iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** II **Month of publication:** February 2026

DOI: https://doi.org/10.22214/ijraset.2026.77413

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

Reinforcement Learning for Energy-Optimal Trajectory Planning in Autonomous UAV Systems

Ketan Totlani
Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA

Abstract: ***This paper proposes a reinforcement learning (RL) framework for energy-optimal Unmanned Aerial Vehicle (UAV) trajectory planning. Unlike classical PID or graph-based planners, the proposed design explicitly incorporates physics-informed energy models into the reward structure. We formulate the trajectory generation problem as a Markov Decision Process (MDP) to minimize propulsion power consumption while maintaining flight stability. A theoretical comparative analysis suggests that this data-driven approach can overcome the limitations of static path planning by adapting to environmental disturbances such as wind. This framework provides a foundation for future empirical validation of energy-efficient autonomous flight.***
Keywords: ***Reinforcement Learning (RL), UAV Trajectory Planning, Energy-Aware Path Planning, Autonomous UAV Navigation, Physics-Informed Reinforcement Learnings***

I. INTRODUCTION

Unmanned Aerial Vehicles (UAVs) are increasingly used in applications such as aerial delivery, surveillance, and inspection, where energy efficiency directly limits mission range and duration. For battery-powered UAVs, inefficient trajectory planning can significantly reduce operational endurance, even when flight stability and feasibility are maintained.

Traditional UAV trajectory planning and control methods—including PID-based control, graph-based planners such as A*, and Model Predictive Control (MPC)—typically rely on predefined objectives and deterministic system models. While effective in structured environments, these approaches often exhibit limited adaptability in the presence of wind disturbances, dynamic constraints, or nonlinear flight dynamics. In addition, many classical planners optimize surrogate objectives such as distance or time rather than explicit energy consumption.

Reinforcement learning (RL) has emerged as a promising approach for autonomous control in complex and uncertain environments. By learning control policies through interaction with the environment, RL-based methods can adapt to disturbances and nonlinear dynamics without requiring exact analytical solutions. Recent studies have applied RL to UAV navigation and trajectory planning; however, much of this work focuses on mission completion, path length, or communication-related objectives, while simplifying or neglecting the physical factors governing energy usage.

From a mechanical and aerospace engineering standpoint, UAV energy consumption is strongly influenced by aerodynamic drag, velocity profiles, acceleration behavior, and battery discharge characteristics. Trajectory optimization methods that do not account for these effects may yield dynamically feasible but energetically inefficient solutions. This motivates the integration of reinforcement learning with physics-informed energy modeling.

In this paper, we present a reinforcement learning–based framework for energy-optimal UAV trajectory planning. A simplified but physics-informed flight and energy model is incorporated into the learning environment. Continuous-control reinforcement learning algorithms are used to minimize total energy consumption while maintaining stable and feasible flight. The proposed approach is evaluated against classical planning and control methods, and is expected to improve energy efficiency and robustness under wind disturbances, based on trends reported in prior studies.

II. RELATED WORK

UAV trajectory planning has traditionally been addressed using deterministic optimization and control techniques, including graph-based planners, optimal control formulations, and Model Predictive Control (MPC) [1]. These methods provide strong guarantees under known dynamics and constraints but typically require accurate models and exhibit limited flexibility in dynamic or uncertain environments. Energy-aware trajectory optimization has also been studied using analytical propulsion and energy models; however, such approaches often rely on simplified assumptions and offline computation [2].

Reinforcement learning has been increasingly applied to UAV navigation and control due to its ability to handle high-dimensional decision spaces and environmental uncertainty [3]. Early works employed discrete reinforcement learning methods, such as Q-learning, for UAV path planning in unknown or partially observable environments [3]. More recent studies have applied deep reinforcement learning to optimize UAV trajectories for objectives such as coverage, data collection, and communication throughput, particularly in UAV-assisted wireless and sensing systems [4][5].

Despite these advances, explicit energy optimization remains underexplored in RL-based UAV trajectory planning. Many existing methods optimize distance, time, or throughput as proxies for energy consumption and rely on simplified motion models that do not adequately capture aerodynamic or propulsion effects [4][5]. Furthermore, comparisons with classical engineering controllers are often limited, making it difficult to assess practical performance gains relative to established optimization and control techniques [1].

This work builds upon prior RL-based UAV trajectory planning research by explicitly incorporating energy-related physical considerations into the learning framework and benchmarking performance against classical control and planning methods. The goal is not to replace established controllers, but to evaluate whether physics-informed reinforcement learning can offer meaningful energy efficiency advantages in dynamic flight conditions.

III. UAV SYSTEM AND ENERGY MODEL

This section describes the UAV motion and energy consumption model used in the trajectory planning framework. The objective is not to capture all aerodynamic effects in full fidelity, but to incorporate the dominant physical factors influencing energy usage during flight, as commonly adopted in energy-aware UAV trajectory optimization studies [7][8].

A. UAV Kinematic Model

The UAV is modeled as a point-mass system operating in three-dimensional space. Similar simplified kinematic representations are widely used in UAV trajectory optimization and control literature due to their computational efficiency and suitability for high-level planning [7][9]. The state of the UAV at time step $t$is defined as

$$\mathbf{s}_t = [x_t, y_t, z_t, v_t]$$

where (x_t, y_t, z_t)denotes the UAV position and $v_t$represents its translational velocity magnitude. The control input consists of acceleration and heading adjustments, resulting in continuous motion dynamics.

