{\rho}_k y_{kb}$ in (70b) is nonconvex. At iteration t, it is upper bounded by the convex AGM inequality introduced in the previous subsection as

$$\tilde{\rho}_k y_{kb} \le \widetilde{\Theta}^{(t)}(\tilde{\rho}_k, y_{kb}) \tag{71}$$

which is convex in $(\tilde{\rho}_k, y_{kb})$ and tight at $(\tilde{\rho}_k^{(t-1)}, y_{kb}^{(t-1)})$. Hence, a convex surrogate of (70b) is

$$\tilde{\rho}_k \sum_{\ell\neq k}\tilde{A}_{k,\ell} + \sigma^2\, \widetilde{\Theta}^{(t)}(\tilde{\rho}_k, y_{kb}) \le \tilde{A}_{k,k}, \qquad \forall k \in \mathcal{K}. \tag{72}$$

The harvested RF power at UEHR j is

$$P_j^{\mathrm{EH}}(\mathbf{q}) = \sum_{\ell=1}^{K}\left|c_{j,\ell} + x_{jb}\tilde{D}_{j,\ell}\right|^2, \qquad \forall j \in \mathcal{J}, \tag{73}$$

where $x_{jb} \triangleq \big(d_j(\mathbf{q})d_b(\mathbf{q})\big)^{-\alpha/2}$ is the cascaded path-loss factor. The nonconvexity of (73) stems from the distance-dependent term x_{jb}. To obtain a numerically stable SCA formulation, we eliminate x_{jb} by expressing the bound in

Algorithm 3: SCA-Based UAV Location Optimization

Input: Initial $\mathbf{q}^{(0)} \in \mathcal{L}$, tolerance $\varepsilon > 0$.
Output: Optimized UAV location $\mathbf{q}^\star$.
Set $t \leftarrow 1$ and initialize $(\tilde{\boldsymbol{\rho}}^{(0)}, \mathbf{y}_b^{(0)}, \mathbf{y}_u^{(0)})$ from $\mathbf{q}^{(0)}$.;
repeat
  Construct $\widetilde{\Theta}^{(t)}(\cdot)$ and $\widetilde{\psi}_j^{(t)}(\cdot)$ using (71) and (75).;
  Solve (77) to obtain $(\mathbf{q}^{(t)}, \zeta^{(t)})$.;
  Update the linearization points and set $t \leftarrow t+1$.;
until $|\zeta^{(t)} - \zeta^{(t-1)}| \le \varepsilon$;
return $\mathbf{q}^{(t-1)}$.;

terms of x_{jb}^2 and using the relation $x_{jb}^2 \ge 1/y_{jb}$, as established in Proposition 2.

Proposition 2. *Consider* (73) *and let* y_{jb} *be defined as in* (65b). *Define the constants* $U_j \triangleq \sum_{\ell=1}^K |c_{j,\ell}|^2$, $S_j \triangleq \sum_{\ell=1}^K |\tilde{D}_{j,\ell}|^2$, *and* $T_j \triangleq \sum_{\ell=1}^K \Re\{c_{j,\ell}^* \tilde{D}_{j,\ell}\}$. *Then, for any* $\varepsilon_j > 1/S_j$ *(when* $S_j > 0$*), the harvested power admits the global lower bound*

$$P_j^{\mathrm{EH}}(\mathbf{q}) \;\ge\; A_j(\varepsilon_j) + B_j(\varepsilon_j)\,\frac{1}{y_{jb}}, \qquad \forall j \in \mathcal{J}, \tag{74}$$

where $A_j(\varepsilon_j) \triangleq U_j - \varepsilon_j T_j^2$, *and* $B_j(\varepsilon_j) \triangleq S_j - \frac{1}{\varepsilon_j} > 0$.

Proof. See Appendix B. ∎

Using (74), the EH requirement can be enforced in SCA by linearizing $1/y_{jb}$ around the previous iterate. Since $h(y) = 1/y$ is convex for $y > 0$, its first-order Taylor approximation at $y_{jb}^{(t-1)}$ is a global affine underestimator

$$\frac{1}{y_{jb}} \ge \frac{1}{y_{jb}^{(t-1)}} - \frac{1}{\big(y_{jb}^{(t-1)}\big)^2}\big(y_{jb} - y_{jb}^{(t-1)}\big) \triangleq \widetilde{\psi}_j^{(t)}(y_{jb}). \tag{75}$$

Therefore, a sufficient convex inner approximation of the EH constraint (37b) is

$$\sum_{j\in\mathcal{J}} \Big(A_j(\varepsilon_j) + B_j(\varepsilon_j)\,\widetilde{\psi}_j^{(t)}(y_{jb})\Big) \;\ge\; \Omega^{-1}(E_h). \tag{76}$$

Remark 2. *A simple and stable choice of* ε_j *is* $\varepsilon_j = \frac{2}{S_j}$ *(for* $S_j > 0$*), which yields* $B_j(\varepsilon_j) = S_j/2$ *and* $A_j(\varepsilon_j) = U_j - 2T_j^2/S_j$.

At SCA iteration t, the UAV location is updated by solving

$$\max_{\tilde{\boldsymbol{\Xi}}} \quad \zeta \tag{77a}$$
$$\text{s.t.} \quad (70a),\ (72),\ (67), (68), (76), (36d). \tag{77b}$$

Here the optimization variable set is $\tilde{\boldsymbol{\Xi}} \triangleq \{\mathbf{q}, \zeta, \tilde{\boldsymbol{\rho}}, \mathbf{y}_b, \mathbf{y}_u\}$, where $\tilde{\boldsymbol{\rho}} = \{\tilde{\rho}_k\}_{k\in\mathcal{K}}$, $\mathbf{y}_b = \{y_{kb}\}_{k=1}^K$, and $\mathbf{y}_u = \{y_{jb}\}_{j=1}^J$. The proposed method for UAV location optimization is summarized in Algorithm 3. Problem (77) is convex and can be efficiently solved using standard convex optimization solvers (e.g., CVX). The linearization points $\{\tilde{\rho}_k^{(t)}, y_{kb}^{(t)}, y_{jb}^{(t)}\}$ are updated iteratively after each solve until convergence.

By alternately optimizing the power allocation, RIS reflection phases, and UAV location, the benchmark problem in (37) can be efficiently solved via a BCD framework, where each block is handled using the SCA-based procedures developed in Sections III-A, III-B, and III-C. The complete procedure is summarized in Algorithm 4.

