applied sciences

MDPI

Article

Two-Layer Non-Alternating Joint Optimization of Trajectory Planning, User Association, and Radio Resource Management for UAV-Assisted Uplink IoT Data Collection: A Weighted Sum Power Minimization Approach

Yang Yu [1,*], Xiaoqing Tang [2] and Guihui Xie [3]

[1] Electronic Information School, Hubei Three Gorges Polytechnic, Yichang 443199, China
[2] School of Artificial Intelligence, Hubei University, Wuhan 430062, China; 2010202120076@whu.edu.cn
[3] School of Automation, China University of Geosciences (Wuhan), Wuhan 430074, China; xieguihui@cug.edu.cn
* Correspondence: yuyang@tgc.edu.cn

Abstract

Unmanned aerial vehicle (UAV)-assisted data collection from battery-constrained Internet of Things (IoT) devices faces a fundamental trade-off between communication reliability and device energy depletion. This paper jointly optimizes UAV trajectory, user association, and radio resource management (RRM) to minimize the uplink weighted sum transmit powers, where each device's weight is dynamically and inversely related to its residual energy, while treating UAV propulsion energy as a feasibility constraint. To circumvent the initialization trap of conventional alternating optimization (AO), we propose a two-layer non-alternating framework. The inner layer solves the per-slot RRM problem analytically via KKT conditions for a fixed UAV position, yielding analytical power allocation and a unique bandwidth solution, while user association is determined by an incremental greedy algorithm. The outer layer formulates trajectory planning (TP) as a Markov decision process (MDP), enabling single-pass trajectory synthesis without cross-layer iteration, thereby inherently avoiding initialization sensitivity. The framework supports the genetic algorithm (GA) and limited depth-first search (DFS) as trajectory solvers, with the deep Q-network (DQN) as a promising future extension, each offering distinct optimality–complexity trade-offs. Simulation results show that the proposed scheme consistently outperforms conventional iterative baselines across various network configurations.

Keywords: unmanned aerial vehicle (UAV); Internet of Things (IoT); trajectory planning (TP); radio resource management (RRM); user association; non-alternating; Markov decision process (MDP); Karush–Kuhn–Tucker (KKT)

Check for updates

Academic Editor: Dimitri Konstantas

Received: 17 August 2026
Revised: 16 September 2026
Accepted: 25 September 2026
Published: 27 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The rapid advancement of 6G wireless networks and the Internet of Things (IoT) has created unprecedented demands for wide-area data collection, emergency communication, and on-demand coverage. Unmanned aerial vehicle (UAV)-assisted communications have emerged as a flexible and cost-effective solution [1], owing to their high mobility, rapid deployment capability, and generally favorable line-of-sight (LoS) channel conditions. Among the diverse UAV-enabled applications, UAV-assisted uplink data collection from ground IoT devices has attracted significant research attention. It serves as a critical enabler for environmental monitoring [2], smart agriculture [3], and disaster response [4]—scenarios where

terrestrial communication infrastructure is either unavailable or temporarily disrupted and where timely data acquisition is essential for decision-making.

A defining characteristic of such scenarios is that IoT devices are typically battery-constrained and deployed in hard-to-reach or hazardous areas, making frequent battery replacement or recharging infeasible [5]. This imposes a fundamental trade-off: higher transmit power improves the communication link quality and achievable data rate but accelerates battery depletion, thereby shortening the operational lifetime of the entire network. Consequently, minimizing the transmit power of ground devices while satisfying their quality-of-service (QoS) requirements—in terms of minimum data rate and transmission reliability—is of paramount importance for sustainable and long-term network operation.

The UAV itself is also energy-limited, with propulsion power consumption dominating its onboard energy budget [6]. Unlike ground devices whose transmit power is the primary energy drain, the UAV consumes significant energy on mobility, which directly constrains the feasible mission duration and coverage range. A practical optimization framework must therefore jointly coordinate user association, radio resource management (RRM), and UAV trajectory planning (TP) to balance the energy expenditure of both ground devices and the UAV, treating the UAV's propulsion energy as a hard feasibility constraint rather than a soft optimization objective. However, achieving this balance is highly nontrivial, as user association, RRM, and TP are deeply intertwined—forming a strongly coupled, mixed-integer non-convex problem that conventional methods struggle to resolve efficiently.

To manage this coupling, the vast majority of existing works adopt alternating optimization (AO) [7,8] frameworks, which iteratively update one set of variables while fixing the others, often employing successive convex approximation (SCA) [9] to handle trajectory-related non-convexity. For instance, Zhang and Ansari [10] proposed a cyclic iterative algorithm that alternately solves 3D drone base station placement and resource allocation subproblems to maximize throughput under backhaul capacity constraints. Liu et al. [11] developed both SCA-based and AO-based algorithms for jointly optimizing UAV trajectory, velocity, and IoT device transmit power in a wireless powered mobile edge learning system, demonstrating that both approaches converge within several iterations. Du et al. [12] applied the block coordinate descent (BCD) method—a classic AO variant—to decouple resource allocation and trajectory optimization, solving each non-convex subproblem via SCA in an iterative manner to minimize mission completion time under UAV energy constraints. Lu et al. [13] employed SCA and BCD to iteratively optimize UAV trajectory and computational resources for maximizing the minimum secure calculation capacity in a dual-UAV mobile edge computing system. Yang et al. [14] jointly utilized bisection search and BCD to alternately solve user association, trajectory, and power subproblems, further employing geometric programming to handle non-convexity in multi-UAV post-disaster data dissemination. Additional AO-based frameworks have been applied across diverse UAV communication scenarios, including joint 3D positioning and channel allocation [15], formation and resource optimization for integrated sensing and communication [16], dual-UAV covert communications [17], secure reconfigurable intelligent surface (RIS)- and non-orthogonal multiple access (NOMA)-assisted multi-UAV systems [18], integrated sensing, communication, and computation in clustered UAV networks [19], and Age of Information (AoI)-minimal TP with wireless power transfer [20].

Despite their popularity, AO methods suffer from two intertwined limitations. First, the nested iterative structure—where AO outer loops alternate between trajectory and resource optimization, while SCA inner loops handle the non-convexity of each subproblem—incurs substantial computational overhead, which grows rapidly with the number of devices and time slots. Second, and more critically, in mixed-integer non-convex problems such as joint trajectory and resource optimization, AO is not guaranteed to converge to a

stationary point and can stagnate at a non-stationary point [21–23]; the solution quality is also highly sensitive to the initial trajectory. This "initialization trap" is particularly severe in UAV TP because the trajectory search space is high-dimensional and riddled with poor local optima; a suboptimal initial trajectory can lock all subsequent AO updates into an inferior region of the solution space from which escape is virtually impossible. While multiple random initializations can mitigate this sensitivity to some extent, they aggravate the already high computational burden and offer no convergence guarantee.

A significant step toward overcoming the initialization trap was recently made by Lyu et al. [24], who proposed a non-iterative framework for downlink proportional fairness (PF) maximization. By reformulating TP as a Markov decision process (MDP) and solving the RRM subproblem via Karush–Kuhn–Tucker (KKT) conditions, their approach eliminates the alternating update loop and does not require an initial trajectory guess. However, that framework is specifically designed for a downlink scenario where the objective is to ensure fair rate distribution among users with static weighting. In uplink IoT data collection, by contrast, the primary concern shifts from throughput fairness to the energy depletion of battery-constrained devices: an effective policy should protect low-residual-energy devices by lowering their transmit power, which introduces a temporal coupling absent in static PF schemes. Current resource allocation decisions affect future device energy states, which in turn should influence future scheduling priorities.

The importance of energy-aware scheduling in uplink IoT data collection has been recognized in the literature. Treating all battery-constrained devices equally in terms of transmit power minimization can lead to premature depletion of the most energy-critical nodes, creating coverage holes and reducing overall network lifetime. To address this, several studies have introduced residual-energy-based prioritization mechanisms—such as weighted sum-power minimization and energy-aware device selection—to balance energy consumption across the network [25,26]. These works demonstrate that the weighted sum power formulation enables computationally efficient energy-aware resource allocation by avoiding the iterative complexity of fractional programming or the nonlinearity of energy efficiency (EE)/weighted sum energy efficiency (WSEE) objectives, while per-device weights provide a direct mechanism for managing power budgets according to device priorities. However, these works assume static network topologies or fixed infrastructure and do not involve UAV trajectory optimization. More recently, a few studies have begun to incorporate energy-aware objectives into UAV trajectory optimization, for instance, by maximizing the minimum residual energy of IoT devices [27] or by introducing energy-state-dependent weights in mobile edge computing (MEC) scenarios [28]. While these works represent an important step forward, they typically operate within conventional AO frameworks that solve each time slot independently and thus do not natively capture the temporal coupling introduced by the dynamic evolution of ground devices' residual energy states. This coupling creates a sequential dependency that most existing methods, whether AO-based or non-iterative, are not designed to handle—AO-based approaches would require repeated full re-optimizations to reflect updated energy states, while non-iterative frameworks such as [24] treat weights as fixed throughout the mission.

To address these challenges, we propose a two-layer non-AO framework tailored for uplink IoT data collection, where the problem is formulated as minimizing the weighted sum of device transmit powers. Crucially, the weight assigned to each device is dynamically and inversely related to its residual battery energy, thereby protecting energy-starved devices to prolong network lifetime. In the inner layer, for any fixed UAV position, the per-slot RRM subproblem is solved analytically via KKT conditions, yielding an analytical power spectral density (PSD) allocation and a unique bandwidth solution obtainable through bisection. In the outer layer, we reformulate the multi-slot TP as an MDP whose state

includes the UAV position, the remaining UAV battery, the residual energies of all devices, as well as their active window and service status indicators. The step reward is defined as the negative of the inner-layer optimal weighted transmit power, and the evolution of device residual energies is embedded directly in the state transition dynamics. In this way, as the residual energies evolve, the weights for subsequent time slots update automatically, natively capturing the resulting temporal coupling without any alternating iterations. The full trajectory is thus generated in a single forward pass, entirely circumventing both the initialization sensitivity and the nested iterative overhead of conventional AO.

The main contributions of this paper are summarized as follows.

1. We propose a two-layer non-AO framework for uplink UAV-assisted IoT data collection that jointly optimizes user association, RRM (including bandwidth allocation and power control), and TP. Unlike existing non-iterative frameworks designed for downlink fairness, our framework integrates dynamically energy-aware weights that are inversely related to the residual energy of each device and evolve across time slots. This design natively introduces and handles the resulting temporal coupling, representing a departure from static-weighting schemes in prior works.
2. For the inner-layer RRM subproblem, we obtain a KKT-based analytical solution that yields an analytical PSD allocation and a unique bandwidth assignment determined via bisection. By eliminating inner iterative loops, this solution attains low computational complexity suitable for real-time deployment.
3. For the outer-layer TP, we formulate the problem as an MDP whose state vector explicitly carries the residual energies of all devices. This formulation naturally captures the time coupling induced by dynamic weights, and it supports multiple solvers—including genetic algorithm (GA), limited depth-first search (DFS), and deep Q-network (DQN) (the latter left for future work)—providing flexible trade-offs between optimality and computational cost. The entire trajectory is obtained in a single forward pass without alternating updates.
4. Extensive simulations demonstrate that the proposed framework consistently outperforms conventional AO baselines across various network configurations. The results also verify that the framework effectively circumvents initialization sensitivity, delivering robust performance regardless of the initial trajectory choice.

The remainder of this paper is organized as follows. Section 2 presents the system model and problem formulation. Section 3 details the proposed two-layer non-alternating framework, including the KKT-based inner-layer RRM solution and the MDP-based outer-layer TP, and the overall algorithm. Section 4 provides numerical simulation results and comparisons, and Section 5 concludes the paper.

2. System Model and Problem Formulation

To formalize the system under consideration, we now specify the network scenario, the communication and energy models, and then formulate the joint optimization problem.

2.1. Network Scenario

We consider a rotary-wing UAV that periodically patrols a target area to collect uplink sensing data (e.g., toxic gas concentration) from I ground IoT devices (users) [29]. The mission spans a finite horizon discretized into T time slots, as illustrated in Figure 1.

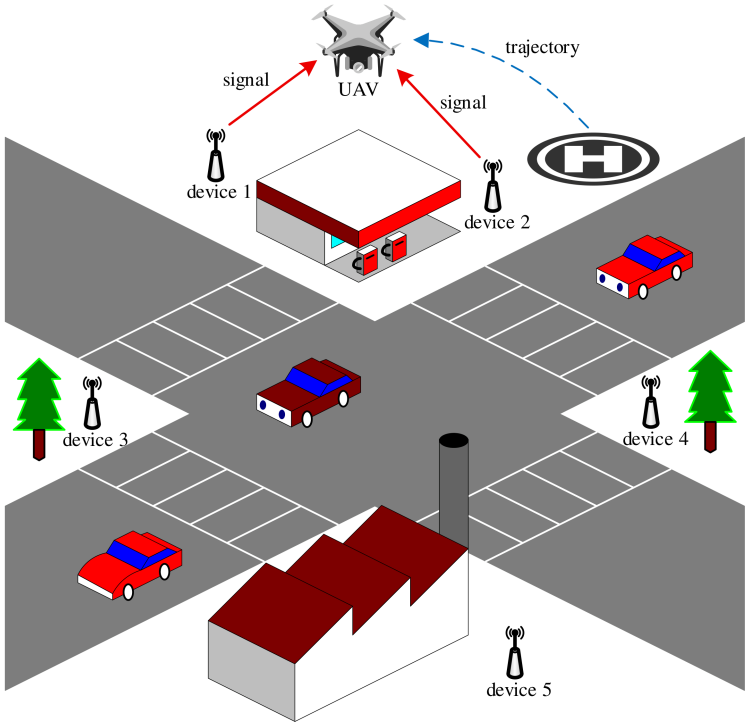

Figure 1. System model with $I = 5$ ground devices.