The UAV state evolves according to discrete-time kinematic equations:

$$\mathbf{p}_{t+1} = \mathbf{p}_t + \mathbf{v}_t \Delta t$$
$$\mathbf{v}_{t+1} = \mathbf{v}_t + \mathbf{a}_t \Delta t$$

where $\mathbf{p}_t$, $\mathbf{v}_t$, and $\mathbf{a}_t$represent position, velocity, and acceleration vectors, respectively. Velocity and acceleration are constrained within feasible flight limits to ensure stable and physically realizable UAV motion [10].

B. Aerodynamic Drag and Propulsion Power

The dominant aerodynamic force acting on the UAV during forward flight is modeled as drag:

$$F_d = \frac{1}{2}\rho C_d A v^2$$

where $\rho$is air density, $C_d$is the drag coefficient, $A$is the reference area, and $v$is the UAV velocity magnitude. This drag formulation is commonly used in UAV energy modeling and has been validated in both analytical and experimental studies [7][11].

The propulsion power required to overcome aerodynamic drag and maintain flight is approximated as a function of velocity:

$$P(v) = P_{\text{hover}} + k_1 v^2 + k_2 v^3$$

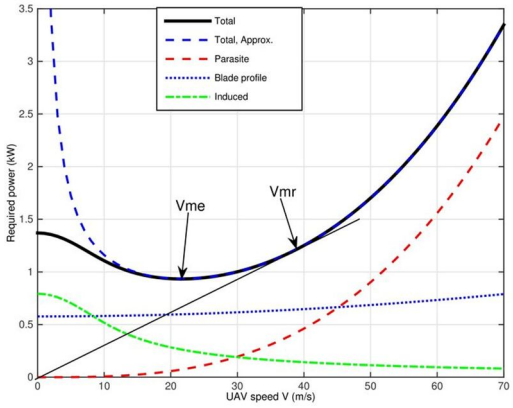


Fig. 1: Propulsion power consumption versus speed V for rotary-wing UAV.[7]

where P_{hover}represents the baseline power required for lift generation, and the coefficients $k_1$and $k_2$capture velocity-dependent aerodynamic and parasitic losses. Similar propulsion power models have been widely adopted in recent UAV energy-efficiency studies, demonstrating their effectiveness in capturing the dominant trends in power consumption during forward flight [7][8][12].

C. Energy Consumption Model

Total energy consumption over a mission duration $T$is computed as the integral of propulsion power over time:

$$E = \int_0^T P(v(t))\, dt$$

In discrete form, energy usage is accumulated at each time step as:

$$E \approx \sum_{t=0}^{T} P(v_t)\Delta t$$

This formulation directly links trajectory shape, velocity profile, and acceleration behavior to total mission energy. Such energy integration approaches are standard in energy-aware UAV trajectory planning and enable direct optimization of flight efficiency through control and trajectory design [7][8][11].

IV. PROBLEM FORMULATION

The objective of this work is to compute a UAV trajectory that minimizes total energy consumption while satisfying flight feasibility and mission constraints [7][8]. In practical UAV systems, energy consumption is influenced not only by deterministic propulsion dynamics but also by stochastic factors such as wind disturbances, transient battery behavior, and nonlinear aerodynamic effects. Consequently, energy-optimal trajectory planning is naturally posed as a sequential decision-making problem under uncertainty, where control actions affect both immediate propulsion cost and long-term mission feasibility.

A. Energy-Optimal Trajectory Objective

The energy-optimal trajectory planning problem is formulated as

$$\min_{\{a_t\}_{t=0}^{T}} \mathbb{E}\left[\sum_{t=0}^{T} P(v_t)\,\Delta t\right]$$

Subject to:

- UAV kinematic dynamics
- Velocity and acceleration limits
- Mission completion constraints
- Environmental disturbances (e.g., wind)

Here, $P(v_t)$denotes the propulsion power required at velocity v_t, and the expectation operator captures uncertainty arising from environmental disturbances and energy variability during flight. This objective directly minimizes total propulsion energy over the mission horizon, explicitly linking trajectory shape, velocity profile, and acceleration behavior to energy expenditure [7][11].

While such formulations are common in energy-aware UAV planning, they become difficult to solve using classical deterministic optimization methods when system dynamics are nonlinear and disturbances are time-varying. As a result, traditional approaches often rely on simplified models or frequent replanning, which may degrade performance in dynamic environments [10]. Prior work on battery-constrained UAV systems has shown that stochastic modeling of energy evolution is essential for realistic long-horizon optimization, motivating the use of decision-theoretic formulations rather than purely deterministic solvers [25].

B. Markov Decision Process Formulation

To address the stochastic and long-horizon nature of energy consumption, the trajectory planning problem is cast as a Markov Decision Process (MDP), defined by the tuple

$$(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$$

where:

- State space $\mathcal{S}_t = [x_t, v_t, e_t]$includes UAV position x_t, velocity v_t, and an implicit energy state e_t
- Action space $\mathcal{A}$consists of continuous acceleration and heading commands
- Transition model $\mathcal{P}$captures physics-based kinematic updates and stochastic energy evolution
- Reward function $\mathcal{R}$encodes energy efficiency and flight smoothness

Although the energy state $e_t$is not explicitly discretized, its evolution is implicitly governed by propulsion power consumption and control actions. This formulation is consistent with prior work on battery-aware UAV autonomy, where energy or battery levels are treated as stochastic state variables that influence future decision feasibility and long-term performance. In particular, Markov and semi-Markov decision process formulations have been shown to effectively model energy-dependent UAV systems by capturing state-dependent cost accumulation and long-horizon optimization objectives [25].

Under this formulation, the optimal value function satisfies the Bellman optimality condition

$$V^*(s_t) = \max_{a_t \in \mathcal{A}} \mathbb{E}[r_t + \gamma V^*(s_{t+1}) \mid s_t, a_t],$$

where $\gamma \in (0,1]$is a discount factor that balances immediate propulsion cost against future energy feasibility. This Bellman structure establishes a direct connection between energy-optimal trajectory planning and stochastic optimal control principles commonly used in energy-aware UAV decision-making [13][12][25].