Algorithm 4: BCD–SCA Benchmark for WCSEE

Input: Init. $\tilde{\mathbf{p}}^{(0)} \succeq 0$, $\mathbf{s}^{(0)}$, $\mathbf{q}^{(0)} \in \mathcal{L}$; ϵ_{out}, $I_{\max}$
Output: $\tilde{\mathbf{p}}^\star$, $\mathbf{s}^\star$, $\mathbf{q}^\star$
$i \leftarrow 0$, $\eta^{(0)} \leftarrow 0$.
repeat
  Update $\mathbf{H}_c$ and compute ZF directions $\{\hat{\mathbf{p}}_k^{(i)}\}$.
  $\tilde{\mathbf{p}}^{(i+1)} \leftarrow$ Alg. 1.
  $\mathbf{s}^{(i+1)} \leftarrow$ Alg. 2.
  $\mathbf{q}^{(i+1)} \leftarrow$ Alg. 3.
  Compute $R^{\mathrm{sec},(\mathrm{i}+1)}$ and update $\eta^{(i+1)} = \frac{R^{\mathrm{sec},(\mathrm{i}+1)}}{\varrho \sum_k \tilde{p}_k^{(i+1)} + P_0}$.
  $i \leftarrow i+1$.
until $|\eta^{(i)} - \eta^{(i-1)}| \le \epsilon_{\mathrm{out}}$ ***or*** $i = I_{\max}$
return $\tilde{\mathbf{p}}^{(i)}, \mathbf{s}^{(i)}, \mathbf{q}^{(i)}$.

IV. SAC-Based Joint ZF Power Allocation, RIS Phase, and Location Optimization

Due to the nonconvexity of the WCSEE maximization problem in Section II-D, conventional AO and SCA-based methods may suffer from poor scalability and convergence to local optima. To address these challenges, we adopt a SAC framework that learns control policies through interaction with the wireless environment. SAC effectively handles high-dimensional continuous action spaces, ensures stable learning via entropy regularization, and accommodates discrete RIS phase constraints through continuous relaxation and projection. In the following, the problem is first reformulated as a Markov decision process (MDP), after which the SAC-based solution is developed.

A. MDP Model

To enable a learning-based solution, the original optimization problem is reformulated as a sequential decision-making process. The joint design variables, including UAV location, RIS phase shifts, and transmit power, are optimized through interaction with the environment. At each time step, the agent observes the current network configuration (e.g., channel conditions and UAV position) and updates the control variables to improve system performance. This naturally leads to an MDP formulation, where the system evolution depends on the current state and action.

The problem is thus modeled as an MDP characterized by the tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R} \rangle$, where the state represents the instantaneous channel conditions and spatial deployment of the UAV-mounted RIS, the action corresponds to the joint control of UAV location, RIS phase shifts, and transmit power, and the reward reflects the achieved WCSEE performance. At each step t, the agent observes $s_t \in \mathcal{S}$, selects an action $a_t \in \mathcal{A}$ according to a policy $\pi(a_t|s_t)$, and receives a reward $r_t = \mathcal{R}(s_t, a_t)$. The components are defined as follows.

- **Action:** The action includes ZF power allocation, RIS phase shifts, and UAV horizontal position:

$$a_t = \big[\{\tilde{p}_k^{(t)}\}_{k=1}^K, \{\theta_m^{(t)}\}_{m=1}^M, \mathbf{q}^{(t)}\big]. \tag{78}$$

The SAC agent outputs a continuous vector in $[-1,1]^{K+M+2}$ via a *tanh* layer, which is mapped to feasible variables. Power is first scaled to $[0,1]$ and normalized to satisfy (36b):

$$\tilde{p}_k^{(t)} = P_{\max}\frac{\hat{p}_k^{(t)}}{\sum_{\ell=1}^K \hat{p}_\ell^{(t)}}, \quad \forall k \in \mathcal{K}, \tag{79}$$

with $\sum_{\ell=1}^K \hat{p}_\ell^{(t)} > 0$. The RIS phases are mapped to $[0, 2\pi)$ and quantized to $\mathcal{F}$, while $\mathbf{q}^{(t)}$ is projected onto $\mathcal{L}$.

- **State:** The state captures channel and geometric information:

$$s_t = \big[\mathfrak{Re}\{\mathbf{H}_{\mathrm{c}}\}, \mathfrak{Im}\{\mathbf{H}_{\mathrm{c}}\}, \mathfrak{Re}\{\hat{\mathbf{u}}_j\}_{j=1}^J, \mathfrak{Im}\{\hat{\mathbf{u}}_j\}_{j=1}^J, \mathbf{q}^{(t-1)}\big]. \tag{80}$$

Here, $\mathbf{H}_{\mathrm{c}}$ and $\{\hat{\mathbf{u}}_j\}_{j=1}^J$ denote the current effective cascaded channels corresponding to the RIS phase configuration determined by the previous action a_{t-1}, while $\mathbf{q}^{(t-1)}$ denotes the UAV location resulting from the previous action. The user and UEHR locations remain fixed within each episode and vary across episodes.

- **Reward:** Given a_t, the ZF precoder is constructed and the worst-case eavesdropping SINR $\gamma_{E,k}^{(t)}$ is computed via (27). The total power is $P_{\mathrm{tot}}^{(t)} = \varrho \sum_{k=1}^K \tilde{p}_k^{(t)} + P_0$. The reward is

$$r_t = \begin{cases} \dfrac{R^{\mathrm{sec},(t)}}{P_{\mathrm{tot}}^{(t)}}, & \text{if (33) holds,} \\ 0, & \text{otherwise,} \end{cases} \tag{81}$$

where $R^{\mathrm{sec},(t)}$ is defined in (35). This feasibility-based reward design ensures that only constraint-satisfying actions contribute to the learning process, preventing the agent from pursuing a trivial policy that achieves high secrecy by reducing or cutting off energy transfer to the UEHRs. Consequently, it guides the SAC agent toward policies that satisfy the mandatory EH operational constraint while maximizing the WCSEE, consistent with constraint-aware reward shaping widely adopted in the literature (see, e.g., [29], [31]).

B. Fundamentals of the SAC Algorithm

Unlike conventional DRL algorithms that only maximize the expected cumulative reward, the SAC algorithm augments the objective with an entropy term to encourage policy exploration and improve training stability. The optimal policy $\pi^\star$ is obtained by solving the maximum-entropy reinforcement learning (RL) problem [33]

$$\pi^\star = \arg\max_{\pi} \mathbb{E}_{(s_t,a_t)\sim\rho_\pi}\left[\sum_{t=0}^{\infty}\big(r_t + \beta\mathcal{H}(\pi(\cdot|s_t))\big)\right], \tag{82}$$

where ρ_π denotes the state–action visitation distribution under policy π, $\mathcal{H}(\pi(\cdot|s_t)) \triangleq -\mathbb{E}_{a_t\sim\pi(\cdot|s_t)}[\log\pi(a_t|s_t)]$ denotes the policy entropy, and $\beta > 0$ is the temperature parameter that balances exploitation and exploration.