The ground devices are indexed by the set $\mathcal{I} = \{1, 2, \ldots, I\}$. Each device has an active uplink period and a minimum rate constraint to guarantee its QoS requirement. We assume that the flight map and service timeline are discretized. The time slots are indexed by $\mathcal{T} = \{1, 2, \ldots, T\}$, each with a duration of ΔT s. The UAV covers a square-shaped ground area of size $w \times w$ m^2, and the flight map is represented by a set $\mathcal{Q}$ of three-dimensional grids discretized by an interval ΔQ m as follows:

$$\mathcal{Q} = \{\Delta Q[l, m, n]^{\mathrm{T}} | [l, m, n]^{\mathrm{T}} \in \mathbb{N}_0^3, 0 \le l\Delta Q \le w, 0 \le m\Delta Q \le w, h_{\min} \le n\Delta Q \le h_{\max}\}, \quad (1)$$

where l, m, n are the grid indices along the x, y, h axes; $h_{\min}$ and $h_{\max}$ are the minimum and maximum altitudes of the UAV, respectively.

We adopt a waypoint-based discrete navigation paradigm [30,31]—a widely adopted standard across various aerial vehicles, from fixed-wing aircraft to rotorcraft—where the UAV resides on a grid point of $\mathcal{Q}$ at the beginning of each time slot. Let $\mathbf{q}^{(t)} = [x^{(t)}, y^{(t)}, h^{(t)}]^{\mathrm{T}} \in \mathcal{Q}$ denote the UAV position at the beginning of the t-th time slot, with $\mathbf{q}^{(0)} \in \mathcal{Q}$ being the initial depot position.

The UAV has a maximum flight speed v, which yields the following mobility constraint:

$$\|\mathbf{q}^{(t)} - \mathbf{q}^{(t-1)}\|_2 \le v\Delta T, \forall t \in \mathcal{T}. \quad (2)$$

Accordingly, all reachable positions at slot t form the set $S(\mathbf{q}^{(t-1)})$, and we have $\mathbf{q}^{(t)} \in S(\mathbf{q}^{(t-1)})$, where $S(\mathbf{q}^{(t-1)}) = \left\{\mathbf{q} \in \mathcal{Q} \middle| \|\mathbf{q} - \mathbf{q}^{(t-1)}\|_2 \le v\Delta T\right\}$.

Since the UAV operates in a periodic patrol mode, it must return to its initial position, i.e., the depot, after completing the data collection mission. This is enforced by the following constraint:

$$\mathbf{q}^{(T)} = \mathbf{q}^{(0)}. \quad (3)$$

On the ground, the i-th device is located at $\mathbf{q}_i = [x_i, y_i, 0]^{\mathrm{T}}$ and remains stationary throughout the mission. Each device periodically generates sensing data that must be uploaded to the UAV within a data validity window $[s_i, e_i)$, where s_i and e_i denote the earliest and latest time slots during which the data remains valid. Data transmitted outside

this window is considered expired and is discarded. To prolong device lifetime, each device only activates its radio within this window. We define a binary indicator $d_i^{(t)}$ that specifies whether device i is active at time slot t:

$$d_i^{(t)} = \begin{cases} 1, & \text{if } s_i \le t < e_i, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

For each time slot t, the UAV determines a set of devices to serve via uplink transmission. We define the user association variables as

$$\alpha_i^{(t)} = \begin{cases} 1, & \text{if device } i \text{ transmits to the UAV at time slot } t, \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

where $\mathcal{U}^{(t)} = \left\{ i \in \mathcal{I} \middle| \alpha_i^{(t)} = 1 \right\}$ denotes the set of associated devices at time slot t.

We denote $\beta_i^{(t)}$ as the bandwidth allocated to device i if it is associated with the UAV at time slot t. The total bandwidth allocated to all associated devices must not exceed the available system bandwidth B, i.e.,

$$\sum_{i \in \mathcal{U}^{(t)}} \beta_i^{(t)} \le B, \forall t \in \mathcal{T}. \tag{6}$$

Let $\rho_i^{(t)}$ denote the transmit PSD of device i at time slot t. Each device has a maximum transmit power limit $P_{\max}$, yielding the individual power constraint:

$$0 \le \rho_i^{(t)} \beta_i^{(t)} \le P_{\max}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}. \tag{7}$$

For devices with $\alpha_i^{(t)} = 0$, the corresponding power and bandwidth allocation are defined as $\rho_i^{(t)} \beta_i^{(t)} = 0$ and $\beta_i^{(t)} = 0$.

2.2. Uplink Communication Model

For the uplink transmission between ground devices and the UAV, we adopt the widely validated probabilistic LoS/non-line-of-sight (NLoS) channel model for low-altitude aerial communications [32]. The probability that device i has an LoS link to the UAV at time slot t is given by

$$\mathbb{P}_i^{(t)} = \frac{1}{1 + a_{\text{env}} \cdot \exp\left(-b_{\text{env}} \left(\theta_i^{(t)} - a_{\text{env}} \right) \right)}, \tag{8}$$

and the corresponding NLoS probability is $1 - \mathbb{P}_i^{(t)}$, where

$$\theta_i^{(t)} = \arcsin\left(\frac{h^{(t)}}{\| \mathbf{q}^{(t)} - \mathbf{q}_i \|_2} \right) \tag{9}$$

is the elevation angle between device i and the UAV. Parameters a_{env} and b_{env} are environment-dependent constants with $\theta_i^{(t)}$ measured in degrees [32,33]. The output of (9), which is in radians, must be converted to degrees before being used in (8).

The average pathloss experienced by device i at time slot t is

$$\xi_i^{(t)} = \text{PL}\left(\mathbf{q}^{(t)}, \mathbf{q}_i \right) + \mathbb{P}_i^{(t)} \eta_{\text{LoS}} + \left(1 - \mathbb{P}_i^{(t)} \right) \eta_{\text{NLoS}}, \tag{10}$$

where

$$\text{PL}(\mathbf{q}^{(t)}, \mathbf{q}_i) = 20 \log_{10}\left(\frac{4\pi f \| \mathbf{q}^{(t)} - \mathbf{q}_i \|_2}{c} \right) \tag{11}$$

is the free-space pathloss, f is the carrier frequency, and c is the speed of light. The constants η_{LoS} and η_{NLoS} denote the mean excessive pathloss values for LoS and NLoS links, respectively [32,34].

Given the allocated bandwidth $\beta_i^{(t)}$ and the transmit PSD $\rho_i^{(t)}$, the uplink data rate of device i at time slot t is expressed as

$$R_i^{(t)} = \beta_i^{(t)} \log_2 \left(1 + \frac{\rho_i^{(t)} g_i^{(t)}}{N_0} \right), \tag{12}$$

where $g_i^{(t)} = 10^{-\xi_i^{(t)}/10}$ is the channel gain between device i and the UAV at time slot t, and N_0 is the PSD of the additive white Gaussian noise (AWGN).

To guarantee the QoS, each associated device, when selected for transmission at time slot t, must achieve a minimum data rate r_i. Therefore, we impose the following constraint:

$$R_i^{(t)} \geq \alpha_i^{(t)} r_i, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}. \tag{13}$$

This formulation does not mandate that every device transmits in each time slot; it only specifies that whenever a transmission occurs, it must meet the QoS rate requirement. This is a standard approach for guaranteeing the reliability of individual transmission sessions in time-critical IoT uplink scenarios.

2.3. Energy Model

We now detail the energy consumption models for the ground devices and UAV. As stated in the Introduction, the system involves two distinct power components: the uplink transmit power of these ground devices, and the propulsion power of the UAV. Given that ground devices are battery-constrained and difficult to recharge, we set the primary objective as minimizing the weighted sum of their transmit power, while treating the UAV's propulsion energy as a hard feasibility constraint.

Let $\rho_i^{(t)} \beta_i^{(t)} \Delta T$ denote the energy consumed by device i at time slot t. We define $E_i^{\text{res}}(t)$ as the residual energy of device i at the beginning of time slot t. Its evolution follows the recursion:

$$E_i^{\text{res}}(t+1) = E_i^{\text{res}}(t) - \rho_i^{(t)} \beta_i^{(t)} \Delta T, \tag{14}$$

$$E_i^{\text{res}}(1) = E_i^{\text{init}}, \tag{15}$$

where E_i^{init} is the initial battery capacity of device i. To ensure physical feasibility, the residual energy should always remain non-negative:

$$E_i^{\text{res}}(t) \geq 0, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}. \tag{16}$$

Turning to the UAV, propulsion power is the dominant component of its total energy consumption. Following a common discretization framework in UAV communication literature, we assume that the UAV flies at constant speed along a straight line within each time slot, and the propulsion power is computed using the average speed over the slot. The time step ΔT is chosen sufficiently small such that this piecewise linear approximation is sufficiently accurate for trajectory optimization purposes. We adopt the widely validated model derived from blade element momentum theory [35], and the propulsion power at time slot t is modeled as

$$P_{\text{prop}}^{(t)} = P_0 \left(1 + \frac{3\|\mathbf{v}^{(t)}\|_2^2}{U_{\text{tip}}^2} \right) + P_{\text{ind}} \left(\sqrt{1 + \frac{\|\mathbf{v}^{(t)}\|_2^4}{4v_0^4}} - \frac{\|\mathbf{v}^{(t)}\|_2^2}{2v_0^2} \right)^{1/2} + \frac{1}{2} d_0 \rho_{\text{air}} \sigma A \|\mathbf{v}^{(t)}\|_2^3, \tag{17}$$

where $\mathbf{v}^{(t)} = \frac{\mathbf{q}^{(t)} - \mathbf{q}^{(t-1)}}{\Delta T}$ is the average flight velocity during slot t, with $\|\mathbf{v}^{(t)}\|_2$ being its magnitude; P_0 and P_{ind} denote the blade profile power and induced power in hovering, respectively; U_{tip} is the rotor blade tip speed; v_0 is the mean rotor induced velocity in hovering; d_0 is the fuselage drag ratio; ρ_{air} is the air density; σ is the rotor solidity; and A is the rotor disc area.

For hovering with $\mathbf{v}^{(t)} = \mathbf{0}$, the propulsion power reduces to $P_{\text{hov}} = P_0 + P_{\text{ind}}$.

The total propulsion energy consumed over the entire patrol cycle is

$$E_{\text{prop}}^{\text{total}} = \sum_{t \in \mathcal{T}} P_{\text{prop}}^{(t)} \Delta T. \tag{18}$$

To ensure the UAV can safely complete the patrol and return to the depot, we impose the battery capacity constraint:

$$E_{\text{prop}}^{\text{total}} \leq E_{\text{battery}}, \tag{19}$$

where E_{battery} is the onboard battery capacity of the UAV. These energy models will be integrated into the complete optimization problem in the following subsection.

2.4. Problem Formulation

By integrating the communication model, energy model, and all constraints defined above, we formulate the joint optimization of TP, user association, bandwidth allocation, and power control. The objective is to minimize the weighted sum transmit power of ground IoT devices with energy-aware dynamic weights, while the UAV's propulsion energy is treated as a hard feasibility constraint.