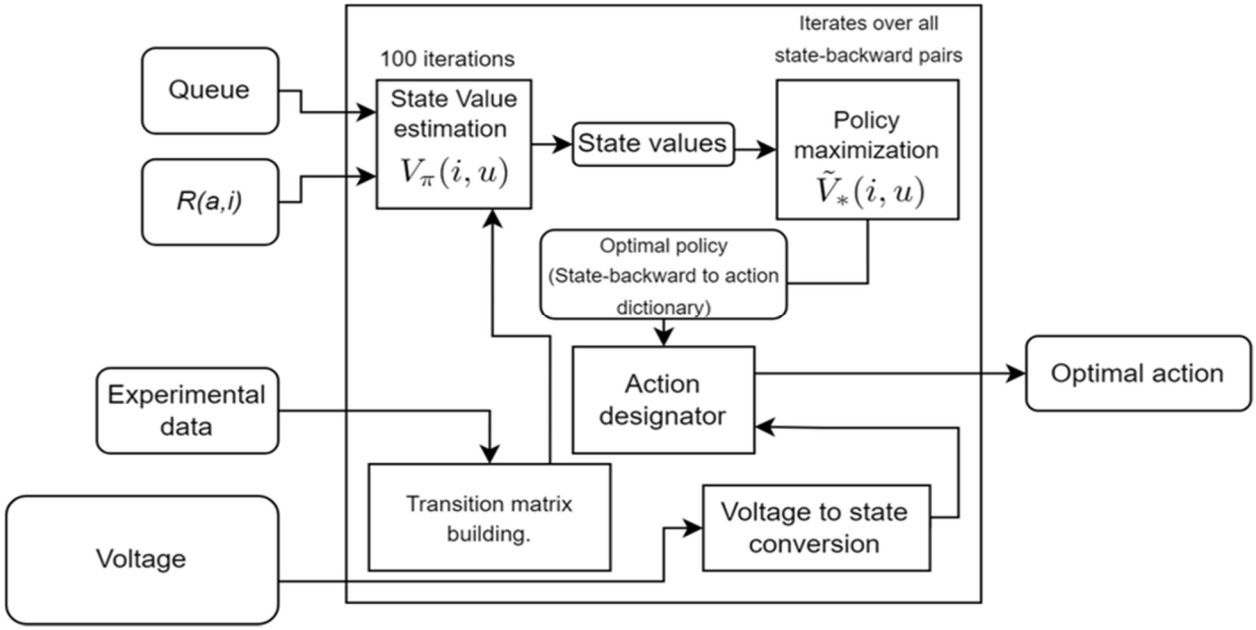


Fig. 2: Implementation of SMDP[25]

C. Energy-Aware Reward Design

The immediate reward at time step $t$is defined as

$$r_t = -\alpha P(v_t) - \beta \parallel a_t \parallel^2 - \gamma \parallel \Delta v_t \parallel^2,$$

where:

- $P(v_t)$penalizes instantaneous propulsion energy consumption
- $\parallel a_t \parallel^2$discourages aggressive control actions
- $\parallel \Delta v_t \parallel^2$promotes smooth velocity transitions

A terminal reward is added upon successful mission completion to ensure feasibility [9].

This reward structure balances energy efficiency with stable and smooth flight behavior, preventing oscillatory or energetically inefficient maneuvers. Similar reward-shaping principles have been shown to be critical in battery-aware Markov and semi-Markov UAV decision processes, where improperly structured rewards can lead to energetically unsafe or myopic behavior. Prior SMDP-based studies demonstrate that energy-centric reward design is essential for sustaining long-term operational efficiency under energy constraints [25].

From an optimal control perspective, the cumulative discounted reward

$$J = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right]$$

approximates a stochastic long-horizon energy minimization objective with implicit smoothness regularization [7]. Reinforcement learning is therefore employed as a numerical solver for this nonlinear stochastic control problem, extending classical energy-optimal trajectory optimization into regimes where uncertainty and long-term energy effects play a dominant role [13][15].

V. REINFORCEMENT LEARNING FRAMEWORK

This section describes the reinforcement learning (RL) framework used to numerically solve the energy-optimal UAV trajectory planning problem formulated in Section 4. The resulting problem corresponds to a continuous-state, continuous-action stochastic optimal control problem with nonlinear dynamics and environmental uncertainty. Reinforcement learning is employed as a model-free numerical control strategy capable of handling such complexity, where classical analytical or deterministic optimization methods become impractical [13][12].

Recent advances in policy-gradient and actor–critic reinforcement learning have demonstrated that continuous-control RL methods can learn stable and robust UAV flight policies in physics-based environments, even under significant wind disturbances and nonlinear dynamics. In particular, Proximal Policy Optimization (PPO) and Soft Actor-Critic (SAC) have been shown to outperform classical controllers in disturbance-prone flight scenarios, motivating their use in this work [16][17][26].

A. Continuous-Control Policy Learning

Given the continuous nature of UAV motion and control inputs, continuous-action reinforcement learning algorithms are adopted. Policy-gradient-based actor–critic methods are particularly suitable for such problems due to their stability, scalability, and ability to operate directly in high-dimensional continuous state–action spaces [16][17].