In SAC, two soft Q-functions $Q_{\phi_1}(s,a)$ and $Q_{\phi_2}(s,a)$, parameterized by neural networks with parameters ϕ_1 and ϕ_2, are employed to approximate the soft state–action value:

$$Q^\pi(s_t,a_t) = \mathbb{E}\big[r_t + \gamma\, V^\pi(s_{t+1})\big], \tag{83}$$

where $\gamma \in [0,1]$ is the discount factor weighting future and instant rewards and the soft value function is given by

$$V^\pi(s_{t+1}) = \mathbb{E}_{a_{t+1}\sim\pi}[Q^\pi(s_{t+1},a_{t+1}) - \beta\log\pi(a_{t+1}|s_{t+1})]\,. \tag{84}$$

In practice, the soft value function is not explicitly parameterized and is instead evaluated implicitly using the current policy and target Q-networks, in accordance with the modern SAC formulation. Given a replay buffer $\mathcal{D}$ containing transition tuples (s_t, a_t, r_t, s_{t+1}), the critic networks are trained by minimizing the following mean squared error (MSE) soft Bellman loss

$$J_Q(\phi_i) = \mathbb{E}_{(s_t,a_t,r_t,s_{t+1})\sim\mathcal{D}}\left[\frac{1}{2}\Big(Q_{\phi_i}(s_t,a_t) - y_t\Big)^2\right], \tag{85}$$

where $i \in \{1,2\}$ and

$$y_t = r_t + \gamma\,\mathbb{E}_{a_{t+1}\sim\pi(\cdot|s_{t+1})}\left[\min_{j=1,2} Q_{\bar{\phi}_j}(s_{t+1},a_{t+1}) - \beta\log\pi(a_{t+1}|s_{t+1})\right], \tag{86}$$

with $\bar{\phi}_j$ denoting the parameters of the corresponding target Q-networks.

The actor (policy) network $\pi_\psi(a|s)$, parameterized by ψ, is updated by minimizing

$$J_\pi(\psi) = \mathbb{E}_{s_t\sim\mathcal{D}}\Big[\mathbb{E}_{a_t\sim\pi_\psi}\big(\beta\log\pi_\psi(a_t|s_t) - \min_{i=1,2} Q_{\phi_i}(s_t,a_t)\big)\Big]. \tag{87}$$

The temperature β is tuned automatically by minimizing [34]

$$J(\beta) = \mathbb{E}_{(s_t,a_t)\sim\mathcal{D}}\Big[-\beta\big(\log\pi_\psi(a_t|s_t) + \bar{\mathcal{H}}\big)\Big], \tag{88}$$

where $\bar{\mathcal{H}}$ is a target entropy level, typically set as a negative constant proportional to the action dimension.

Beyond this practical perspective, SAC admits a useful interpretation based on probabilistic inference: the policy improvement step corresponds to projecting the policy distribution toward an energy–based target distribution. Formally, the optimal policy solves

$$J_\pi(\psi) = \mathbb{E}_{s_t\sim\mathcal{D}}\left[D_{\mathrm{KL}}\left(\pi_\psi(\cdot|s_t)\,\middle\|\,\frac{\exp(Q_\phi(s_t,\cdot)/\beta)}{Z_\phi(s_t)}\right)\right], \tag{89}$$

where $D_{\mathrm{KL}}(\cdot\|\cdot)$ denotes the Kullback–Leibler divergence and $Z_\theta(s_t)$ is a normalizing partition function. This viewpoint reveals that SAC implicitly projects the policy toward a Boltzmann distribution induced by the critic values, which explains its stable convergence and robust exploration behaviour.

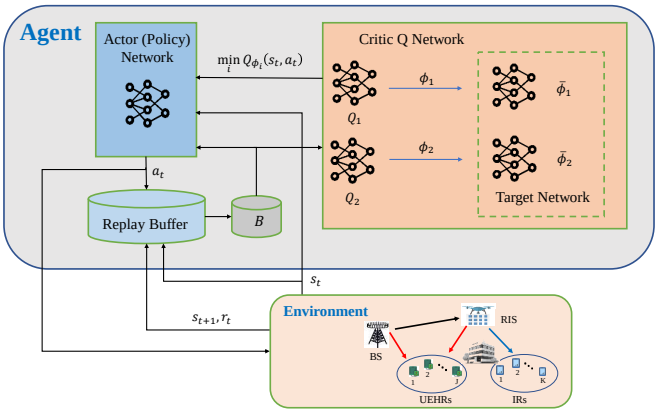


Fig. 2. SAC framework for joint power allocation, RIS phase optimization, and UAV positioning with twin critics and target Q-networks.

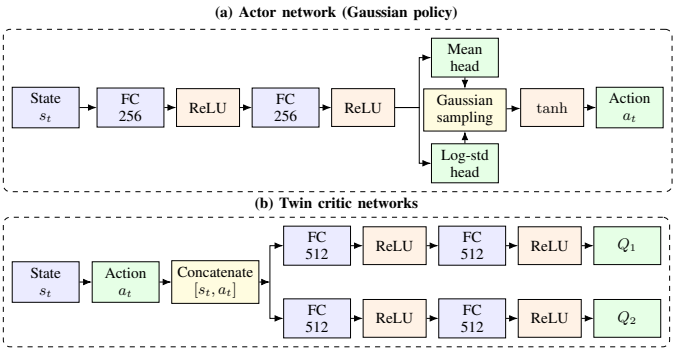


Fig. 3. Structure of the proposed SAC networks: (a) Gaussian actor network and (b) twin Q-critic networks.

C. SAC-Based Joint Optimization Framework

Unlike deterministic actor–critic baselines, SAC updates the policy via an entropy-regularized objective without a target actor network. Fig. 2 depicts the interaction between the SAC agent and the wireless environment. Moreover, Fig. 3 illustrates the architecture of the proposed SAC networks, where the actor employs fully connected (FC) layers with rectified linear unit (ReLU) activations to output the mean and log-standard deviation of a Gaussian policy, followed by Gaussian sampling and $\tanh$ squashing, while each critic takes the concatenated state–action pair as input and outputs a scalar Q-value. The training protocol follows an off-policy paradigm with experience replay and mini-batch updates, where the actor, twin critics, and entropy temperature are iteratively optimized using sampled transitions. At each step, the agent observes the state s_t and samples a continuous action $a_t \sim \pi_\psi(\cdot|s_t)$ that jointly controls the power allocation, RIS phase-shift vector, and UAV horizontal position. The critic employs twin Q-networks Q_{ϕ_1} and Q_{ϕ_2}, and the minimum Q-value is used to mitigate overestimation bias. Target critic networks $\{Q_{\bar{\phi}_1}, Q_{\bar{\phi}_2}\}$ are updated via Polyak averaging to stabilize Bellman backups, while the entropy temperature β is automatically tuned to match a target entropy level. The environment maps actions to feasible variables via power normalization, RIS phase quantization, and location projection, and returns the reward based on the hard EH constraint and the resulting WCSEE. Transitions are stored in a replay buffer and mini-batches are sampled to update the actor, critics, and β as summarized in Algorithm 5. After training, π_ψ enables real-time joint control under different channel realizations and network geometries.