To facilitate concise notation, we define the aggregated variable vectors as follows:

$$\mathbf{Q} = \left[\mathbf{q}^{(t)}\right]_{t \in \mathcal{T}}, \boldsymbol{\alpha} = \left[\alpha_i^{(t)}\right]_{i \in \mathcal{I}, t \in \mathcal{T}}, \boldsymbol{\beta} = \left[\beta_i^{(t)}\right]_{i \in \mathcal{I}, t \in \mathcal{T}}, \boldsymbol{\rho} = \left[\rho_i^{(t)}\right]_{i \in \mathcal{I}, t \in \mathcal{T}}. \tag{20}$$

The complete optimization problem is formally stated as

$$\min_{\mathbf{Q}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\rho}} \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}} w_i^{(t)} \rho_i^{(t)} \beta_i^{(t)} \Delta T \tag{21a}$$

$$\text{s.t. } \mathbf{q}^{(t)} \in S\left(\mathbf{q}^{(t-1)}\right), \forall t \in \mathcal{T}, \tag{21b}$$

$$\mathbf{q}^{(T)} = \mathbf{q}^{(0)}, \tag{21c}$$

$$\sum_{t \in \mathcal{T}} P_{\text{prop}}^{(t)} \Delta T \leq E_{\text{battery}}, \tag{21d}$$

$$\alpha_i^{(t)} \leq d_i^{(t)}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \tag{21e}$$

$$\sum_{t \in \mathcal{T}} \alpha_i^{(t)} \geq 1, \forall i \in \mathcal{I}, \tag{21f}$$

$$\alpha_i^{(t)} \in \{0, 1\}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}. \tag{21g}$$

$$0 \leq \beta_i^{(t)} \leq \alpha_i^{(t)} B, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \tag{21h}$$

$$\sum_{i \in \mathcal{U}^{(t)}} \beta_i^{(t)} \leq B, \forall t \in \mathcal{T}, \tag{21i}$$

$$0 \leq \rho_i^{(t)} \beta_i^{(t)} \leq P_{\max}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \tag{21j}$$

$$R_i^{(t)} \geq \alpha_i^{(t)} r_i, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \tag{21k}$$

$$E_i^{\text{res}}(t) \geq 0, \forall i \in \mathcal{I}, \forall t = 1, ..., T+1, \tag{21l}$$

$$E_i^{\text{res}}(t+1) = E_i^{\text{res}}(t) - \rho_i^{(t)} \beta_i^{(t)} \Delta T, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \tag{21m}$$

$$E_i^{\text{res}}(1) = E_i^{\text{init}}, \forall i \in \mathcal{I}. \tag{21n}$$

The weight $w_i^{(t)}$ in the objective function (21a) is defined as

$$w_i^{(t)} = \frac{1/\left(E_i^{\text{res}}(t) + \delta\right)}{\sum_{j\in\mathcal{I}} 1/\left(E_j^{\text{res}}(t) + \delta\right)}, \tag{22}$$

where δ is a small positive constant to avoid division by zero. Note that the weights $w_i^{(t)}$ in (21a) are themselves functions of the optimization variables through (22) and (21m), reflecting the temporal coupling of the problem.

This weight is energy-aware: devices with lower residual battery are assigned higher weights, incentivizing the algorithm to reduce their transmit power to preserve scarce energy. The residual energy evolves according to (14) and (15), which couples current power decisions with future weights and naturally motivates a sequential decision-making framework. Since the optimization is performed centrally at the UAV, the weights are computed with full knowledge of all devices' residual energies.

The roles of the individual constraints are as follows.

(21b,c) are the UAV mobility constraints, limiting the flight speed and enforcing the return to the depot.

(21d) restricts the total propulsion energy within the UAV's battery capacity.

(21e–g) are the user association constraints: (21e) limits association to the active window of each device, (21f) requires each device to be served at least once per patrol cycle (this hard constraint is relaxed in the algorithmic implementation via a penalty term to avoid infeasibility under tight resource budgets), and (21g) specifies the binary nature of the association variables.

(21h–k) are the per-slot RRM constraints: (21h,i) enforce the bandwidth allocation limits, (21j) bounds the transmit power, and (21k) ensures the QoS requirement for each associated device.

Finally, (21l–n) characterize the device energy dynamics and enforce physical feasibility: (21l) guarantees that the residual energy of each device remains non-negative, (21m) defines the evolution of device residual energy driven by the transmit power decisions, and (21n) specifies the initial energy states.

(21) is a mixed-integer non-convex problem. The main challenges stem from three aspects: binary association variables introduce combinatorial complexity; data rate and propulsion power are non-convex in the UAV's position and velocity; and dynamic energy-aware weights couple decisions across slots via residual energy evolution, rendering the objective nonlinear in historical decisions. Conventional iterative methods alternating between trajectory and resource optimization are prone to the initialization trap, often converging to poor local optima heavily dependent on initial points [21–24].

3. Proposed Two-Layer Non-AO Framework

Building on Section 2, we propose a two-layer non-alternating framework to solve (21). The core idea is to decouple the problem into per-slot RRM for a fixed position and multi-slot TP. This decomposition eliminates the alternating update loop—the root cause of both initialization sensitivity and high computational overhead—and thus circumvents the initialization trap. Numerical searches within each layer are non-alternating, further reducing complexity.

The inner layer provides a KKT-based analytical RRM solution for any fixed UAV position, with complexity $O(I^2)$ in practice ($O(I^3)$ worst-case). The outer layer reformulates

TP as an MDP, where the state and reward are designed to capture time coupling from dynamic weights, enabling full trajectory generation in a single forward pass.

3.1. Inner-Layer RRM: KKT-Based Analytical Solution

The inner-layer subproblem minimizes the weighted sum of device transmit powers via joint optimization of association, bandwidth, and power, subject to (21h–k). To balance optimality against computational efficiency, we decouple it into two steps. First, we fix the association and optimize resource allocation, reducing the problem to a single-variable convex form solvable via KKT conditions. Based on this rule, we design an incremental greedy algorithm that selects users by evaluating the marginal cost increase. For a fixed UAV position, channel gains are constant within a slot, allowing per-slot independent solutions. Hence, we omit the time index t for brevity.

3.1.1. Bandwidth and Power Allocation for Fixed User Association

We now consider the bandwidth and power allocation subproblem for a given user association set $\mathcal{U} \subseteq \mathcal{I}$ at any given time slot. Under fixed association, the binary constraint (21g) is no longer needed. The remaining problem is formulated as

$$\min_{\{\rho_i,\beta_i\}_{i\in\mathcal{U}}} \sum_{i\in\mathcal{U}} w_i\rho_i\beta_i \tag{23a}$$

$$\text{s.t.} \sum\nolimits_{i\in\mathcal{U}}\beta_i \le B, \tag{23b}$$

$$0 \le \rho_i\beta_i \le P_{\max}, \forall i \in \mathcal{U}, \tag{23c}$$

$$\beta_i \ge 0, \forall i \in \mathcal{U}, \tag{23d}$$

$$\beta_i \log_2\left(1+\frac{\rho_i g_i}{N_0}\right) \ge r_i, \forall i \in \mathcal{U}. \tag{23e}$$

For a fixed bandwidth allocation $\{\beta_i\}_{i\in\mathcal{U}}$, the rate constraint (23e) must hold with equality at optimality. Indeed, since the objective term $w_i\rho_i\beta_i$ is strictly increasing in ρ_i, the optimal solution for user i must adopt the minimum ρ_i that satisfies (23e). This minimum is attained when the inequality in (23e) is tight. Solving $\beta_i\log_2(1+\rho_i g_i/N_0) = r_i$ gives the optimal PSD for a given bandwidth allocation:

$$\rho_i^*(\beta_i) = \frac{N_0}{g_i}\left(2^{r_i/\beta_i}-1\right), \forall i \in \mathcal{U}. \tag{24}$$

The corresponding transmit power is $\rho_i^*(\beta_i)\beta_i = \frac{N_0\beta_i}{g_i}\left(2^{r_i/\beta_i}-1\right)$, and the user is infeasible if it exceeds $P_{\max}$.

We note that the objective here is the weighted transmit power rather than the total communication energy. For a given user association, the circuit power of radio transceivers is approximately constant and does not affect the bandwidth and power allocation. Since orthogonal bandwidth allocation is assumed, a larger bandwidth reduces the required transmit PSD and thus the transmit power, which is consistent with the observation that higher bandwidth improves communication efficiency.

We next establish a useful monotonic property of the total transmit power with respect to bandwidth.

Proposition 1. *For any user i, the actual transmit power $\rho_i\beta_i$ is strictly decreasing with respect to β_i. The proof is deferred to Appendix A.*

With Proposition 1, we can now prove that the total bandwidth constraint (23b) must be tight at optimality. To see this, consider any feasible solution with $\sum_{i\in\mathcal{U}}\beta_i < B$. Since there is unused bandwidth, we may slightly increase β_i for any user i without reducing the bandwidth of other users. By Proposition 1, this strictly decreases $\rho_i\beta_i$ and hence the weighted term $w_i\rho_i\beta_i$ in the objective, while all other terms remain unchanged. Thus, any solution with unused bandwidth can be strictly improved, so the optimal solution must satisfy $\sum_{i\in\mathcal{U}}\beta_i = B$.

We now derive the minimum bandwidth constraint imposed by the maximum transmit power limit. By Proposition 1, the transmit power $\rho_i^*(\beta_i)\beta_i$ is strictly decreasing in β_i. Therefore, the condition $\rho_i^*(\beta_i)\beta_i \le P_{\max}$ is equivalent to $\beta_i \ge \beta_i^{\min}$, where $\beta_i^{\min}$ is the unique solution of

$$r_i = \beta_i^{\min}\log_2\left(1+\frac{P_{\max}g_i}{N_0\beta_i^{\min}}\right). \tag{25}$$

If $\beta_i^{\min} > B$, user i cannot be served under the available system bandwidth and is therefore excluded from any feasible candidate association set $\mathcal{U}$.

By substituting (24) into (23a), replacing the transmit power constraint (23c) by the equivalent bandwidth lower bound $\beta_i \ge \beta_i^{\min}$, and using $\sum_{i\in\mathcal{U}}\beta_i = B$, we transform the original two-variable optimization problem (23) into a single-variable problem with respect to bandwidth allocation:

$$\min_{\{\beta_i\}_{i\in\mathcal{U}}} \sum_{i\in\mathcal{U}} \frac{w_iN_0\beta_i}{g_i}\left(2^{r_i/\beta_i}-1\right) \tag{26a}$$

$$\text{s.t. } \sum_{i\in\mathcal{U}}\beta_i = B, \tag{26b}$$

$$\beta_i \ge \beta_i^{\min}, \forall i\in\mathcal{U}. \tag{26c}$$

The objective function (26a) is convex in $\beta_i > 0$ (with $\beta_i = 0$ being infeasible due to (23e)). Indeed, each term $\beta_i\left(2^{r_i/\beta_i}-1\right)$ has a positive second-order derivative over the positive domain. Since (26a) is a positive weighted sum of such convex terms (with weights $w_iN_0/g_i > 0$), the objective is convex. Furthermore, the feasible set defined by (26b) and (26c) is convex. Therefore, problem (26) is a convex optimization problem and can be optimally solved via KKT conditions [36].

To obtain the optimal bandwidth allocation, we construct the Lagrangian of (26) as follows:

$$\mathcal{L} = \sum_{i\in\mathcal{U}}\frac{w_iN_0\beta_i}{g_i}\left(2^{r_i/\beta_i}-1\right) + \lambda\left(\sum_{i\in\mathcal{U}}\beta_i - B\right) - \sum_{i\in\mathcal{U}}\mu_i\left(\beta_i-\beta_i^{\min}\right), \tag{27}$$

where λ is the Lagrange multiplier for the total bandwidth constraint, and $\mu_i \ge 0$ are the multipliers for the lower-bound constraints $\beta_i \ge \beta_i^{\min}$. Applying the KKT stationarity condition gives:

$$\frac{\partial\mathcal{L}}{\partial\beta_i} = \frac{w_iN_0}{g_i}\left[2^{r_i/\beta_i}\left(1-\frac{r_i\ln 2}{\beta_i}\right)-1\right] + \lambda - \mu_i = 0. \tag{28}$$

For users with $\beta_i > \beta_i^{\min}$, the lower-bound constraint is inactive, so $\mu_i = 0$ by complementary slackness, which yields

$$\frac{w_iN_0}{g_i}\left[1-2^{r_i/\beta_i}\left(1-\frac{r_i\ln 2}{\beta_i}\right)\right] = \lambda. \tag{29}$$

For users at the lower bound $\beta_i = \beta_i^{\min}$, the bandwidth is determined directly by (25), and $\mu_i \ge 0$ adjusts accordingly.

Let $\widetilde{\beta}_i(\lambda)$ denote the unique solution of (29) for a given $\lambda > 0$. Then, the optimal bandwidth allocation is

$$\beta_i^*(\lambda) = \max\{\beta_i^{\min}, \widetilde{\beta}_i(\lambda)\}. \tag{30}$$

The multiplier λ is determined by the total bandwidth equation

$$\sum_{i\in\mathcal{U}} \max\{\beta_i^{\min}, \widetilde{\beta}_i(\lambda)\} = B. \tag{31}$$

Since each term is non-increasing in λ and strictly decreasing when $\widetilde{\beta}_i(\lambda) > \beta_i^{\min}$, the left-hand side is strictly decreasing, ensuring a unique λ that can be found by bisection search.

3.1.2. User Association Selection via Incremental Greedy Algorithm

With the ability to compute the optimal bandwidth and power allocation for any given user set $\mathcal{U}$ established in Section 3.1.1, we now turn to the problem of selecting the user association set $\mathcal{U}^{(t)} \subseteq \mathcal{I}$ at each time slot.

The inner-layer RRM focuses on minimizing the instantaneous weighted sum power, leaving the long-term obligation of serving all devices to the outer MDP layer. Under this separation of concerns, the per-slot association strategy incrementally admits users with smallest marginal cost increase it adds users incurring the smallest marginal weighted sum power increase, until no more users can be feasibly served.