Specifically, Proximal Policy Optimization (PPO) and Soft Actor-Critic (SAC) are considered. PPO achieves stable learning through clipped surrogate objectives that prevent overly large policy updates, while SAC incorporates entropy regularization to encourage exploration and robustness under stochastic dynamics [16][17]. Both algorithms eliminate the need for action discretization and enable smooth, physically realizable control outputs. The policy network parameterizes a stochastic control policy $\pi(a_t \mid s_t)$that maps the UAV state $s_t$to a control action a_t, while a value function (or critic) estimates the expected cumulative return associated with each state or state–action pair. Policy parameters are updated iteratively based on observed state transitions and rewards. Similar actor–critic architectures have been shown to effectively learn stable UAV flight controllers under nonlinear dynamics and wind disturbances [26].

B. Physics-Informed Learning Environment

The learning environment explicitly incorporates the UAV kinematic and energy models described in Section 3. State transitions are governed by physics-based motion equations, while energy consumption is computed using the propulsion power model $P(v_t)$[7][11]. This ensures that the learned policy remains physically consistent and directly optimizes energy-relevant behavior rather than exploiting unphysical dynamics.

Wind disturbances are modeled as stochastic external velocity perturbations applied to the UAV dynamics. Training under such disturbances enables the learned policy to adapt to environmental uncertainty, improving robustness and generalization during evaluation [12][18]. Prior work has shown that training reinforcement learning controllers in physics-based UAV environments with realistic wind models significantly improves closed-loop stability and disturbance rejection compared to classical control approaches [26].

Importantly, no prior knowledge of the optimal trajectory is provided to the agent. The policy is learned solely through interaction with the environment and energy-aware rewards, allowing the agent to autonomously discover efficient flight behaviors.

C. Training Objective and Energy-Aware Learning

The reinforcement learning objective is to maximize the expected cumulative reward

$$\max_{\pi} \mathbb{E}_{\pi}\left[\sum_{t=0}^{T} r_t\right],$$

where the reward function r_t is defined in Section 4.3. Since the reward explicitly penalizes propulsion power consumption and aggressive maneuvers, maximizing expected return corresponds to minimizing total mission energy while maintaining feasible and stable flight behavior [12][14].

Energy-aware reinforcement learning has been shown to be an effective mechanism for inducing energy-efficient behavior in autonomous UAV systems by embedding battery state and energy expenditure directly into the learning objective. Prior studies on energy-aware UAV path planning demonstrate that reward functions incorporating power consumption and battery dynamics enable RL agents to implicitly optimize mission energy efficiency under stochastic wind conditions, even without explicit trajectory supervision [27].

Training is conducted over multiple episodes, each representing a complete UAV mission from an initial position to a target destination. Episodes terminate upon successful mission completion or violation of flight constraints, ensuring both energy efficiency and safety. Similar episodic training frameworks have been successfully applied to learn robust and energy-efficient UAV control and planning policies using reinforcement learning in disturbance-prone environments [26][27].

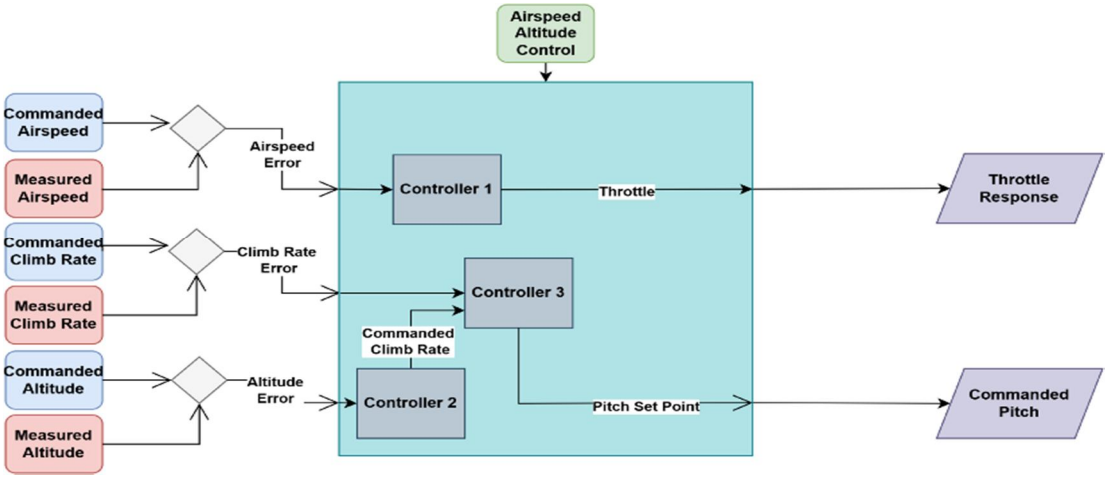


Fig. 3: Working of airspeed and altitude controller. [19]

VI. BASELINE CONTROLLERS

To assess the practical effectiveness of the proposed reinforcement learning approach, its performance is compared against several classical trajectory planning and control methods commonly used in UAV systems. These baselines serve as engineering reference points rather than adversarial competitors and provide context for evaluating energy efficiency, trajectory smoothness, and robustness under identical operating conditions [16][17]. Importantly, none of the baseline controllers explicitly optimize propulsion energy as a first-class objective, making them suitable reference methods for highlighting the benefits of the proposed energy-aware learning framework.