D. Computational Complexity

The computational complexity of the proposed SAC-based framework consists of two components: i) environment interaction and reward computation, and ii) SAC network updates.

At each decision step, mapping the action to $\tilde{\mathbf{p}}^{(t)}$, $\boldsymbol{\theta}^{(t)}$, and $\mathbf{q}^{(t)}$ requires $O(K+M)$ operations. Constructing the cascaded channels for K IRs and J UEHRs incurs $O\big((K+J)N_tM\big)$ complexity. The ZF precoder computation requires $O(N_tK^2+K^3)$, while secrecy-rate and eavesdropping evaluation costs

Algorithm 5: SAC-Based Joint Optimization

Input : $\gamma, \lambda_Q, \lambda_\pi, \lambda_\beta, \bar{\mathcal{H}}, \tau, B$
Output: Policy $\pi_\psi(a|s)$
Initialize replay buffer $\mathcal{D}$, critics Q_{ϕ_1}, Q_{ϕ_2}, targets $Q_{\bar{\phi}_1}, Q_{\bar{\phi}_2}$ with $\bar{\phi}_i \leftarrow \phi_i$, actor π_ψ, and temperature β.
for *each episode* **do**
  Reset environment and observe s_0.
  for *each step* t **do**
    Sample $a_t \sim \pi_\psi(\cdot|s_t)$, map it to $\{\tilde{p}_k^{(t)}\}, \{\theta_m^{(t)}\}, \mathbf{q}^{(t)}$, execute it, and obtain (r_t, s_{t+1}).
    Store (s_t, a_t, r_t, s_{t+1}) in $\mathcal{D}$.
    Sample a mini-batch $\{(s, a, r, s')\}$ from $\mathcal{D}$.
    Update critics: $\phi_i \leftarrow \phi_i - \lambda_Q \nabla_{\phi_i} J_Q(\phi_i),\ i = 1, 2.$
    Update actor: $\psi \leftarrow \psi - \lambda_\pi \nabla_\psi J_\pi(\psi)$.
    Update temperature: $\beta \leftarrow \beta - \lambda_\beta \nabla_\beta J(\beta)$.
    Soft-update targets: $\bar{\phi}_i \leftarrow \tau\phi_i + (1-\tau)\bar{\phi}_i,\ i = 1, 2.$

$O(JKN_t + JK^2)$. Hence, the per-step environment complexity is

$$C_{\text{env}} = O\big((K+J)N_tM + N_tK^2 + K^3 + JKN_t + JK^2\big). \quad (90)$$

The SAC algorithm employs one actor and two critic networks. Let P_π and P_Q denote the number of parameters of the actor and each critic, respectively. For a minibatch of size B, each update requires

$$C_{\text{SAC}} = O\big(B(2P_Q + P_\pi)\big). \quad (91)$$

For an episode of T steps, the overall training complexity is

$$O\big(T\big[C_{\text{env}} + C_{\text{SAC}}\big]\big). \quad (92)$$

Despite the higher offline training cost, the online decision-making complexity is low, requiring only a single actor forward pass, i.e., $O(P_\pi)$, per step.

For the SCA–BCD benchmark, the complexity is governed by the outer BCD iterations and the internal iterations for solving the convex subproblems of power allocation, RIS phase optimization, and UAV location. Using a standard interior-point worst-case approximation, their complexities

TABLE II
HYPERPARAMETERS FOR DRL-BASED ALGORITHMS

Parameter	Description	Value
Loss	Critic loss function	MSE
γ	Discount factor	0.99
τ	Target-network soft update factor	5×10^{-3}
B	Mini-batch size	256
D	Replay buffer size	10^5
T	Training steps per episode	2×10^4
Activation	Hidden-layer activation	ReLU
Hidden layers	Number of hidden layers	2
Hidden units	Units per hidden layer	256
λ_π	Actor learning rate	10^{-4}
λ_Q	Critic learning rate	10^{-3}
λ_β	Temperature learning rate	10^{-3} (SAC)

scale as $O\big(I_{\rm p}(K+2)^3\big)$, $O\big(I_{\rm r}(2M+2JK+2K+1)^3\big)$, and $O\big(I_{\rm q}(2K+J+3)^3\big)$, respectively, where $I_{\rm p}$, $I_{\rm r}$, and $I_{\rm q}$ denote the corresponding iteration numbers. Hence, the overall complexity of Algorithm 4 is

$$C_{\rm SCA} = O\Big(I_{\rm BCD}\big[I_{\rm p}(K+2)^3 + I_{\rm r}(2M+2JK+2K+1)^3 + I_{\rm q}(2K+J+3)^3\big]\Big), \tag{93}$$

where $I_{\rm BCD}$ denotes the number of outer BCD iterations. The cost of channel construction and ZF beamforming is omitted as it is dominated by the convex subproblem solves. Compared with the proposed SAC framework, the SCA–BCD method incurs significantly higher online complexity due to per-realization iterative optimization.

V. NUMERICAL RESULTS

This section evaluates the performance of the proposed SAC-based algorithm and compares it with SCA- and DRL-based benchmarks, including DDPG and TD3, under various system settings. The results demonstrate that the proposed SAC approach consistently achieves superior performance in terms of convergence behavior and average reward, highlighting its effectiveness in addressing the considered optimization problem.

A. Simulation and Benchmark Settings

Unless otherwise specified, simulations are conducted in a 1500×3000 m^2 area with the BS located at $(0,0)$ and a UAV-mounted RIS initially positioned at $(1000, 0, 100)$ m. The UAV moves horizontally within a 50×50 m^2 region at a fixed altitude of $H = 100$ m. The BS employs $N_t = 6$ antennas to serve $K = 4$ legitimate users in the presence of $J = 3$ UEHRs, randomly distributed within circles of radius 500 m. The RIS consists of $M = 10$ elements with discrete phase resolution $L = 8$. Unless stated otherwise, $P_{\max} = 10$ dBm, $\sigma^2 = 10^{-3}$, $P_0 = 1$ W, and $\alpha_b = \alpha_k = \alpha_j = \alpha_{bj} = 2.5$. Both perfect and ICSI are considered; for ICSI, the uncertainty radius is $\nu \in [0.01, 0.1]$, and the minimum harvested energy requirement is $E_h \in [0.005, 0.02]$. For SCA, results are averaged over 100 channel realizations with tolerance $\epsilon = 0.01$. For DRL-based methods, each run consists of $T = 2 \times 10^4$ interaction steps, and results are averaged over multiple random seeds. The main hyperparameters are summarized in Table II.