To formalize the selection process, consider a given active user set $\mathcal{U} \subseteq \mathcal{I}_{\text{active}}$, where $\mathcal{I}_{\text{active}} = \{i | d_i = 1\}$ at the current time slot. The optimal bandwidth and power allocation is obtained by solving (26) for $\{\beta_i\}$, with $\{\rho_i\}$ then given by (24). To quantify the cost of selecting a particular user set, we define the cost function as

$$C(\mathcal{U}) = \sum_{i\in\mathcal{U}} w_i \rho_i^*(\mathcal{U}) \beta_i^*(\mathcal{U}), \tag{32}$$

which is exactly the optimal value of problem (23) for the given user set $\mathcal{U}$.

To evaluate the marginal cost increase, we first need an efficient way to verify the feasibility of candidate sets. We observe the following monotonicity property: if a user set $\mathcal{U}$ is feasible—i.e., there exists a bandwidth and power allocation satisfying all constraints in (23)—then any subset $\mathcal{U}' \subseteq \mathcal{U}$ is also feasible. On the other hand, adding users to a feasible set may render it infeasible if the bandwidth budget is insufficient. This property enables us to construct feasible sets incrementally: once $\mathcal{U}_{\text{current}}$ is feasible, only the feasibility of $\mathcal{U}_{\text{current}} \cup \{i\}$ needs to be verified for each candidate i.

The optimization over $\mathcal{U}$ is combinatorial in nature. Exhaustive search over all 2^I subsets is computationally prohibitive for large I. We therefore adopt an incremental greedy algorithm (Algorithm A1) that starts from the empty set and iteratively adds the user that yields the smallest marginal cost increase while maintaining feasibility, until no additional feasible additions are possible.

A key component of this greedy procedure is the verification of candidate set feasibility. To reduce the computational overhead of KKT verification, we first derive a lightweight necessary and sufficient condition for feasibility based on the minimum bandwidth $\beta_i^{\min}$ defined in (25). For any candidate user set $\mathcal{U}_{\text{candidate}}$, a necessary and sufficient condition for feasibility is

$$\sum_{j\in\mathcal{U}_{\text{candidate}}} \beta_j^{\min} \leq B. \tag{33}$$

This condition follows directly from the fact that the total bandwidth must be at least the sum of individual minimum bandwidths, and if (33) holds, the remaining bandwidth

$B - \sum_{j \in \mathcal{U}_{\text{candidate}}} \beta_j^{\min}$ can always be allocated among users to satisfy all constraints. We use (33) as a fast pre-check before solving the exact KKT allocation in (26).

Algorithm A1 summarizes the incremental greedy procedure, and its detailed pseudo-code is provided in Appendix B. Since at most one user is added per iteration, the outer loop runs at most I times. In the k-th iteration ($|\mathcal{U}_{\text{current}}| = k$), the algorithm evaluates at most $I - k$ candidate sets. For each candidate, solving the bandwidth allocation via (30) and (31) involves solving the KKT conditions for the k users in the candidate set, requiring $O(k)$ operations per evaluation. The worst-case complexity is therefore $\sum_{k=0}^{I-1} (I-k) \cdot O(k) = O(I^3)$.

In practice, however, the fast pruning step filters out a large fraction of candidates, especially as k increases, substantially reducing the number of KKT evaluations. When the bandwidth constraint is tight, and most candidates are pruned in later iterations, the empirical complexity scales roughly as $O(I^2)$.

For a fixed user association set $\mathcal{U}$, problem (26) is convex in $\{\beta_i\}$, so the KKT conditions yield the globally optimal bandwidth and power allocation. For user association selection, the incremental greedy algorithm provides a polynomial-time approximate solution. While the combinatorial association problem is generally NP-hard [37], the greedy selection based on marginal cost increment achieves near-optimal performance in practice. Overall, the inner-layer RRM delivers a computationally efficient solution with guaranteed optimality in the resource allocation subproblem and near-optimality in user association. A rigorous theoretical analysis of the greedy approximation ratio is left as future work.

3.2. Outer-Layer TP: MDP Reformulation

For notational convenience, we define $f(\mathbf{q}^{(t)}) = C\left(\mathcal{U}^*(\mathbf{q}^{(t)})\right)$ as the minimum weighted sum power obtained from the inner-layer RRM at a given $\mathbf{q}^{(t)}$, where $\mathcal{U}^*(\mathbf{q}^{(t)})$ is the near-optimal user association set from Algorithm A1. This function bridges the inner-layer RRM and the outer-layer TP.

With the inner-layer RRM solution established in Section 3.1, which yields the interface function $f(\mathbf{q}^{(t)})$ for any given UAV position $\mathbf{q}^{(t)}$, we now turn to the outer-layer problem: finding the optimal trajectory $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(T)}$ that minimizes the weighted sum power over the entire patrol cycle while satisfying all mobility and energy constraints.

The trajectory optimization subproblem can be extracted from (21) as follows:

$$\min_{\mathbf{q}^{(1)},\ldots,\mathbf{q}^{(T)}} \sum_{t=1}^{T} f(\mathbf{q}^{(t)}) \tag{34a}$$

$$\text{s.t. } \mathbf{q}^{(t)} \in S\left(\mathbf{q}^{(t-1)}\right), \forall t \in \mathcal{T}, \tag{34b}$$

$$\mathbf{q}^{(T)} = \mathbf{q}^{(0)}, \tag{34c}$$

$$\sum_{t=1}^{T} P_{\text{prop}}^{(t)} \Delta T \leq E_{\text{battery}}, \tag{34d}$$

$$E_i^{\text{res}}(t+1) = E_i^{\text{res}}(t) - \rho_i^{(t)} \beta_i^{(t)} \Delta T, \forall i \in \mathcal{I}. \tag{34e}$$

In (34e), $\rho_i^{(t)} \beta_i^{(t)}$ is the optimal transmit power from the inner layer at position $\mathbf{q}^{(t)}$. Note that (34e) describes the evolution of device residual energy, which is handled as a state transition rather than a hard constraint at this stage.

Problem (34) is a sequential decision-making problem with strong temporal dependencies: the current position decision affects future feasible positions via the velocity constraint, the remaining UAV battery through propulsion energy consumption, and the device residual energies via the power allocation at each visited position. Moreover, the dynamic weight $w_i^{(t)}$ at future slots depends on the cumulative power consumption from previ-

ous decisions, which further intensifies the temporal coupling. These inherent sequential dependencies naturally motivate an MDP formulation [38].

3.2.1. MDP Formulation

We formulate the TP problem as a finite-horizon MDP defined by the tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{P}$ is the state transition probability, $\mathcal{R}$ is the reward function, and $\gamma \in [0,1]$ is the discount factor. Since the patrol mission has a fixed finite time horizon T, we set $\gamma = 1$, which is equivalent to optimizing the total undiscounted reward over the finite horizon. Moreover, the environment is deterministic given the UAV's actions, so the transition probabilities degenerate to $\mathcal{P}(s'|s,a) = 1$ for the unique next state.

The state at time slot t includes all information necessary for decision-making: the current UAV position $\mathbf{q}^{(t-1)}$, the remaining UAV battery $E_{\text{UAV}}^{\text{res}}(t)$, the residual energies $\{E_i^{\text{res}}(t)\}_{i\in\mathcal{I}}$ of each ground device, the active window indicators $\left\{d_i^{(t)}\right\}_{i\in\mathcal{I}}$, and the service status flags $\left\{m_i^{(t)}\right\}_{i\in\mathcal{I}}$, where $m_i^{(t)} \in \{0,1\}$ indicates whether device i has been served at least once by the beginning of slot t. Formally, the state is defined as

$$s^{(t)} = \left(\mathbf{q}^{(t-1)}, E_{\text{UAV}}^{\text{res}}(t), \{E_i^{\text{res}}(t)\}_{i\in\mathcal{I}}, \left\{d_i^{(t)}\right\}_{i\in\mathcal{I}}, \left\{m_i^{(t)}\right\}_{i\in\mathcal{I}}\right) \in \mathcal{S}. \tag{35}$$

The dimension of the state vector is $3I + 4$, which scales linearly with the number of devices. However, the size of the discrete state space grows exponentially with I, which motivates the use of heuristic and learning-based solution methods for the MDP. The initial state $s^{(1)}$ at the beginning of the patrol cycle is given by

$$s^{(1)} = \left(\mathbf{q}^{(0)}, E_{\text{battery}}, \left\{E_i^{\text{init}}\right\}_{i\in\mathcal{I}}, \left\{d_i^{(1)}\right\}_{i\in\mathcal{I}}, \{0\}_{i\in\mathcal{I}}\right), \tag{36}$$

where all devices have $m_i^{(1)} = 0$ since none have been served yet.

At each time slot t, the UAV selects its next position from the set of reachable grid points as its action:

$$a^{(t)} = \mathbf{q}^{(t)} \in S(\mathbf{q}^{(t-1)}), \tag{37}$$

where $S(\cdot)$ is the feasible position set defined in Section 2.1. The return-to-depot constraint is not imposed as a hard restriction on the action space at the final step; instead, it is enforced through a sufficiently large penalty in the reward function. This incentivizes the MDP agent to terminate the patrol at the depot while maintaining a non-empty action space at every state.

Given the current state $s^{(t)}$ and action $\mathbf{q}^{(t)}$, the next state $s^{(t+1)}$ is deterministically updated as follows. The UAV position is updated to $\mathbf{q}^{(t)}$. The UAV remaining energy evolves as $E_{\text{UAV}}^{\text{res}}(t+1) = E_{\text{UAV}}^{\text{res}}(t) - P_{\text{prop}}^{(t)}\Delta T$, where $P_{\text{prop}}^{(t)}$ is computed via (17). For each device i, the residual energy follows $E_i^{\text{res}}(t+1) = E_i^{\text{res}}(t) - \rho_i^{(t)}\beta_i^{(t)}\Delta T$, where $\rho_i^{(t)}\beta_i^{(t)}$ is the optimal power from the inner-layer RRM at position $\mathbf{q}^{(t)}$. The service status is updated as $m_i^{(t+1)} = \max(m_i^{(t)}, \alpha_i^{(t)})$, where $\alpha_i^{(t)}$ is the association decision from the inner layer. This ensures that $m_i^{(t)} = 1$ once device i has been served at least once up to slot t. Finally, the active window indicators $d_i^{(t+1)}$ are updated according to (4) based on the next time slot.

The single-step reward is defined as the negative of the minimum weighted sum power achieved at the chosen position, with additional penalty terms for constraint violations:

$$\begin{aligned} r(s^{(t)}, a^{(t)}) = &-f(\mathbf{q}^{(t)}) - \lambda_{\text{UAV}} \cdot \mathbf{1}\{E_{\text{UAV}}^{\text{res}}(t+1) < 0\} - \lambda_{\text{dev}} \cdot \textstyle\sum_{i=1}^{I} \mathbf{1}\{E_i^{\text{res}}(t+1) < 0\} - \\ &\lambda_{\text{serve}} \cdot \textstyle\sum_{i=1}^{I} \mathbf{1}\left\{m_i^{(t+1)} = 0 \text{ and } t = e_i - 1\right\} - \lambda_{\text{depot}} \cdot \mathbf{1}\left\{t = T \text{ and } \mathbf{q}^{(T)} \neq \mathbf{q}^{(0)}\right\}. \end{aligned} \tag{38}$$

The penalty weights λ_{UAV}, λ_{dev}, λ_{serve}, and λ_{depot} are all positive and correspond respectively to UAV battery depletion, device battery depletion (enforcing the non-negative residual energy constraint (21l)), missed service deadlines (evaluated at the last time slot of each device's active window, assuming $e_i \leq T+1$ for all i, i.e., all validity windows fall within the patrol horizon), and failure to return to the depot at the end of the patrol.

These weights are set to sufficiently large values such that any constraint-violating or non-returning trajectory yields a total reward worse than all feasible trajectories. Under this setting, the optimal policy of the MDP automatically satisfies the hard constraints in (34). The total accumulated reward over the episode is

$$R_{\text{total}} = \sum_{t=1}^{T} r\left(s^{(t)}, a^{(t)}\right), \tag{39}$$

and maximizing this total reward is equivalent to minimizing the weighted sum power while satisfying all constraints.

3.2.2. Solution Methods for the MDP

The MDP formulated in the previous subsection can be solved using various decision-making algorithms. Following the framework of [24], three representative approaches are introduced: a GA for near-global exploration, a limited DFS for exact search under short horizons, and a DQN for learning-based approximation suitable for large-scale deployment. These methods cover different trade-offs between optimality and computational complexity. We elaborate on each scheme in detail below; the DQN-based implementation is left for future work.

GA: The trajectory is encoded as a gene $\mathbf{Q} = [\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(T)}]$, where each position $\mathbf{q}^{(t)}$ is constrained to the reachable set $S(\mathbf{q}^{(t-1)})$. The fitness of a gene is defined as the total accumulated reward R_{total} in (39). Figure 2 illustrates the gene encoding, fitness evaluation, and elite selection process. A canonical GA with elitism is employed [39], where the top elite individuals are preserved across generations to ensure monotonic improvement. The initial population is generated by randomly sampling kinematically feasible trajectories, and crossover is performed by exchanging sub-trajectories between two parent genes while preserving kinematic feasibility by ensuring all waypoints remain within their respective reachable sets. Mutation randomly perturbs a subset of waypoints within their reachable sets with a prespecified mutation probability. The algorithm terminates when the maximum number of generations is reached. GA serves as a near-upper-bound baseline, as it explores the trajectory space globally and can approach the global optimum given sufficient generations.