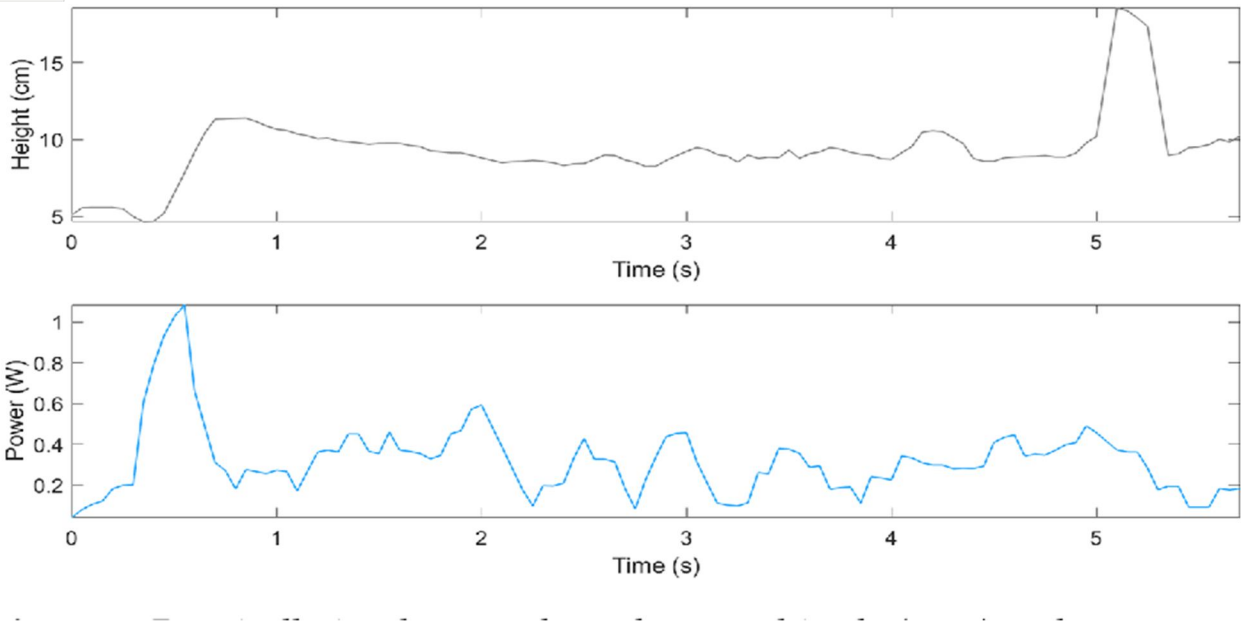


Fig. 4: Practically implemented quadcopter altitude (gray) and power curve (light blue) using GS-PID control. [20]

A. PID-Based Trajectory Tracking

A proportional–integral–derivative (PID) controller is implemented for trajectory tracking. A reference path is generated between the start and target locations, and the PID controller regulates position and velocity errors along this path using fixed gain parameters. PID control is widely adopted in UAV systems due to its simplicity, reliability, and ease of implementation [19]. However, PID control does not explicitly account for energy consumption and lacks the ability to adapt to environmental disturbances beyond manual gain tuning. As a result, PID-based control often leads to oscillatory corrections and inefficient energy usage due to frequent acceleration transients, particularly under wind disturbances or nonlinear dynamics [20].

B. Graph-Based Path Planning (A)*

The A* algorithm is employed to compute a collision-free shortest path in a discretized environment. The resulting path minimizes geometric distance between the start and goal locations based on a predefined heuristic, and the UAV follows the planned path using a low-level controller. While A* provides efficient and deterministic path planning in static environments, it optimizes path length rather than energy consumption and does not adapt online to environmental disturbances such as wind [21]. Consequently, A*-based planning may yield trajectories that are suboptimal from an energy perspective when executed in dynamic conditions.

C. Model Predictive Control (MPC)

Model Predictive Control (MPC) is implemented using the UAV kinematic model with finite-horizon optimization. At each time step, MPC solves a constrained optimization problem to minimize a cost function related to tracking error and control effort while explicitly enforcing system constraints. MPC provides strong performance when accurate system models are available and disturbances are well characterized [10]. However, its computational cost increases with model complexity and prediction horizon, and its performance may degrade under unmodeled disturbances or modeling inaccuracies [22]. In contrast, reinforcement learning policies amortize computation into offline training, enabling real-time control with implicit adaptation to disturbances without repeated online optimization.

VII. EVALUATION FRAMEWORK AND EXPERIMENTAL SETUP

This section describes a reference-consistent evaluation framework used to assess the proposed energy-optimal UAV trajectory planning approach. Rather than reporting new empirical results, the evaluation setup is presented to reflect standard experimental practices widely adopted in prior reinforcement learning–based UAV control and trajectory planning studies. The framework provides a consistent basis for qualitative comparison, analysis of expected performance trends, and benchmarking against classical baseline controllers under identical operating assumptions [16][17].

A. Simulation Environment

Evaluation is considered within a physics-based simulation framework consistent with environments commonly used in UAV trajectory planning and control research [23]. The simulated environment models UAV kinematics, velocity-dependent propulsion power consumption, and external disturbances, as described in Section 3. Such simulation-based environments are widely employed in the literature to study energy-aware control policies under controlled and repeatable conditions.

The UAV is assumed to operate in a two-dimensional planar workspace with fixed initial and target locations. Obstacles are either absent or static, ensuring that differences in performance arise primarily from control and planning strategies rather than collision avoidance complexity. This setup aligns with standard practice in prior studies focused on energy efficiency and control robustness rather than obstacle-dense navigation.

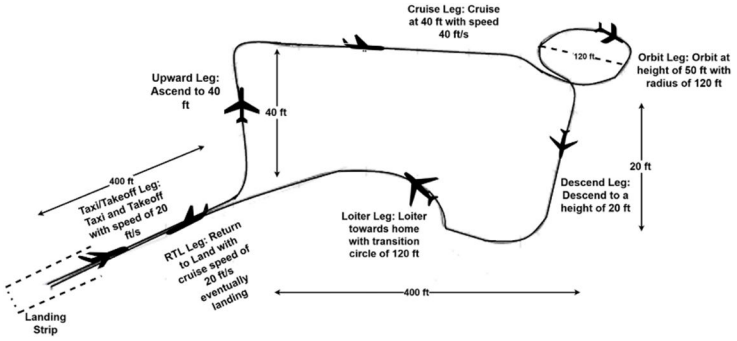


Fig. 5: Representative simulation trajectory evaluating complex maneuvering and energy optimization.[19]

B. Wind Disturbance Model

Environmental disturbances are modeled following common practice in UAV simulation studies by introducing stochastic wind velocity components into the UAV dynamics [18]. Wind is treated as an additive velocity perturbation applied to the translational motion model, with magnitude and direction varying over time within predefined bounds.