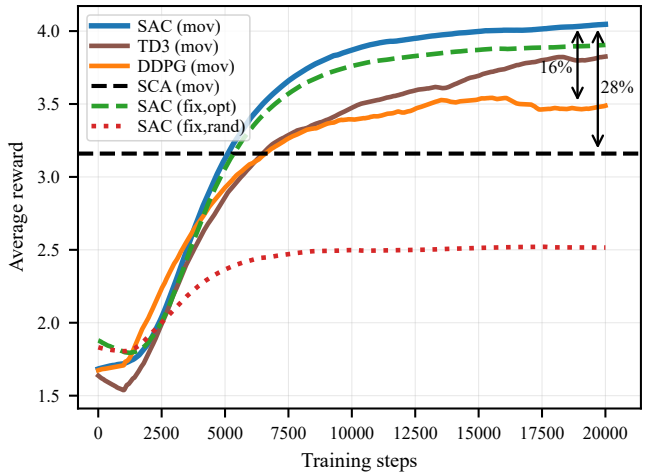


Fig. 4. Convergence behavior under the ideal setting at $P_{\max} = 10$ dBm.

B. Performance in the Ideal Case

We first evaluate the performance of the SCA, DDPG, TD3 and the proposed SAC algorithms under an ideal setting with PCSI and continuous RIS phase shifts, which serves as a performance benchmark under ideal conditions.

As shown in Fig. 4, the proposed SAC-based method with a movable RIS achieves the highest total reward and the fastest convergence. In particular, SAC consistently outperforms TD3, and further improves upon DDPG, achieving about 16% higher reward at convergence. Compared with the SCA benchmark, the performance gain increases to approximately 28%. This improvement is attributed to SAC's ability to jointly optimize RIS phases, UAV location, and transmit power within a unified learning framework, whereas SCA relies on sequential updates that may lead to suboptimal solutions. The figure also includes two additional benchmarks with a fixed RIS: (i) *SAC (fix,opt)* with optimized RIS phases and (ii) *SAC (fix,rand)* with random phases. The results highlight the importance of RIS phase optimization, as random phases significantly degrade performance due to the loss of control over the cascaded BS–RIS–IR/UEHR links.

Fig. 5 depicts SEE versus the transmit power $P_{\max}$ for the SCA, DDPG, and the proposed SAC-based schemes. As expected, increasing $P_{\max}$ improves SEE for all methods due to higher achievable secrecy rates. Nevertheless, the SAC-based approach consistently outperforms both SCA and DDPG across the entire power range. At high transmit power levels, SCA becomes comparable to DDPG; however, both remain inferior to SAC. Fig. 6 illustrates the impact of the number of RIS elements M on SEE under the ideal PCSI setting. SEE increases monotonically with M for all schemes due to the enhanced passive beamforming gain and stronger cascaded links. The proposed SAC-based approach consistently achieves the highest SEE across the entire range of M, highlighting its effectiveness in jointly optimizing RIS phase shifts, UAV location, and transmit power. In contrast, the SCA-based benchmark improves more rapidly at larger M due to increased optimization accuracy under PCSI, while DDPG saturates earlier, indicating limited adaptability to the

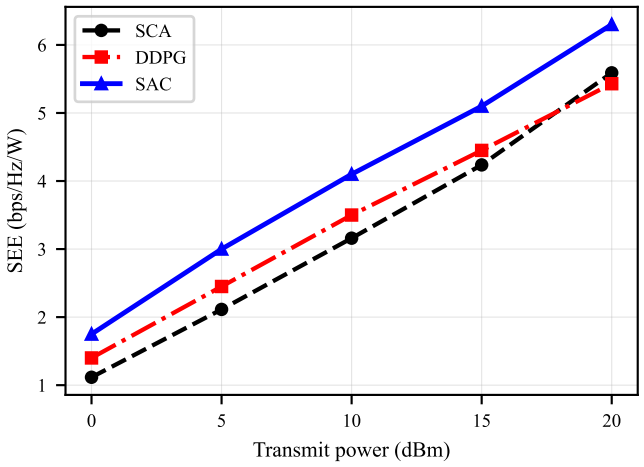


Fig. 5. SEE versus transmit power for SCA, DDPG, and the proposed SAC-based scheme.

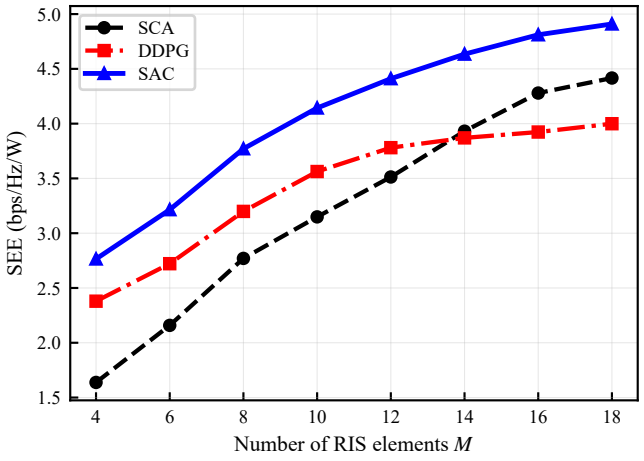


Fig. 6. SEE versus the number of RIS elements M at $P_{\max} = 10$ dBm for different schemes.

enlarged action space. Overall, the results confirm the superior scalability and robustness of SAC with increasing RIS size.

C. Performance in the Non-Ideal Case

We next evaluate a non-ideal setting with ICSI and discrete RIS phases ($L = 8$), where $E_h = 0.02$ and $P_{\max} = 10$ dBm. Fig. 7 compares SAC, TD3, and DDPG under different CSI uncertainty radii $\nu \in \{0.01, 0.1\}$, with PCSI curves and the SCA baseline shown for reference. As ν increases, the achievable reward decreases due to the more conservative robust design. Nevertheless, SAC consistently outperforms both TD3 and DDPG and exhibits faster convergence. Notably, SAC with $\nu = 0.1$ still surpasses DDPG with $\nu = 0.01$, demonstrating strong robustness under severe CSI uncertainty. This gain is attributed to the entropy-regularized policy and twin-critic structure of SAC, which improve exploration and reduce value overestimation under CSI uncertainty.