DFS: A rolling-horizon depth-first search aims to find the best sub-trajectory of length n (search depth) via exhaustive tree search over the discretized grid space and repeats the computation until the entire time horizon is covered [40]. At each step, the algorithm enumerates all possible n-step trajectories starting from the current position, evaluates the accumulated reward of each partial trajectory, and commits to the first action of the best partial trajectory. Figure 3 illustrates the search tree structure for a representative scenario. When the search depth covers the entire horizon and the state space is fully discretized, DFS yields the globally optimal trajectory for the deterministic finite-horizon MDP. In practice, however, n is limited by memory and computation time due to the exponential growth of the search tree, which has a branching factor of up to $M = \max_{\mathbf{q} \in \mathcal{Q}} |S(\mathbf{q})|$. DFS is therefore best suited for small-scale scenarios where exact optimality is desired.

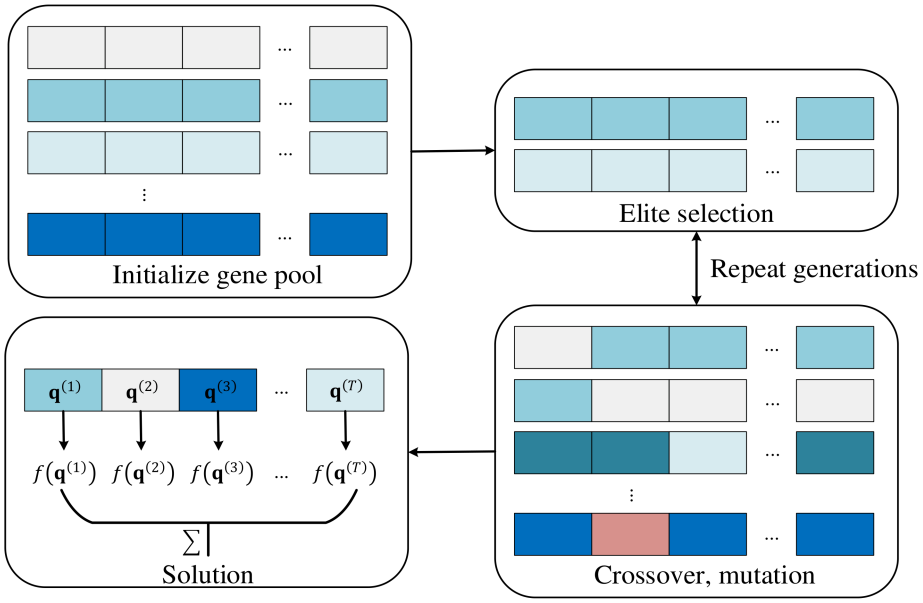

Figure 2. Illustration of the GA-based trajectory encoding and fitness evaluation.

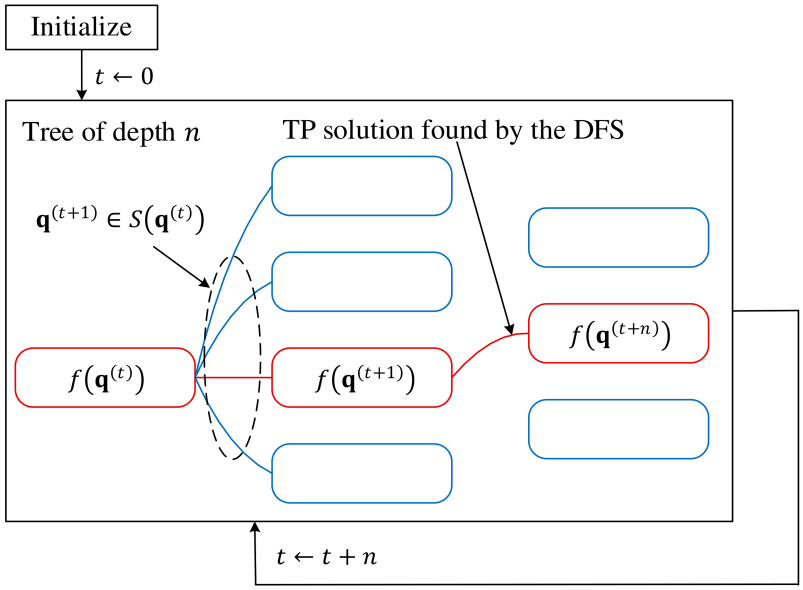

Figure 3. Tree representation of the DFS space with depth n.

DQN: A reinforcement learning algorithm approximates the optimal action-value function $Q(s, a)$ using a deep neural network [41], making it suitable for large-scale scenarios where the state space is too large for exhaustive search.

The DQN-based approach is compatible with the proposed framework; however, its full implementation and evaluation are left as future work due to the significant training overhead and the need for extensive hyperparameter tuning. This paper focuses on GA and DFS as representative solvers, which already provide compelling validation of the proposed framework.

3.3. Overall Algorithm

With the inner-layer RRM solution and the outer-layer MDP formulation established in the preceding subsections, we now assemble them into a complete non-AO framework for Problem (21), which is summarized as Algorithm A2, with its detailed pseudo-code provided in Appendix B. For each candidate trajectory, the RRM solution is computed

on-the-fly via Algorithm A1 and the KKT-based analytical resource allocation in (30), (31), and (24), without any alternating updates between TP and RRM.

The MDP solver then uses the returned total reward to evaluate and improve candidate trajectories, and finally selects the best trajectory found.

We now analyze the computational complexity of the proposed MDP-based TP framework and discuss its practical implications. The overall complexity of the outer-layer MDP solver is determined by two factors: the cost of evaluating the reward function at each visited position, and the number of position evaluations required by the chosen solution algorithm.

Evaluating the term $f(\mathbf{q}^{(t)})$ in the reward function (38) requires invoking the inner-layer RRM procedure (i.e., the analytical bandwidth and power allocation in Section 3.1.1 together with the greedy user association in Algorithm A1) at the candidate position $\mathbf{q}^{(t)}$. This procedure has a worst-case complexity of $O(I^3)$ and an empirical complexity of $O(I^2)$ in practice. The state transition and penalty computations are $O(I)$. Therefore, the per-step cost is dominated by the inner-layer RRM, denoted as C_{RRM}, which is $O(I^3)$ in the worst case and $O(I^2)$ in practice.

The overall computational complexity is therefore

$$C_{\text{total}} = N_{\text{eval}} T C_{\text{RRM}}, \tag{40}$$

where N_{eval} is the number of position evaluations per time step.

The per-step position evaluations for each solver are summarized in Table 1.

Table 1. Complexity of MDP solution methods.

Algorithm	N_{eval}	Total Evaluations	Complexity	Dominant Factor
GA	GP	GPT	$O(GPTC_{\text{RRM}})$	Generations G and population P
DFS	M^n	$M^n T$	$O(M^n TC_{\text{RRM}})$	Search depth n (branching factor M)
* DQN	N_{ep}	$N_{\text{ep}} T$	$O(N_{\text{ep}} TC_{\text{RRM}})$	Training episodes N_{ep}

* DQN complexity is provided for completeness; its practical evaluation is left for future work. Where G is the number of generations, P is the population size, $M = \max_{\mathbf{q}\in\mathcal{Q}} |S(\mathbf{q})|$ is the maximum number of reachable positions per step, n is the DFS depth, and N_{ep} is the number of training episodes.

Compared with conventional AO approaches, the proposed MDP-based framework offers a fundamentally different complexity profile. AO methods require multiple outer iterations between trajectory optimization and resource allocation (typically 10–20 iterations), where each outer iteration typically involves solving a non-convex trajectory subproblem using SCA or similar techniques, with its own inner iterations. In contrast, in our framework each candidate trajectory is evaluated in a single forward pass, and the RRM solution is computed only once per visited position, without any alternating updates between TP and RRM. This structural difference not only reduces computational overhead but also eliminates the initialization sensitivity inherent in AO methods.

4. Simulation Results and Discussion

We evaluate the proposed framework through numerical simulations on a square-shaped ground area of size $w \times w = 600 \times 600$ m^2, where I IoT devices are randomly and uniformly deployed. The UAV operates at a fixed altitude of $h = 50$ m throughout the simulations, with a maximum flight speed of $v = 15$ m/s. This fixed-altitude setting is adopted to simplify the simulation and focus on the core contributions of the proposed framework; all benchmark schemes operate under the same altitude condition to ensure fair comparison. The flight map is discretized with grid resolution $\Delta Q = 40$ m, and the

mission horizon is divided into $T = 20$ time slots, each with duration $\Delta T = 3$ s. The system bandwidth is set to $B = 5$ MHz, the carrier frequency is $f = 2$ GHz, and the noise PSD is $N_0 = -174$ dBm/Hz. Each device has a maximum transmit power of $P_{\max} = 23$ dBm. The probabilistic LoS/NLoS channel model follows the parameters in [32–34]: $a_{\text{env}} = 9.64$, $b_{\text{env}} = 0.06$, $\eta_{\text{LoS}} = 1$ dB, and $\eta_{\text{NLoS}} = 40$ dB. The UAV propulsion model adopts the rotary-wing parameters from [35]: $P_0 = 79.8$ W, $P_{\text{ind}} = 88.6$ W, $U_{\text{tip}} = 120$ m/s, $v_0 = 4.03$ m/s, $d_0 = 0.6$, $\rho_{\text{air}} = 1.225$ kg/m^3, $\sigma = 0.05$, and $A = 0.503$ m^2. The UAV starts from the center of the service area with initial position $\mathbf{q}^{(0)} = [300, 300, 50]^{\text{T}}$. The UAV battery capacity is set to $E_{\text{battery}} = 2 \times 10^4$ J (approximately 5.56 Wh), which is a conservative value for small rotary-wing UAVs. This choice ensures that the propulsion energy constraint is active and meaningful within the mission horizon, thereby allowing the energy-limited nature of the UAV to be properly reflected in the trajectory planning evaluation.

For simplicity, each device i has a fixed initial residual energy of $E_i^{\text{init}} = 0.5$ J and a full active window spanning the entire mission horizon, i.e., $[s_i, e_i) = [1, T+1)$, meaning all devices are active throughout the patrol cycle. The QoS requirement is set to $r_i = 5$ Mbps for all devices. The main simulation parameters are summarized in Table 2. All simulations are performed using a custom Python simulator (Python version 3.12), and the figures are generated with MATLAB R2022b. All runtime measurements were conducted on a workstation equipped with an AMD Ryzen 9 7945HX CPU and 40 GB RAM. Orthogonal bandwidth allocation is adopted, so no inter-device interference is present within each time slot; the only noise source is AWGN. All results are averaged over 50 independent random deployments with different random seeds, and the reported values are subject to a 95% confidence interval unless otherwise stated.

Table 2. Simulation parameter configurations.

Parameter	Symbol	Value
Map width	w	600 m
UAV altitude	h	50 m
UAV maximum speed	v	15 m/s
Grid resolution	ΔQ	40 m
Number of time slots	T	20
Slot duration	ΔT	3 s
Initial UAV position	$\mathbf{q}^{(0)}$	$[300, 300, 50]^{\text{T}}$ m
UAV battery capacity	E_{battery}	2×10^4 J
System bandwidth	B	5 MHz
Carrier frequency	f	2 GHz
Noise PSD	N_0	-174 dBm/Hz
Maximum transmit power	$P_{\max}$	23 dBm
LoS/NLoS environment parameter	$(a_{\text{env}}, b_{\text{env}})$	(9.64,0.06)
LoS excessive pathloss	η_{LoS}	1 dB
NLoS excessive pathloss	η_{NLoS}	40 dB
Blade profile power	P_0	79.8 W
Induced power in hover	P_{ind}	88.6 W
Rotor blade tip speed	U_{tip}	120 m/s
Mean rotor induced velocity in hover	v_0	4.03 m/s
Fuselage drag ratio	d_0	0.6
Air density	ρ_{air}	1.225 kg/m^3
Rotor solidity	σ	0.05
Rotor disc area	A	0.503 m^2
Initial device energy	E_i^{init}	0.5 J
QoS rate requirement	r_i	5 Mbps

We evaluate the proposed framework using two trajectory solvers introduced in Section 3.2.2 and denote them as Proposed-GA and Proposed-DFS, representing global exploration and limited exact search. For Proposed-DFS, the search depth is set to $n = 3$. Both schemes employ the same inner-layer RRM procedure detailed in Section 3.1 to ensure consistent resource allocation. The DQN-based trajectory solver, while conceptually supported by the framework, is left as future work.

The proposed schemes are compared against the following baselines.

GA-RRM serves as a GA-based benchmark for the inner-layer RRM: it uses a GA to jointly optimize user association, bandwidth, and power allocation directly for a fixed UAV position, independent of the KKT-based analytical solution. GA-RRM is used to evaluate the performance of the inner-layer RRM in Section 4.1.

EX-RRM performs exhaustive search over all possible user association subsets for a fixed UAV position and, for each subset, solves the bandwidth and power allocation optimally via the KKT-based method in Section 3.1.1. It therefore provides the globally optimal solution for the inner-layer RRM. Due to its exponential complexity $O(2^I)$, EX-RRM is only evaluated in small-scale scenarios (e.g., $I = 10$) as an optimality benchmark.