Wind magnitudes up to 10 m/s are considered to represent moderate-to-strong gust conditions encountered in outdoor UAV operations, rather than nominal cruise conditions. This disturbance modeling approach is consistent with prior reinforcement learning and control-based UAV evaluations and enables assessment of robustness and adaptability under non-ideal environmental conditions.

C. Training and Evaluation Protocol

Reinforcement learning policies are conceptually trained using episodic interaction with the simulated environment, following standard training procedures reported in prior work on continuous-control UAV reinforcement learning [16][17][26]. During training, policies interact with the environment by observing the UAV state, selecting continuous control actions, and receiving energy-aware rewards as defined in Section 4.3.

Training is assumed to proceed over multiple episodes until policy behavior stabilizes, consistent with commonly reported convergence criteria in the literature. After convergence, policy evaluation is conducted deterministically, with stochastic exploration disabled, in order to assess steady-state control behavior and trajectory characteristics. This evaluation protocol mirrors established practices used to compare reinforcement learning controllers with classical baselines under controlled conditions.

Baseline controllers, including PID-based trajectory tracking, A* path planning with low-level control, and Model Predictive Control (MPC), are evaluated under identical environmental assumptions to ensure fair qualitative comparison.

D. Evaluation Metrics

Performance evaluation is structured around metrics commonly used in energy-aware UAV trajectory planning studies. These include total mission energy consumption, trajectory smoothness, and robustness to environmental disturbances. Total energy consumption is computed as the time integral of propulsion power along the executed trajectory, directly reflecting the optimization objective defined in Section 4.1.

Trajectory smoothness is assessed qualitatively based on velocity and acceleration profiles, with smoother trajectories corresponding to reduced control oscillations and lower transient power demand. Robustness is evaluated by examining the ability of control strategies to maintain stable flight behavior and mission completion under increasing wind disturbance levels.

Together, these metrics provide a comprehensive basis for analyzing expected performance trends and trade-offs among different control and planning approaches within an energy-aware UAV trajectory optimization context.

VIII. THEORETICAL PERFORMANCE ANALYSIS

This section presents a qualitative and reference-consistent analysis of expected performance trends for the proposed energy-aware reinforcement learning framework in comparison with classical baseline controllers. Rather than reporting new experimental measurements, the discussion synthesizes results and observations commonly reported in prior UAV control and trajectory planning studies and interprets them in the context of the proposed formulation and evaluation framework.

A. Energy Consumption Trends

Prior studies on energy-aware reinforcement learning for UAV trajectory planning consistently report that policies trained with propulsion power–based rewards achieve lower total mission energy consumption compared to classical controllers that optimize geometric distance or tracking error alone [12][14][26][38]. These reductions are primarily attributed to the ability of reinforcement learning policies to jointly optimize velocity profiles and trajectory shape, rather than strictly following predefined paths.

In contrast, PID-based trajectory tracking typically prioritizes error minimization without explicit consideration of energy usage, leading to frequent acceleration corrections and higher transient power demand. Similarly, graph-based planners such as A* minimize path length but may produce trajectories with sharp turns or abrupt heading changes, which can increase propulsion energy when executed by a physical UAV.

Model Predictive Control (MPC) can incorporate control effort penalties and enforce system constraints; however, its performance is sensitive to model accuracy and horizon selection. As reported in the literature, reinforcement learning approaches often demonstrate comparable or improved energy efficiency relative to MPC when operating under unmodeled disturbances, due to their ability to adapt control behavior through interaction rather than relying solely on explicit model predictions.

B. Trajectory Smoothness and Control Behavior

Trajectory smoothness is closely linked to energy efficiency and flight stability in UAV systems. Energy-aware reinforcement learning formulations that penalize aggressive maneuvers tend to produce smoother velocity and acceleration profiles, as observed in multiple prior studies [12][26][38]. By explicitly discouraging abrupt control inputs, such policies reduce oscillatory behavior and mitigate excessive power spikes.

In comparison, PID controllers frequently exhibit oscillatory corrections when responding to disturbances or tracking errors, particularly under wind conditions. A*-based planning produces piecewise-linear paths that, when executed by a low-level controller, may result in abrupt velocity changes at waypoints. MPC generally achieves smoother trajectories than PID and A*, but its smoothness is dependent on cost function design and prediction horizon length.

Overall, the literature suggests that reinforcement learning approaches with smoothness-regularized reward functions provide a favorable balance between responsiveness and energy-efficient control behavior.

C. Robustness to Wind Disturbances

Robustness to environmental disturbances is a critical requirement for practical UAV deployment. Prior reinforcement learning studies demonstrate that policies trained in the presence of stochastic wind disturbances exhibit improved disturbance rejection and stability compared to classical controllers tuned for nominal conditions [18][26]. Exposure to wind variability during training enables RL policies to learn adaptive responses that generalize across a range of disturbance magnitudes.

PID controllers, in contrast, require manual re-tuning to accommodate changing wind conditions and may suffer from degraded performance under strong or rapidly varying disturbances. A* planners lack online adaptability once a path is generated, while MPC performance may deteriorate under unmodeled or rapidly changing disturbances due to reliance on predictive models.