Fig. 8 compares the learning performance of SAC and DDPG under different mini-batch sizes B with $\nu = 0.01$, $P_{\max} = 10$ dBm, and $E_h = 0.01$. It is observed that

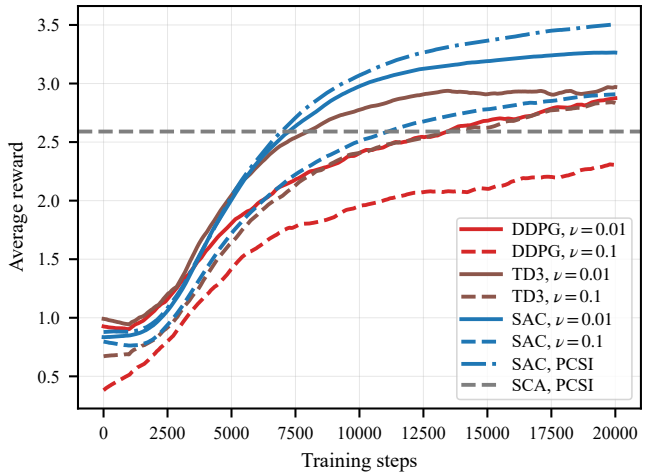


Fig. 7. Convergence comparison under ICSI with discrete RIS phases ($L = 8$) and $E_h = 0.02$ at $P_{\max} = 10$ dBm.

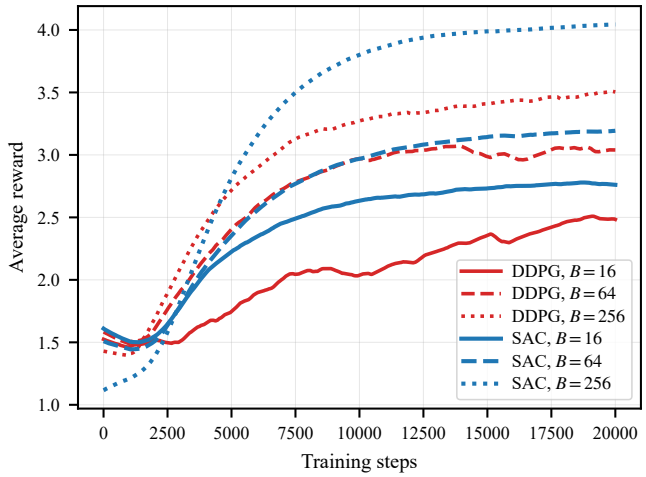


Fig. 8. Convergence comparison of SAC and DDPG under different mini-batch sizes B with $\nu = 0.01$, $P_{\max} = 10$ dBm, and $E_h = 0.01$.

SAC consistently achieves higher average rewards and faster convergence than DDPG across all batch sizes. Moreover, increasing B results in smoother learning curves and improved convergence stability for both methods, as larger batches provide more accurate gradient estimates and reduce update variance. Nevertheless, even with small mini-batch sizes, SAC demonstrates superior robustness and stability, highlighting its stronger sample efficiency and suitability for robust UAV–RIS optimization.

To further evaluate the scalability of the proposed SAC framework, Fig. 9 illustrates the WCSEE versus the number of legitimate users $K \in \{2, 4, 6, 8, 10\}$ for different numbers of UEHRs $J \in \{2, 3, 4, 5\}$. The results are obtained under the ICSI setting with $\nu = 0.01$, $E_h = 0.01$, $N_t = 10$, and $P_{\max} = 30$ dBm. As shown in Fig. 9, the WCSEE decreases monotonically with increasing K for all considered values of J. This degradation is mainly attributed to the limited transmit-power budget and the increasing spatial degrees-of-freedom (DoFs) required to suppress inter-user interference through ZF

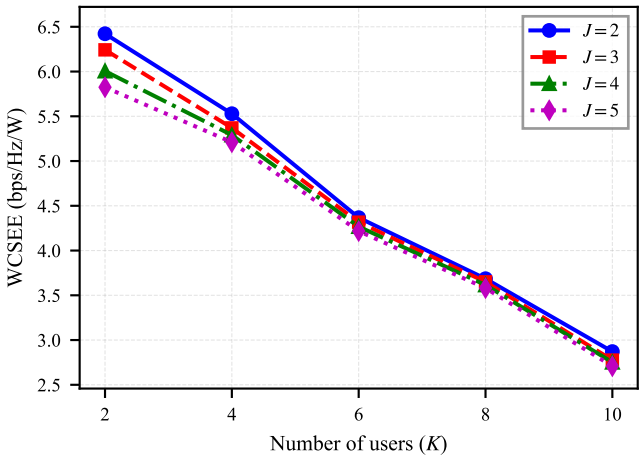


Fig. 9. WCSEE versus the number of legitimate users K for different UEHR counts $J \in \{2,3,4,5\}$ with $\nu = 0.01$, $E_h = 0.01$, and $P_{\max} = 30$ dBm.

precoding. As the number of legitimate users increases, less power and spatial resources are available for improving the individual secrecy rates, resulting in lower overall WCSEE.

For a fixed K, increasing the number of UEHRs also reduces WCSEE. A larger J introduces additional EH constraints and potential information-leakage channels, requiring more conservative resource allocation and RIS configuration to satisfy the worst-case secrecy and EH requirements. Notably, the performance gap among different J values becomes progressively smaller as K increases. For $K = 2$, the WCSEE decreases from approximately 6.42 to 5.85 bps/Hz/W as J increases from 2 to 5, indicating a more pronounced impact of additional UEHRs when sufficient spatial resources are available. In contrast, when $K = 10$, all curves converge to approximately 2.8 bps/Hz/W, suggesting that the performance becomes primarily constrained by the increasing multi-user interference-management burden rather than by the additional UEHRs. Overall, these results demonstrate that the proposed SAC framework maintains stable performance as the network becomes increasingly dense.

VI. Conclusions and Future Work

This paper developed a robust framework for secure and energy-efficient resource allocation in UAV-mounted RIS-assisted multi-user MISO SWIPT networks with distributed UEHRs. The proposed formulation jointly considers the UAV hovering position, BS transmit power allocation, discrete RIS phase shifts, nonlinear EH, and norm-bounded CSI uncertainty, resulting in a challenging WCSEE maximization problem. A BCD–SCA framework was developed as a tractable benchmark for continuous RIS phase shifts under perfect CSI, while a customized SAC framework was proposed for the general robust setting with discrete phase control and imperfect CSI. Simulation results demonstrated that the proposed SAC approach achieves up to 28% and 16% higher WCSEE than the SCA benchmark and DDPG, respectively, while providing faster convergence and robust performance under CSI uncertainty and different system configurations. The results also showed that SAC maintains stable learning across different mini-batch sizes. Furthermore, the scalability study revealed that the WCSEE decreases with increasing numbers of legitimate users and UEHRs, whereas the performance gap associated with different UEHR densities becomes less pronounced in dense-user regimes. These results highlight the effectiveness of SAC for jointly optimizing the highly coupled continuous and discrete variables in secure UAV–RIS-enabled SWIPT networks. Despite these promising results, several practical aspects remain to be addressed. The current framework considers a hovering UAV and therefore does not explicitly capture dynamic trajectory, mobility, or propulsion-energy constraints. Moreover, the bounded CSI uncertainty model does not account for potentially time-varying or statistically correlated channel-estimation errors, and practical RIS hardware impairments such as mutual coupling are not considered.