GA-AO adopts a conventional AO strategy, where TP (using GA) and RRM (using the same inner-layer procedure as the proposed schemes) are iteratively optimized in an alternating fashion until convergence. This serves as the AO baseline to validate the effectiveness of the non-alternating structure.

Fixed-TP keeps the UAV stationary at the center of the service area, i.e., $\mathbf{q}^{(t)} = [300, 300, 50]^{\mathrm{T}}$ for all $t \in \mathcal{T}$, while the RRM is optimized via the same inner-layer procedure in Section 3.1. This baseline isolates the benefit of adaptive TP.

GA parameters are identical across GA-based schemes: population size 50, elite count 10, and mutation probability 10%, consistent with [24]. The number of generations is set to 10,000 for the evaluations in Sections 4.1 and 4.2, and reduced to 1000 for the scalability evaluations in Sections 4.3 and 4.4 to manage computational cost while preserving the relative performance trends among the compared schemes.

In all experiments, the penalty weights in (38) are set to $\lambda_{\mathrm{UAV}} = \lambda_{\mathrm{dev}} = \lambda_{\mathrm{serve}} = \lambda_{\mathrm{depot}} = 5$, which is comparable in magnitude to the weighted sum power objective and thus penalizes constraint violations without dominating the optimization objective.

Throughout this section, all reported values of average weighted transmit power are defined as the total weighted transmit energy in (21a) divided by $T\Delta T$.

4.1. Near-Optimality of the Inner-Layer RRM

We first evaluate the performance of the inner-layer RRM solution by comparing it against two benchmarks: GA-RRM and EX-RRM. This subsection focuses on a single representative time slot with a fixed UAV position; therefore, the metric considered here is the instantaneous weighted transmit power, not the average power over the patrol cycle. Unlike the KKT-based analytical solution, GA-RRM performs a stochastic search over the joint space of all RRM variables via a genetic algorithm, while EX-RRM performs exhaustive search over all 2^I possible user association subsets and, for each subset, solves the bandwidth and power allocation optimally via the KKT-based method in Section 3.1.1. Table 3 reports the weighted sum power achieved by Algorithm A1 and EX-RRM as percentages of that achieved by GA-RRM, for $I \in \{10, 20, 30, 40, 50\}$, so a value of 100% indicates identical performance to GA-RRM.

Table 3. Weighted sum power of different inner-layer RRM methods relative to GA-RRM (%).

Algorithm	$I = 10$	$I = 20$	$I = 30$	$I = 40$	$I = 50$
GA-RRM	100	100	100	100	100
Algorithm A1	100.00	94.44	88.27	73.05	58.19
EX-RRM	99.67	—	—	—	—

Several observations can be drawn from Table 3. First, for $I = 10$, Algorithm A1 matches GA-RRM exactly. Moreover, EX-RRM achieves 99.67% of the GA-RRM value, indicating that the globally optimal solution is only 0.33% lower than the result obtained by both GA-RRM and Algorithm A1. This confirms that Algorithm A1 attains near-optimal performance in this small-scale setting, with a negligible gap to the exhaustive optimum. Second, for all tested values of I, the relative value of Algorithm A1 is always less than or equal to 100%, meaning that Algorithm A1 consistently achieves a weighted sum power no worse than GA-RRM. This behavior is expected: for any fixed user association set, problem (26) is convex, and the KKT conditions yield the globally optimal bandwidth and power allocation. However, the user association subproblem itself is combinatorial, and GA-RRM is a heuristic that may fail to converge to the global optimum, especially as I increases. The fact that Algorithm A1 consistently achieves no worse performance than GA-RRM, together with the tiny gap to EX-RRM at $I = 10$, confirms that the greedy association strategy is highly effective in practice, at least in the small-scale regime where the global optimum is known.

The decreasing trend of Algorithm A1 relative to GA-RRM as I increases reflects the growing difficulty for GA-RRM to converge within the allocated generations as the search space dimensionality grows. In low-dimensional settings, GA-RRM can fully converge and thus serves as a reliable reference; as I increases, the search space expands, and GA-RRM frequently fails to converge, as evidenced by the decreasing relative values. This confirms that Algorithm A1 consistently achieves no worse performance than GA-RRM. However, because EX-RRM is computationally prohibitive for $I > 10$, the near-optimality of Algorithm A1 in larger-scale settings is supported only by empirical comparisons and should not be interpreted as a formal guarantee. The small optimality gap verified at $I = 10$ provides encouraging evidence for the greedy strategy's effectiveness, but further theoretical analysis is needed to establish its approximation ratio.

In terms of computational complexity, Algorithm A1 achieves this performance—which is provably optimal for fixed user associations and near-optimal at small scales—with an empirical complexity of $O(I^2)$, as analyzed in Section 3.1.2. GA-RRM, by contrast, requires $O(GPI)$ operations, where $G = 10{,}000$ is the number of generations and $P = 50$ is the population size. Even for a moderate network with $I = 20$, this translates to approximately 10^7 operations for GA-RRM, whereas Algorithm A1 requires only a few hundred evaluations per time slot. EX-RRM, while providing the exact optimum, has a prohibitive complexity of $O(2^I)$ and is thus limited to $I = 10$ in our simulations. This stark contrast confirms that the proposed inner-layer RRM offers an exceptional trade-off between near-optimality and computational efficiency, making it highly promising for real-time IoT deployments.

4.2. Effectiveness of the Non-Alternating Framework

Next, we evaluate the sensitivity of the proposed framework to the initial trajectory choice to validate the effectiveness of the non-alternating structure. To isolate the effect of the initial position from other sources of randomness, we fix a single device topology with $I = 20$ devices (randomly generated and kept constant throughout this experiment).

Instead of fixing the initial UAV position, we generate 15 random initial positions uniformly sampled within the 600×600 m^2 service area. For each initial position, we independently run Proposed-GA (non-alternating) and GA-AO (alternating), recording the resulting average weighted transmit power over the patrol cycle. Both schemes use the same GA parameters for TP and the same inner-layer RRM procedure, so any performance difference is attributed solely to the optimization structure.

Figure 4 shows the box plots of the average weighted transmit power achieved by Proposed-GA and GA-AO across the 15 initial positions, and Table 4 summarizes the corresponding statistics. GA-AO exhibits a wide interquartile range and substantial spread between the best and worst cases, indicating high sensitivity to the initial trajectory. In several initial positions, GA-AO converges to significantly higher average weighted transmit power, confirming that the alternating structure can trap the optimization in inferior local minima. By contrast, Proposed-GA achieves consistently low average weighted transmit power with minimal variance, demonstrating that the non-alternating structure is inherently robust to the initial choice.

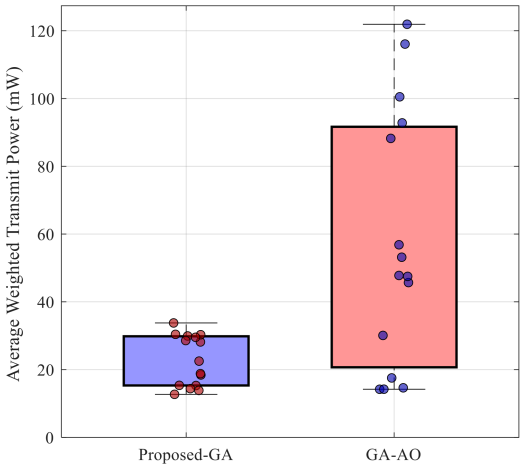

Figure 4. Box plots of average weighted transmit power achieved by Proposed-GA and GA-AO across 15 random initial positions.

Table 4. Performance summary of Proposed-GA and GA-AO across 15 random initial positions.

Algorithm	Mean (mW)	Std (mW)	Min (mW)	Max (mW)
Proposed-GA	22.80	7.27	12.70	33.76
GA-AO	57.41	36.37	14.20	121.92

The worst-case performance of Proposed-GA (33.76 mW) remains competitive with the mean performance of GA-AO (57.41 mW), and is substantially lower than GA-AO's worst-case (121.92 mW). On average, GA-AO achieves approximately 2.5 times higher average weighted transmit power than Proposed-GA. GA-AO also exhibits a substantially larger standard deviation (36.37 mW) than Proposed-GA (7.27 mW), confirming that its performance is highly sensitive to the initial trajectory, whereas the proposed framework delivers consistently reliable results regardless of the starting point.

These results provide direct experimental evidence that the proposed two-layer non-alternating framework effectively circumvents the initialization trap inherent in conventional AO methods, validating the design rationale presented in Section 3. The framework delivers robust near-optimal performance regardless of the starting point.

4.3. Performance Under Varying Network Sizes

We now evaluate the scalability and robustness of the proposed framework by varying the number of devices $I \in \{10, 20, 30, 40, 50\}$ while keeping all other simulation parameters identical. For each configuration, we evaluate four schemes: Proposed-GA, Proposed-DFS, GA-AO, and Fixed-TP.

In addition to the average weighted transmit power, this subsection also evaluates two additional metrics that reflect the service fairness and energy balance of the network: (i) the average service count per device, defined as $\frac{1}{I}\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}}\alpha_i^{(t)}$, which reflects the service efficiency of the TP and scheduling policy; and (ii) the standard deviation of residual energies at the end of the patrol, defined as $\sigma_E = \sqrt{\frac{1}{I}\sum_{i\in\mathcal{I}}\left[E_i^{\text{res}}(T+1) - \overline{E}\right]^2}$, where $\overline{E} = \frac{1}{I}\sum_{i\in\mathcal{I}} E_i^{\text{res}}(T+1)$ and a smaller σ_E indicates more balanced energy depletion and thus prolonged network lifetime.

Figures 5–7 present the three performance metrics—average weighted transmit power, average service count per device, and standard deviation of residual energies—as functions of the number of devices I. The proposed non-alternating framework consistently outperforms the baseline schemes across all three metrics, validating its effectiveness in balancing power minimization, service fairness, and energy balance.

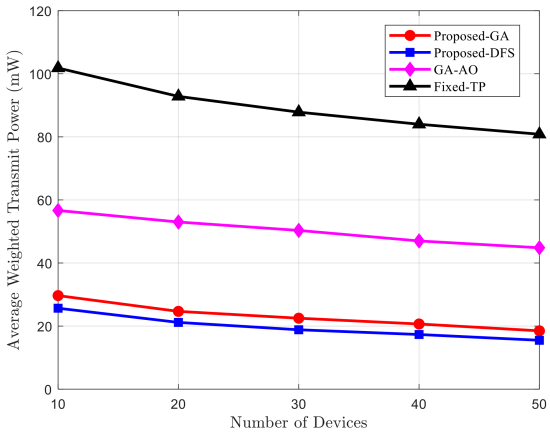

Figure 5. Average weighted transmit power vs. number of devices.

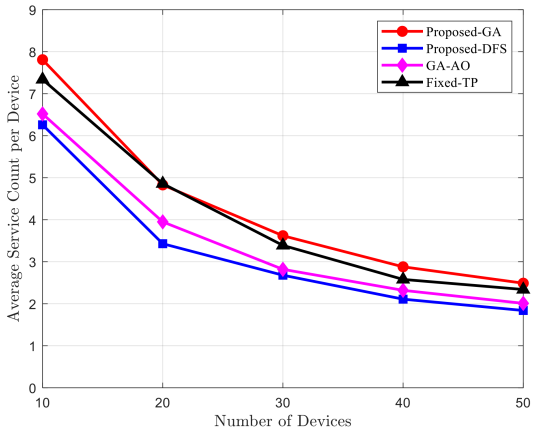

Figure 6. Average service count per device vs. number of devices.

As shown in Figure 5, the average weighted transmit power decreases with I for all schemes, which, when examined together with Figure 6, reflects the reduced per-device transmission opportunities as the number of contending devices increases under limited bandwidth. Among all schemes, Proposed-DFS achieves the lowest average weighted transmit power across the entire range of I, followed closely by Proposed-GA. However,

Figure 7 reveals that this power advantage of Proposed-DFS comes at a cost: it serves the fewest devices among all schemes. By contrast, Proposed-GA achieves near-lowest average weighted transmit power while maintaining the highest service counts across most network sizes (with only a marginal exception at $I = 20$), offering the best overall trade-off between power efficiency and service fairness. GA-AO exhibits both higher average weighted transmit power and lower service counts than Proposed-GA across all configurations, confirming the inferiority of the alternating structure. Fixed-TP shows the highest average weighted transmit power among all schemes, highlighting the critical role of adaptive TP in reducing device transmit power.

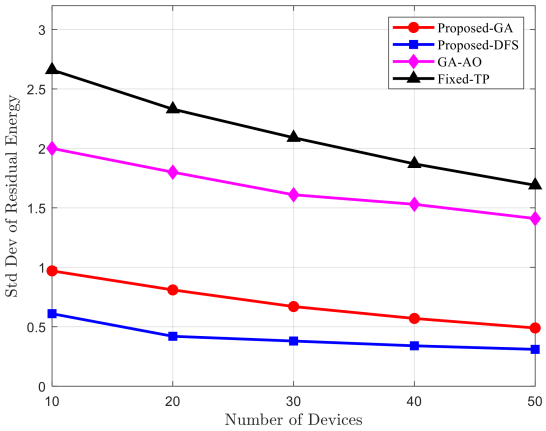

Figure 7. Standard deviation of residual energies σ_E vs. number of devices.