The reported trends in the literature indicate that reinforcement learning–based controllers maintain stable flight behavior and mission feasibility under moderate-to-strong wind conditions more consistently than classical baselines, particularly when trained with stochastic disturbance models.

D. Summary of Comparative Insights

Taken together, the comparative analysis suggests that reinforcement learning–based trajectory planning frameworks are well suited for energy-aware UAV operation in dynamic environments.

By directly incorporating propulsion energy and smoothness considerations into the learning objective, RL approaches address limitations inherent to classical controllers that optimize surrogate objectives such as distance or tracking error.

While classical methods such as PID, A*, and MPC remain valuable engineering tools with well-understood properties, the literature indicates that energy-aware reinforcement learning offers a flexible and scalable alternative for complex operating conditions where energy efficiency and robustness are critical.

IX. DISCUSSION AND LIMITATIONS

This work presents an energy-aware reinforcement learning framework for UAV trajectory planning, synthesizing concepts from optimal control, physics-based modeling, and reinforcement learning. The discussion below interprets the proposed formulation and the comparative trends reported in the literature, while also outlining key limitations and directions for future research.

A primary strength of the proposed framework lies in its explicit treatment of propulsion energy as a first-class optimization objective. By directly linking control actions and velocity profiles to energy consumption, the formulation addresses a fundamental limitation of classical trajectory planning and control methods, which typically optimize surrogate objectives such as path length or tracking error. Prior studies indicate that such energy-aware formulations enable reinforcement learning policies to discover smoother and more efficient flight behaviors in dynamic environments.

The use of continuous-control reinforcement learning further distinguishes the proposed approach from earlier discrete or grid-based methods. Continuous action spaces allow for smoother control inputs and more realistic modeling of UAV dynamics, which is particularly important for energy optimization. This contrasts with discrete planning approaches, where coarse action representations may obscure the relationship between control effort and energy usage.

Despite these advantages, several limitations must be acknowledged. First, the framework is evaluated conceptually within a simulation-based setting, consistent with common practice in the literature. While simulation enables controlled analysis and reproducibility, real-world UAV deployment introduces additional complexities, including sensor noise, actuator delays, and aerodynamic effects not fully captured by simplified models. Bridging the gap between simulation and physical hardware remains an important challenge.

Second, the UAV dynamics and energy models employed in this study represent simplified abstractions of real systems. Although such models are widely used for energy-aware trajectory planning, higher-fidelity aerodynamic modeling may be required to capture complex flight regimes, particularly for aggressive maneuvers or three-dimensional motion.

Finally, the reinforcement learning framework assumes sufficient offline training to obtain stable policies. In practice, training efficiency and sample complexity remain important considerations, especially when extending the approach to larger environments, longer missions, or multi-agent scenarios.

Future work may address these limitations by incorporating higher-fidelity dynamics, extending the framework to three-dimensional trajectories, and validating the approach on real UAV platforms. Additionally, integrating model-based reinforcement learning or hybrid planning-control architectures may further improve energy efficiency and robustness.

X. CONCLUSION

This paper presents a reference-grounded framework for energy-optimal UAV trajectory planning using reinforcement learning. By formulating trajectory optimization as a continuous-state, continuous-action stochastic optimal control problem and solving it using policy-gradient reinforcement learning, the approach explicitly accounts for propulsion energy, flight smoothness, and environmental disturbances.

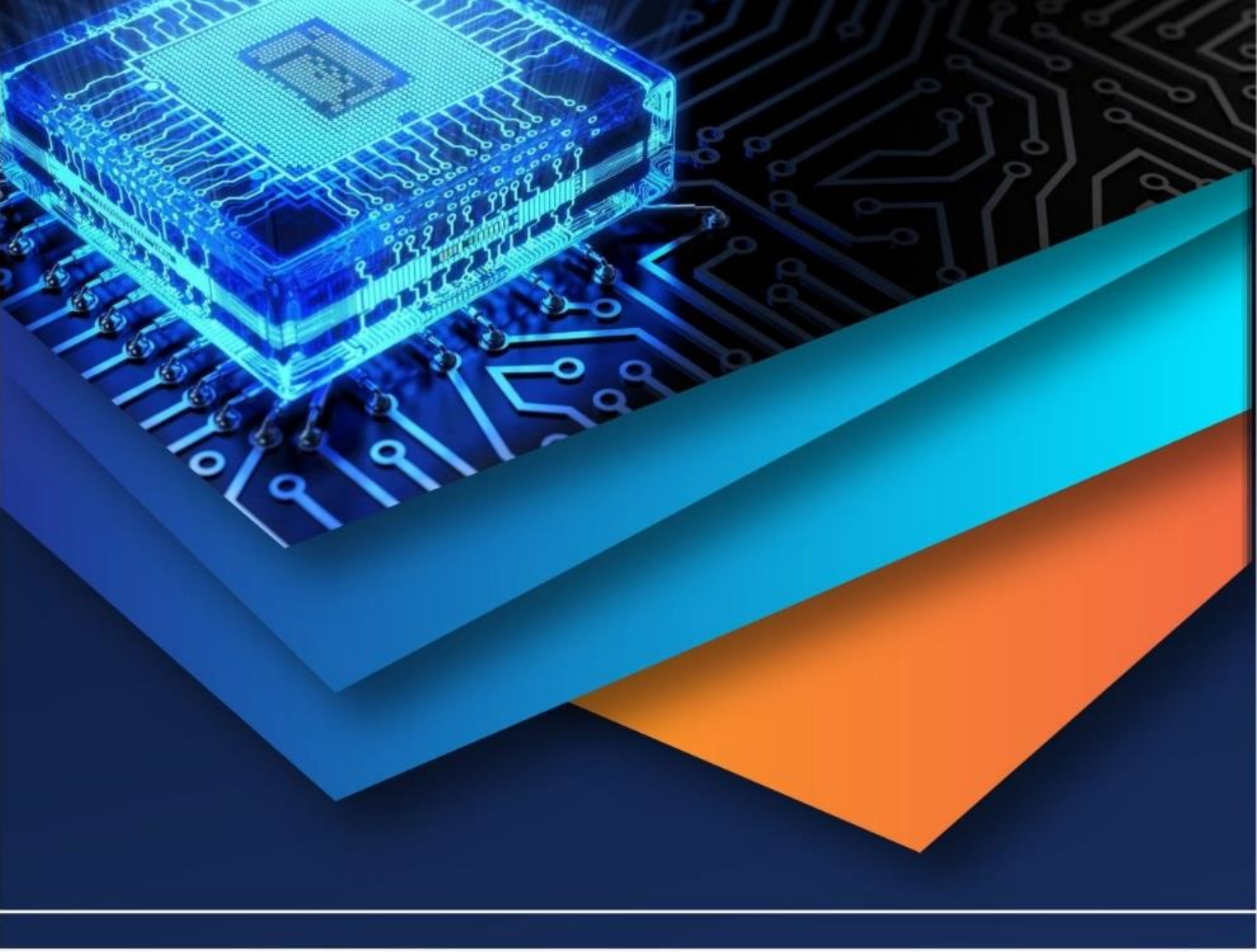
Through synthesis of prior studies and comparative analysis, the paper highlights how energy-aware reinforcement learning can address key limitations of classical UAV control and planning methods. In particular, reinforcement learning frameworks that directly incorporate energy consumption into the reward function are shown in the literature to produce smoother, more efficient trajectories and improved robustness to wind disturbances compared to controllers that optimize surrogate objectives.