Future work will extend the framework to dynamic 3D UAV trajectory optimization with mobility and propulsion-energy constraints, as well as cooperative multi-UAV/multi-RIS systems using multi-agent reinforcement learning. Other promising directions include active and STAR-RIS architectures, more realistic RIS hardware models, data-driven CSI uncertainty, and time-varying user mobility.

Appendix A
Proof of Proposition 1

We outline the derivations of (54a)–(54e) as follows.

The legitimate SINR constraint in (52a) can be expressed (after introducing ρ_k) as

$$\rho_k \le \frac{\left|\mathbf{t}_{kk}^H \mathbf{s}\right|^2}{\sum_{\ell \ne k} \left|\mathbf{t}_{k\ell}^H \mathbf{s}\right|^2 + \sigma^2}, \quad \forall k \in \mathcal{K}. \tag{A.1}$$

Rearranging (A.1) yields the DC form

$$\mathcal{M}_k(\mathbf{s}) - \mathcal{N}_k(\mathbf{s}, \rho_k) \le 0, \tag{A.2}$$

where $\mathcal{M}_k(\mathbf{s}) \triangleq \sum_{\ell \ne k} \left|\mathbf{t}_{k\ell}^H \mathbf{s}\right|^2 + \sigma^2$ and $\mathcal{N}_k(\mathbf{s}, \rho_k) \triangleq \frac{\left|\mathbf{t}_{kk}^H \mathbf{s}\right|^2}{\rho_k}$. The nonconvexity is due to $\mathcal{N}_k(\mathbf{s}, \rho_k)$. Thus, at iteration t, we replace $\mathcal{N}_k(\mathbf{s}, \rho_k)$ by its first-order Taylor approximation $\Psi^{(t)}(\mathbf{s}, \rho_k; \mathbf{t}_{kk})$, which directly gives (54a).

For UEHR j decoding user k, introduce auxiliary variables $\rho_{E,j,k}$ and $\xi_{j,k}$ such that

$$\begin{cases} \left|c_{j,k} + \mathbf{t}_{j,k}^H \mathbf{s}\right|^2 \le \xi_{j,k} \rho_{E,j,k}, \\ \xi_{j,k} \le \sum_{\ell \ne k} \left|c_{j,\ell} + \mathbf{t}_{j,\ell}^H \mathbf{s}\right|^2 + \sigma^2. \end{cases} \tag{A.3}$$

The product $\xi_{j,k}\rho_{E,j,k}$ in (A.3) is bilinear and nonconvex; we apply the standard affine lower bound $\Theta^{(t)}(\xi_{j,k}, \rho_{E,j,k})$, which yields (54b). Moreover, each quadratic term $\left|c_{j,\ell} + \mathbf{t}_{j,\ell}^H \mathbf{s}\right|^2$ in the second inequality of (A.3) is convex; thus, its first-order Taylor expansion provides a global lower bound. Accordingly, we construct a convex surrogate constraint via inner approximation, leading to (54c).

Constraint (54d) follows by rewriting $f_{E,k} \ge \log_2(1 + \rho_{E,j,k})$ into $2^{f_{E,k}} \ge 1 + \rho_{E,j,k}$ and then applying the first-order Taylor approximation $\Gamma^{(t)}(f_{E,k})$ of $2^{f_{E,k}}$. Finally, (54e) is the epigraph form of the secrecy constraint. This completes the proof.

APPENDIX B
PROOF OF PROPOSITION 2

For fixed $\{c_{j,\ell}, \tilde{D}_{j,\ell}\}$ and real $x_{jb} \geq 0$, expanding (73) yields the convex quadratic

$$P_j^{\mathrm{EH}}(x_{jb}) = U_j + S_j x_{jb}^2 + 2T_j x_{jb}, \tag{B.1}$$

where U_j, S_j, and T_j are defined in Proposition 2. Next, we lower bound the linear term $2T_j x_{jb}$.

Lemma 1. *For any $x_{jb} \in \mathbb{R}$ and any $\varepsilon_j > 0$,*

$$2T_j x_{jb} \;\geq\; -\varepsilon_j T_j^2 - \frac{1}{\varepsilon_j} x_{jb}^2. \tag{B.2}$$

Proof. Let $a = \sqrt{\varepsilon_j}\, T_j$ and $b = x_{jb}/\sqrt{\varepsilon_j}$. Since $(a+b)^2 \geq 0$, we have $2ab \geq -a^2 - b^2$, i.e., $2T_j x_{jb} \geq -\varepsilon_j T_j^2 - x_{jb}^2/\varepsilon_j$, which proves (B.2). ∎

Applying Lemma 1 to (B.1) yields, for any $\varepsilon_j > 0$,

$$P_j^{\mathrm{EH}} \geq \big(U_j - \varepsilon_j T_j^2\big) + \Big(S_j - \frac{1}{\varepsilon_j}\Big) x_{jb}^2 = A_j(\varepsilon_j) + B_j(\varepsilon_j) x_{jb}^2. \tag{B.3}$$

Choosing $\varepsilon_j > 1/S_j$ (for $S_j > 0$) ensures $B_j(\varepsilon_j) > 0$. Finally, from (65b) and $x_{jb} = \big(d_j^2(\mathbf{q}) d_b^2(\mathbf{q})\big)^{-\alpha/4}$, we obtain

$$x_{jb}^2 = \big(d_j^2(\mathbf{q}) d_b^2(\mathbf{q})\big)^{-\alpha/2} \geq \frac{1}{y_{jb}}. \tag{B.4}$$

Substituting (B.4) into (B.3) yields $P_j^{\mathrm{EH}} \geq A_j(\varepsilon_j) + B_j(\varepsilon_j)\, 1/y_{jb}$, which coincides with (74) and completes the proof.

REFERENCES

[1] R. Kaur *et al.*, "A survey on reconfigurable intelligent surface for physical layer security of next-generation wireless communications," *IEEE Open J. Veh. Technol.*, vol. 5, pp. 172–199, 2024.

[2] C. Pan *et al.*, "An overview of signal processing techniques for RIS/IRS-aided wireless systems," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 5, pp. 883–917, Aug. 2022.

[3] Y. Zeng, Q. Wu, and R. Zhang, "Accessing from the sky: A tutorial on UAV communications for 5G and beyond," *Proc. IEEE*, vol. 107, no. 12, pp. 2327–2375, Dec. 2019.