The standard deviation of residual energies in Figure 7 further distinguishes the schemes. Proposed-GA and Proposed-DFS achieve substantially lower σ_E than GA-AO and Fixed-TP across all network sizes, confirming that the non-alternating framework enables the dynamic weighting mechanism to more effectively balance energy consumption across devices. Notably, although Proposed-DFS achieves slightly better energy balance, Proposed-GA offers substantially higher service fairness with only moderately higher σ_E, reinforcing its role as the most balanced scheme overall. The consistently low σ_E of the proposed schemes validates that the non-alternating structure, combined with energy-aware dynamic weights, successfully prevents premature energy depletion of critical devices.

In particular, Proposed-DFS achieves the lowest average weighted transmit power and the best energy balance, making it well-suited for scenarios where power efficiency and energy fairness are paramount. However, its low service count may limit its applicability in mission-critical or delay-sensitive IoT applications. By contrast, Proposed-GA, while slightly inferior in power and energy balance, offers substantially higher service fairness, making it a more balanced choice for typical IoT data collection missions where both fairness and longevity matter. The choice between the two solvers thus depends on the specific network requirements and performance priorities. This validates the core design rationale of the proposed framework: by decoupling TP and RRM and embedding dynamic energy-aware weights into the MDP state, the framework achieves a favorable trade-off among competing objectives that conventional AO and fixed-trajectory schemes fail to attain.

4.4. Convergence and Complexity Analysis

We first examine the convergence behavior of the inner-layer RRM solution (Algorithm A1). Recall from Section 3.1.1 that the optimal bandwidth allocation is obtained by solving the KKT conditions in (29)–(31) via bisection over λ, where the inner fixed-point Equation (29) for each user converges within a few steps. Figure 8 plots the instantaneous weighted

transmit power at a single time slot as a function of the number of KKT iterations for a representative deployment with $I = 20$. The objective value decreases rapidly from 111.83 mW to 1.20 mW within the first 10 iterations—a reduction of nearly two orders of magnitude—and stabilizes at 0.92 mW after about 15 iterations. This confirms the analytical claim in Section 3.1.1 that the KKT-based solution converges efficiently within a small number of iterations to practical numerical precision. The linear convergence pattern observed in the logarithmic-scale plot is consistent with the bisection search over λ, while the fast convergence is further supported by the convexity of problem (26), which ensures that the KKT conditions yield the global optimum.

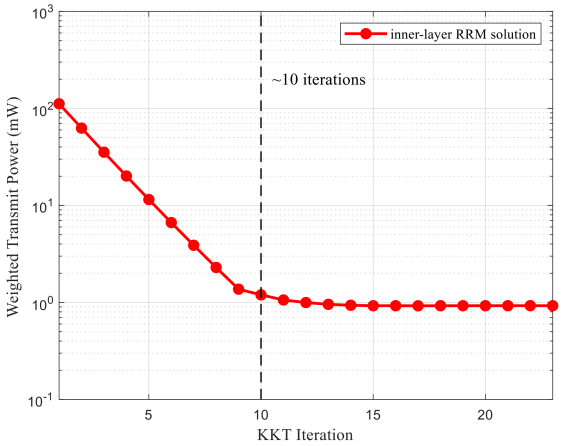

Figure 8. Convergence of the inner-layer RRM solution with respect to KKT iterations for a single representative time slot, measured in instantaneous weighted transmit power (mW).

We now compare the computational complexity of the four schemes. Figure 9 presents the runtime of Proposed-GA, Proposed-DFS, GA-AO, and Fixed-TP as a function of the number of devices I, with the vertical axis on a logarithmic scale to accommodate the wide range of runtime values.

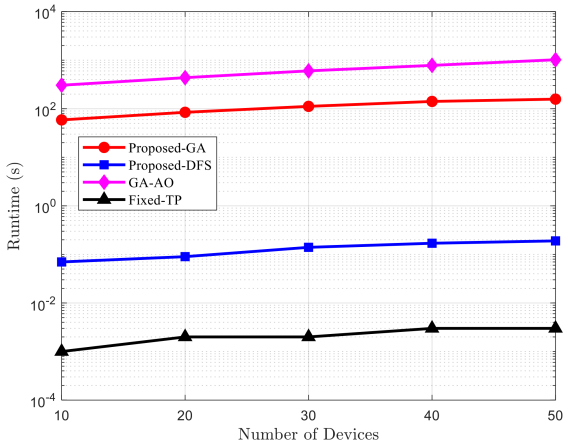

Figure 9. Runtime vs. number of devices.

Several observations can be made. First, Fixed-TP incurs the lowest runtime among all schemes, requiring less than 0.01 s regardless of the network size, as it only performs RRM optimization without TP.

Second, Proposed-DFS exhibits remarkably low runtime, consistently below 0.2 s, due to its deterministic rolling-horizon search with a fixed depth of $n = 3$. This makes it highly attractive for deployment scenarios where computational resources are severely limited.

Third, Proposed-GA achieves a runtime of approximately 60–160 s, which is significantly lower than that of GA-AO across all network sizes. For instance, at $I = 50$, Proposed-GA requires 157.7 s, whereas GA-AO takes 1020.5 s—over six times longer. This stark contrast confirms the theoretical expectation that the non-alternating structure of Proposed-GA avoids the repeated RRM re-optimizations required by the alternating updates in GA-AO, leading to substantially lower computational overhead.

However, despite the superiority of Proposed-GA over GA-AO, among the proposed schemes, Proposed-DFS offers the best trade-off for practical deployment. With runtime below 0.2 s—three orders of magnitude faster than Proposed-GA—while still delivering superior average weighted transmit power performance (as shown in Figure 5), Proposed-DFS is particularly suitable for large-scale or time-critical IoT data collection scenarios where real-time decision-making is essential, and the modest sacrifice in service fairness is acceptable.

In summary, the runtime results confirm the complexity analysis in Section 3.2.2: Proposed-GA provides a favorable balance between solution quality and computational cost relative to GA-AO, while Proposed-DFS stands out as the most practical choice for deployment due to its ultra-low latency and deterministic behavior.

5. Conclusions

This paper has proposed a two-layer non-AO framework for UAV-assisted uplink IoT data collection, jointly optimizing trajectory, user association, bandwidth, and power control to minimize the weighted transmit power of the ground devices with energy-aware dynamic weights. Unlike conventional AO that suffers from initialization sensitivity and high computational overhead, our framework decouples per-slot RRM and multi-slot TP: the inner layer jointly optimizes user association, bandwidth, and power allocation via a KKT-based analytical solution with provable optimality for fixed association and near-optimal user association, achieving polynomial-time complexity ($O(I^3)$ worst-case and $O(I^2)$ in practice), while the outer layer reformulates TP as an MDP whose state explicitly tracks device residual energies, capturing the temporal coupling induced by dynamic weights without cross-layer iterations.

Simulation results demonstrated that the proposed framework consistently outperforms AO baselines across various network configurations, achieving robust performance regardless of the initial trajectory—with a standard deviation of 7.27 mW versus 36.37 mW for GA-AO—and reducing computational time by over sixfold for $I = 50$. The dynamic weighting mechanism effectively balances energy consumption, as reflected by lower residual energy variance. These findings confirm that eliminating alternating updates effectively mitigates the initialization trap inherent in prior works, while the MDP-based formulation offers a principled way to handle sequential dependencies that static-weighting schemes fail to address.

The framework supports flexible solvers, such as the GA and DFS, which offer distinct optimality–complexity trade-offs and make the framework adaptable to different deployment requirements; in particular, the limited-depth DFS variant is attractive for latency-sensitive scenarios. Overall, this work provides a scalable and robust paradigm for energy-efficient UAV-assisted data collection in dynamic IoT environments.

Although the proposed framework achieves promising performance, several limitations should be acknowledged. First, this work focuses on a single UAV, and the simulation adopts a fixed-altitude setting for simplicity; although the proposed formulation supports altitude variation, a systematic evaluation of three-dimensional trajectory and altitude optimization is not performed. Multi-UAV cooperation and inter-UAV coordination are also not addressed. Second, the framework assumes perfect channel state information, static ground

devices, and reliable residual-energy feedback; channel estimation errors, device mobility, and feedback delay may degrade performance in practice. Third, the energy-aware weights are designed as a heuristic inverse function of residual energy, and their optimality in terms of long-term network lifetime and fairness has not been theoretically established. Fourth, the greedy user association achieves near-optimal performance in simulations but lacks a rigorous approximation-ratio guarantee. Fifth, the MDP state space grows exponentially with the number of devices, which limits scalability; the GA is computationally expensive for large networks, DFS with limited depth cannot guarantee global optimality, and the DQN-based solver has not been fully evaluated. Finally, the propulsion energy model and discrete waypoint navigation are approximations of real flight dynamics. Future work will address multi-UAV extensions, altitude optimization, robust optimization under imperfect CSI, distributed implementation and DQN-based learning, theoretical analysis of the greedy association algorithm, and hardware/field experiments.

Author Contributions: Y.Y. provided the idea, designed the architecture, and wrote the paper. X.T. performed simulations. G.X. analyzed the experimental results and proofread the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by Hubei Provincial Natural Science Foundation of China under Grant 2025AFC093; the 2025 Yichang Natural Science Research Project under Grant A25-3-013; the Research Platform Construction Project of Hubei Three Gorges Polytechnic under Grant 2024KYPT0201.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IoT	Internet of Things
UAV	Unmanned aerial vehicle
LoS	Line-of-sight
QoS	Quality-of-service
RRM	Radio resource management
TP	Trajectory planning
AO	Alternating optimization
SCA	Successive convex approximation
BCD	Block coordinate descent
RIS	Reconfigurable intelligent surface
NOMA	Non-orthogonal multiple access
AoI	Age of Information
PF	Proportional fairness
MDP	Markov decision process
KKT	Karush–Kuhn–Tucker
EE	Energy efficiency
WSEE	Weighted sum energy efficiency
MEC	Mobile edge computing
PSD	Power spectral density

GA	Genetic algorithm
DFS	Depth-first search
DQN	Deep Q-network
NLoS	Non-line-of-sight
AWGN	Additive white Gaussian noise

Appendix A

Proof of Proposition 1. Let $x = r_i/\beta_i$. From (24), the total transmit power becomes $\rho_i\beta_i = \frac{N_0 r_i}{g_i} \cdot \frac{2^x-1}{x}$. Define $\phi(x) = \frac{2^x-1}{x}$. Its derivative is $\phi'(x) = \frac{x\cdot 2^x \ln 2-(2^x-1)}{x^2}$. Let $h(x) = x \cdot 2^x \ln 2 - 2^x + 1$. Since $h'(x) = x \cdot 2^x (\ln 2)^2 > 0$ for $x > 0$, $h(x)$ is strictly increasing on $(0, \infty)$. Moreover, $\lim_{x\to 0^+} h(x) = 0$, so $h(x) > 0$ and thus $\phi'(x) > 0$ for all $x > 0$. Hence, $\phi(x)$ is strictly increasing in x. As β_i increases, $x = r_i/\beta_i$ decreases, so $\phi(x)$ decreases. Therefore, $\rho_i\beta_i$ strictly decreases with β_i. □

Appendix B

This appendix presents the detailed pseudo-codes of the proposed algorithms. Algorithm A1 describes the incremental greedy user association for the inner-layer RRM, and Algorithm A2 summarizes the overall two-layer non-alternating framework.