[4] M. K. Banafaa, I. Shayea, J. Din, M. H. Azmi, A. Alhammadi, and Y. I. Al-sshehri, "A comprehensive survey on 5G-and-beyond networks with UAVs: Applications, emerging technologies, regulatory aspects, research trends and challenges," *IEEE Access*, vol. 12, pp. 7786–7826, 2024.

[5] Q. Wu *et al.*, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, May 2021.

[6] M. Misbah *et al.*, "Phase and 3-D placement optimization for rate enhancement in RIS-assisted UAV networks," *IEEE Wireless Commun. Lett.*, vol. 12, no. 7, pp. 1135–1138, Jul. 2023.

[7] Q. Deng *et al.*, "IRS-assisted cognitive UAV networks: Joint sensing duration, passive beamforming, and 3-D location optimization," *IEEE Internet Things J.*, vol. 11, no. 2, pp. 2767–2782, Jan. 2024.

[8] B. Clerckx *et al.*, "Fundamentals of wireless information and power transfer: From RF energy harvester models to signal and system designs," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 1, pp. 4–33, Jan. 2019.

[9] I. Krikidis *et al.*, "Simultaneous wireless information and power transfer in modern communication systems," *IEEE Commun. Mag.*, vol. 52, no. 11, pp. 104–110, Nov. 2014.

[10] B. Zhang *et al.*, "Performance analysis of RIS-assisted wireless communications with energy harvesting," *IEEE Trans. Veh. Technol.*, vol. 72, no. 1, pp. 1325–1330, Jan. 2023.

[11] M. Mohammadi, H. Q. Ngo, and M. Matthaiou, "Phase-shift and transmit power optimization for RIS-aided massive MIMO SWIPT IoT networks," *IEEE Trans. Commun.*, vol. 73, no. 1, pp. 631–647, Jan. 2025.

[12] 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, May 2013.

[13] H. V. Poor and R. F. Schaefer, "Wireless physical layer security," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 1, pp. 19–26, Jan. 2017.

[14] J. Ouyang *et al.*, "Secrecy energy efficiency in full-duplex AF relay systems with untrusted energy harvesters," *IEEE Commun. Lett.*, vol. 25, no. 11, pp. 3493–3497, Nov. 2021.

[15] Y. Lu *et al.*, "Global energy efficiency in secure MISO SWIPT systems with non-linear power-splitting EH model," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 1, pp. 216–232, Jan. 2019.

[16] H. R. Hashempour *et al.*, "Secure SWIPT in the multiuser STAR-RIS aided MISO rate splitting downlink," *IEEE Trans. Veh. Technol.*, vol. 73, no. 9, pp. 13 466–13 481, Sep. 2024.

[17] C. Luo, W. Jiang, J. Nie, D. Niyato, and C. Pan, "Robust beamforming design for the STAR-RIS-assisted secure SWIPT system," *IEEE Trans. Veh. Technol.*, vol. 73, no. 12, pp. 18 605–18 619, Dec. 2024.

[18] H. Fu *et al.*, "Robust secure beamforming design for two-user downlink MISO rate-splitting systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8351–8365, Dec. 2020.

[19] E. Boshkovska *et al.*, "Robust resource allocation for MIMO wireless powered communication networks based on a non-linear EH model," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 1984–1999, May 2017.

[20] 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, Mar. 2020.

[21] S. K. Singh *et al.*, "NOMA enhanced hybrid RIS-UAV-assisted full-duplex communication system with imperfect SIC and CSI," *IEEE Trans. Commun.*, vol. 70, no. 11, pp. 7609–7627, Nov. 2022.

[22] Y. Su *et al.*, "Joint location and beamforming optimization for STAR-RIS aided NOMA-UAV networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 8, pp. 11 023–11 028, Aug. 2023.

[23] S. Li *et al.*, "Reconfigurable intelligent surface assisted UAV communication: Joint trajectory design and passive beamforming," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 716–720, May 2020.

[24] E. A. Gharavol *et al.*, "Robust downlink beamforming in multiuser MISO cognitive radio networks with imperfect channel-state information," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 2852–2860, Jul. 2010.

[25] H. R. Hashempour *et al.*, "Power efficient cooperative communication within IIoT subnetworks: Relay or RIS?" *IEEE Internet Things J.*, vol. 11, no. 21, pp. 35 260–35 274, Nov. 2024.

[26] V. Kumar *et al.*, "A novel SCA-based method for beamforming optimization in IRS/RIS-assisted MU-MISO downlink," *IEEE Wireless Commun. Lett.*, vol. 12, no. 2, pp. 297–301, Feb. 2023.

[27] L. D. Nguyen *et al.*, "Energy efficiency in cell-free massive MIMO with zero-forcing precoding design," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1871–1874, Aug. 2017.

[28] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, no. 7, pp. 492–498, 1967.

[29] Y. Zhang *et al.*, "Deep reinforcement learning for secrecy energy efficiency maximization in RIS-assisted networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 9, pp. 12 413–12 418, Sep. 2023.

[30] C. Luo *et al.*, "Optimization and DRL-based joint beamforming design for active-RIS enabled cognitive multicast systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 11, pp. 16 234–16 247, Nov. 2024.

[31] Z. Ma *et al.*, "Deep reinforcement learning for energy efficiency maximization in RSMA-IRS-assisted ISAC systems," *IEEE Trans. Veh. Technol.*, vol. 74, no. 11, pp. 18 273–18 278, Nov. 2025.

[32] M. M. Razaq *et al.*, "DRL-based physical-layer security optimization in near-field MIMO systems," *IEEE Internet Things J.*, vol. 12, no. 12, pp. 18 606–18 615, Jun. 2025.

[33] T. Haarnoja *et al.*, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2018, pp. 1861–1870.

[34] ——, "Soft actor-critic algorithms and applications," *arXiv preprint arXiv:1812.05905*, 2018.

[35] B. Zheng, C. You, and R. Zhang, "Fast channel estimation for IRS-assisted OFDM," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 580–584, Mar. 2021.

[36] C. You, B. Zheng, and R. Zhang, "Channel estimation and passive beamforming for intelligent reflecting surface: Discrete phase shift and progressive refinement," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2604–2620, Nov. 2020.

[37] H. Guo and V. K. N. Lau, "Uplink cascaded channel estimation for intelligent reflecting surface assisted multiuser MISO systems," *IEEE Trans. Signal Process.*, vol. 70, pp. 3964–3977, 2022.

[38] E. Boshkovska *et al.*, "Practical non-linear energy harvesting model and resource allocation for SWIPT systems," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2082–2085, Dec. 2015.