Algorithm A1. Incremental Greedy User Association for Weighted Sum Power Minimization

Input: $\mathcal{I}$, $\{w_i\}$, $\{r_i\}$, $\{g_i\}$, B, $P_{\max}$, N_0, $\{d_i\}$
Output: Near-optimal user association set $\mathcal{U}^*$, bandwidth allocation $\boldsymbol{\beta}^*$, PSD allocation $\boldsymbol{\rho}^*$

1. Initialize $\mathcal{U}_{\text{current}} \leftarrow \emptyset$
2. **while** true **do**
3. $\quad \mathcal{U}_{\text{next}} \leftarrow \mathcal{U}_{\text{current}}$
4. $\quad \Delta_{\min} \leftarrow \infty$, $i_{\text{best}} \leftarrow$ null
5. $\quad$ **for** each $i \in \mathcal{I} \backslash \mathcal{U}_{\text{current}}$ with $d_i = 1$ **do**
6. $\quad\quad \mathcal{U}_{\text{candidate}} \leftarrow \mathcal{U}_{\text{current}} \cup \{i\}$
7. $\quad\quad$ **if** $\sum_{j\in\mathcal{U}_{\text{candidate}}} \beta_j^{\min} \le B$ **then** ▷ Fast pruning via (33)
8. $\quad\quad\quad$ Solve bandwidth and power allocation for $\mathcal{U}_{\text{candidate}}$ via (30), (31) and (24)
9. $\quad\quad\quad$ **if** a solution exists **then**
10. $\quad\quad\quad\quad \Delta \leftarrow C(\mathcal{U}_{\text{candidate}}) - C(\mathcal{U}_{\text{current}})$ ▷ Marginal cost increase
11. $\quad\quad\quad\quad$ **if** $\Delta < \Delta_{\min}$ **then**
12. $\quad\quad\quad\quad\quad \Delta_{\min} \leftarrow \Delta$, $i_{\text{best}} \leftarrow i$
13. $\quad\quad\quad\quad$ **end if**
14. $\quad\quad\quad$ **end if**
15. $\quad\quad$ **end if**
16. $\quad$ **end for**
17. $\quad$ **if** $i_{\text{best}} =$ null **then**
18. $\quad\quad$ **break**
19. $\quad$ **end if**
20. $\quad \mathcal{U}_{\text{next}} \leftarrow \mathcal{U}_{\text{current}} \cup \{i_{\text{best}}\}$
21. $\quad \mathcal{U}_{\text{current}} \leftarrow \mathcal{U}_{\text{next}}$
22. **end while**
23. $\mathcal{U}^* \leftarrow \mathcal{U}_{\text{current}}$, compute final $\boldsymbol{\beta}^*$ and $\boldsymbol{\rho}^*$ via (26) and (24)
24. **return** $\mathcal{U}^*, \boldsymbol{\beta}^*, \boldsymbol{\rho}^*$

Algorithm A2. Overall Non-AO Framework

Input: grid $\mathcal{Q}$, device parameters ($\{\mathbf{q}_i\}$, $\{r_i\}$, $\{E_i^{\text{init}}\}$, $\{s_i\}$, $\{e_i\}$), UAV parameters (v, E_{battery}, $P_{\max}$, B, N_0)

Output: Trajectory $\mathbf{Q} = [\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(T)}]$, corresponding RRM decisions $\{\boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t)}, \boldsymbol{\rho}^{(t)}\}_{t=1}^{T}$

1. Initialize state: reset to initial state $s^{(1)}$ via (36), set $R_{\text{total}} \leftarrow 0$
2. **for** $t = 1$ to T **do**
3. Take action $\mathbf{q}^{(t)}$ (selected by the MDP solver from $S\left(\mathbf{q}^{(t-1)}\right)$)
4. Compute propulsion power $P_{\text{prop}}^{(t)}$ via (17)
5. Compute current energy-aware weights $\left\{w_i^{(t)}\right\}$ via (22)
6. Run **Algorithm A1** at position $\mathbf{q}^{(t)}$
7. Obtain $\left\{\alpha_i^{(t)}\right\}$, $\left\{\beta_i^{(t)}\right\}$, $\left\{\rho_i^{(t)}\right\}$
8. Update state: $E_{\text{UAV}}^{\text{res}} \leftarrow E_{\text{UAV}}^{\text{res}} - P_{\text{prop}}^{(t)}\Delta T$, $E_i^{\text{res}} \leftarrow E_i^{\text{res}} - \rho_i^{(t)}\beta_i^{(t)}\Delta T$, $m_i \leftarrow \max\left(m_i, \alpha_i^{(t)}\right)$, and update $d_i^{(t+1)}$ according to (4)
9. Compute step reward $r^{(t)}$ via (38)
10. $R_{\text{total}} \leftarrow R_{\text{total}} + r^{(t)}$
11. **end for**
12. **return** $\mathbf{Q}$, $\{\boldsymbol{\alpha}^{(t)}\}$, $\{\boldsymbol{\beta}^{(t)}\}$, $\{\boldsymbol{\rho}^{(t)}\}$

References

1. Santoso, F. AI-Driven Energy-Efficient UAV-Assisted 6G Communications: A Comprehensive Survey and Research Roadmap. *IEEE Open J. Commun. Soc.* **2026**, *7*, 2838–2875. [CrossRef]
2. Meng, X.; Wu, X.; Xie, Z.; Zhang, T.; Xu, T. Secure Resource Allocation and Trajectory Design for UAV-Assisted IoT with Double Cluster Head. *IEEE Trans. Green Commun. Netw.* **2024**, *8*, 1661–1675. [CrossRef]
3. Akbari, M.; Syed, A.; Kennedy, W.S.; Erol-Kantarci, M. AoI-Aware Energy-Efficient SFC in UAV-Aided Smart Agriculture Using Asynchronous Federated Learning. *IEEE Open J. Commun. Soc.* **2024**, *5*, 1222–1242. [CrossRef]
4. Wang, C.; Zhang, L.; Luo, J.; Dai, S. QoS-Aware Deployment Optimization for Capsule Airport–UAV Emergency Communication Networks. *Drones* **2026**, *10*, 544. [CrossRef]
5. Van Mulders, J.; Cappelle, J.; Goossens, S.; De Strycker, L.; Van der Perre, L. UAV-Based Servicing of IoT Nodes: Assessment of Ecological Impact. *Sensors* **2023**, *23*, 2291. [CrossRef]
6. Zeng, Y.; Zhang, R.; Lim, T.J. Wireless communications with unmanned aerial vehicles: Opportunities and challenges. *IEEE Commun. Mag.* **2016**, *54*, 36–42. [CrossRef]
7. Zeng, Y.; Wu, Q.; Zhang, R. Accessing from the Sky: A Tutorial on UAV Communications for 5G and Beyond. *Proc. IEEE* **2019**, *107*, 2327–2375. [CrossRef]
8. Mozaffari, M.; Saad, W.; Bennis, M.; Nam, Y.-H.; Debbah, M. A Tutorial on UAVs for Wireless Networks: Applications, Challenges, and Open Problems. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 2334–2360. [CrossRef]
9. Liu, A.; Lau, V.K.N.; Kananian, B. Stochastic Successive Convex Approximation for Non-Convex Constrained Stochastic Optimization. *IEEE Trans. Signal Process.* **2019**, *67*, 4189–4203. [CrossRef]
10. Zhang, S.; Ansari, N. 3D drone base station placement and resource allocation with FSO-based backhaul in hotspots. *IEEE Trans. Veh. Technol.* **2020**, *69*, 3322–3329. [CrossRef]
11. Liu, J.; Xu, Z.; Wen, Z. Joint data transmission and trajectory optimization in UAV-enabled wireless powered mobile edge learning systems. *IEEE Trans. Veh. Technol.* **2023**, *72*, 11617–11630. [CrossRef]
12. Du, W.; Wang, T.; Zhang, H.; Dong, Y.; Li, Y. Joint resource allocation and trajectory optimization for completion time minimization for energy-constrained UAV communications. *IEEE Trans. Veh. Technol.* **2023**, *72*, 4568–4579. [CrossRef]
13. Lu, W.; Ding, Y.; Gao, Y.; Hu, S.; Wu, Y.; Zhao, N.; Gong, Y. Resource and trajectory optimization for secure communications in dual unmanned aerial vehicle mobile edge computing systems. *IEEE Trans. Ind. Informat.* **2022**, *18*, 2704–2713. [CrossRef]
14. Yang, H.; Ruby, R.; Pham, Q.-V.; Wu, K. Aiding a disaster spot via multi-UAV-based IoT networks: Energy and mission completion time-aware trajectory optimization. *IEEE Internet Things J.* **2022**, *9*, 5853–5867. [CrossRef]
15. Ren, Y.; Chi, K.; Su, T.; Yang, Y.; Gao, J. BLLL-based distributed alternating optimization for joint 3D UAV positioning and channel allocation in dense urban scenarios. *IEEE Access* **2025**, *13*, 217660–217674. [CrossRef]

16. Lahmeri, M.-A.; Mustieles-Pérez, V.; Vossiek, M.; Krieger, G.; Schober, R. UAV formation and resource allocation optimization for communication-assisted 3D InSAR sensing. *IEEE Trans. Commun.* **2025**, *73*, 5788–5804. [CrossRef]
17. Liu, P.; Si, J.; Li, Z.; Al-Dhahir, N.; Gao, Y. Joint 3-D trajectory and power optimization for dual-UAV-assisted short-packet covert communications. *IEEE Internet Things J.* **2024**, *11*, 17388–17401. [CrossRef]
18. Chen, J.; Zheng, K.; Jia, J.; Deng, Y.; Wang, X. Secure resource allocation and trajectory design for RIS and NOMA assisted multi-UAV systems. *IEEE Internet Things J.* **2025**, *12*, 36436–36450. [CrossRef]
19. Li, S.; Gong, P.; Li, C.; Kim, D.K.; Wu, D.O.; Gao, X. Interference-constrained integrated sensing, communication, and computation for clustered UAV networks: Hierarchical alternating optimization for the delay-value trade-off. *Chin. J. Electron.* **2026**, *35*, 978–995. [CrossRef]
20. Hu, H.; Xiong, K.; Qu, G.; Ni, Q.; Fan, P.; Letaief, K.B. AoI-minimal trajectory planning and data collection in UAV-assisted wireless powered IoT networks. *IEEE Internet Things J.* **2021**, *8*, 1211–1223. [CrossRef]
21. Tseng, P. Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization. *J. Optim. Theory Appl.* **2001**, *109*, 475–494. [CrossRef]
22. Razaviyayn, M.; Hong, M.; Luo, Z.-Q. A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM J. Optim.* **2013**, *23*, 1126–1153. [CrossRef]
23. Xu, Y.; Yin, W. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM J. Imaging Sci.* **2013**, *6*, 1758–1789. [CrossRef]
24. Lyu, H.; Jang, J.; Lee, H.; Yang, H.J. Non-Iterative Optimization of Trajectory and Radio Resource for Aerial Network. *IEEE Trans. Wirel. Commun.* **2025**, *24*, 1555–1567. [CrossRef]
25. Yu, Y.; Tang, X. Weighted Sum Power Minimization for Cooperative Spectrum Sharing in Cognitive Radio Networks. *IEEE Commun. Lett.* **2025**, *29*, 160–164. [CrossRef]
26. Yu, Y.; Tang, X.; Xie, G. Geometric Weighted Sum Power Minimization for Energy-Efficient D2D Underlay Networks. *IEEE Commun. Lett.* **2025**, *29*, 2909–2913. [CrossRef]
27. Du, Y.; Guo, Y.; Hao, J.; Zhu, H. Residual energy maximization for NOMA-enabled UAV-assisted data collection network: Trajectory optimization and resource allocation. *J. Wirel. Com. Netw.* **2023**, *2023*, 92.
28. Al-Bakhrani, A.A.; Li, M.; Obaidat, M.S.; Alotaibi, M.; Amran, G.A. ERA-UAV-MEC: Energy-resilient adaptive framework for sustainable UAV-assisted mobile edge computing. *Comput. Netw.* **2026**, *282*, 112274. [CrossRef]
29. Wei, Z.; Zhu, M.; Zhang, N.; Wang, L.; Zou, Y.; Meng, Z.; Wu, H.; Feng, Z. UAV-Assisted Data Collection for Internet of Things: A Survey. *IEEE Internet Things J.* **2022**, *9*, 15460–15483. [CrossRef]
30. Ostroumov, I.; Kuzmenko, N. Statistical Analysis and Flight Route Extraction from Automatic Dependent Surveillance-Broadcast Data. In Proceedings of the 2022 Integrated Communication, Navigation and Surveillance Conference (ICNS), Dulles, VA, USA, 5–7 April 2022.
31. Meier, L.; Honegger, D.; Pollefeys, M. PX4: A node-based multithreaded open source robotics framework for deeply embedded platforms. In Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015.
32. Al-Hourani, A.; Kandeepan, S.; Lardner, S. Optimal LAP Altitude for Maximum Coverage. *IEEE Wirel. Commun. Lett.* **2014**, *3*, 569–572. [CrossRef]
33. Holis, J.; Pechac, P. Elevation Dependent Shadowing Model for Mobile Communications via High Altitude Platforms in Built-Up Areas. *IEEE Trans. Antennas Propag.* **2008**, *56*, 1078–1084. [CrossRef]
34. Al-Hourani, A.; Kandeepan, S.; Jamalipour, A. Modeling air-to-ground path loss for low altitude platforms in urban environments. In Proceedings of the 2014 IEEE Global Communications Conference, Austin, TX, USA, 8–12 December 2014.
35. Zeng, Y.; Xu, J.; Zhang, R. Energy Minimization for Wireless Communication with Rotary-Wing UAV. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 2329–2345. [CrossRef]
36. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
37. Korte, B.; Vygen, J. *Combinatorial Optimization: Theory and Algorithms*, 6th ed.; Springer: Berlin, Germany, 2018.
38. Amodu, O.A.; Mahmood, R.A.R.; Althumali, H.; Bukar, U.A.; Abdullah, N.F.; Jarray, C. *Markov Decision Processes and Reinforcement Learning for Timely UAV-IoT Data Collection Applications*; Springer: Cham, Switzerland, 2025.
39. Goldberg, D.E. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley: Reading, MA, USA, 1989.
40. Cormen, T.H.; Leiserson, C.E.; Rivest, R.L.; Stein, C. *Introduction to Algorithms*, 3rd ed.; MIT Press: Cambridge, MA, USA, 2009.
41. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.