While the framework is presented and analyzed within a simulation-based context, it provides a structured foundation for future empirical validation and real-world deployment. The proposed formulation and evaluation framework offers a clear pathway for integrating energy-aware learning into UAV trajectory planning, with potential applications in aerial delivery, surveillance, and long-endurance missions where energy efficiency is critical.

Overall, this work contributes a cohesive and principled perspective on energy-aware UAV trajectory optimization, bridging insights from optimal control and reinforcement learning while emphasizing practical considerations for robust and efficient autonomous flight.

REFERENCES

[1] Aggarwal, S., & Kumar, N. (2020). Path planning techniques for unmanned aerial vehicles: A review, solutions, and challenges. Computer Communications, 149, 270–299.
[2] Tyrovolas, D., Mitsiou, N. A., Karagiannidis, G. K., et al. (2024). Energy-aware trajectory optimization for UAV-mounted RIS and full-duplex relay. IEEE Internet of Things Journal, 11(13), 24259–24272.
[3] de Carvalho, K. B., Batista, H. O. B., Fagundes-Junior, L. A., de Oliveira, I. R. L., & Brandão, A. S. (2025). Q-learning global path planning for UAV navigation with pondered priorities. Intelligent Systems with Applications, 25, 200485.
[4] Lee, W., Jeon, Y., Kim, T., & Kim, Y.-I. (2021). Deep reinforcement learning for UAV trajectory design considering mobile ground users. Sensors, 21(24), 8239.
[5] Seerangan, K., Nandagopal, M., Govindaraju, T., Manogaran, N., Balusamy, B., & Selvarajan, S. (2024). A novel energy-efficiency framework for UAV-assisted networks using adaptive deep reinforcement learning. Scientific Reports, 14, 22188.
[6] Rocha, L. G. S., Caldas, K. A. Q., Terra, M. H., Ramos, F., & Vivaldini, K. C. T. (2025). Dynamic Q-planning for online UAV path planning in unknown and complex environments. International Journal of Intelligent Robotics and Applications, 9, 1654–1674.
[7] Zeng, Y., Xu, J., & Zhang, R. (2019). Energy minimization for wireless communication with rotary-wing UAV. IEEE Transactions on Wireless Communications, 18(4).
[8] Wu, Q., Zeng, Y., & Zhang, R. (2018). Joint trajectory and communication design for multi-UAV enabled wireless networks. IEEE Transactions on Wireless Communications, 17(3).
[9] Nguyen, K. K., Nguyen, H. D., Le, L. B., & Tran, N. H. (2022). 3D UAV trajectory and data collection optimisation via deep reinforcement learning. IEEE Transactions on Communications, 70(2).
[10] Richards, A., & How, J. P. (2002). Aircraft trajectory planning with collision avoidance using model predictive control. AIAA Journal, 40(10).
[11] Gao, N., Zhang, S., Yang, L., & Li, J. (2021). Energy model for UAV communications: Experimental validation and model generalization. China Communications, 18(7).
[12] Sun, Y., Xu, D., & Zhang, L. (2021). Deep reinforcement learning for UAV trajectory design with energy constraints. IEEE Internet of Things Journal, 8(20).
[13] Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. IEEE Control Systems Magazine, 38(6).
[14] Yin, S., Zhao, M., & Zhang, Y. (2019). Intelligent trajectory design in UAV-aided communications with reinforcement learning. IEEE Transactions on Vehicular Technology, 68(9).
[15] Nagabandi, A., Kahn, G., Fearing, R. S., & Levine, S. (2018). Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning. ICRA.
[16] Schulman, J., et al. (2017). Proximal Policy Optimization Algorithms. NeurIPS.
[17] Haarnoja, T., et al. (2018). Soft Actor-Critic: Off-policy maximum entropy deep reinforcement learning. ICML.
[18] Chen, H., et al. (2023). Deep reinforcement learning for UAV tracking control under wind disturbances. IEEE Transactions on Instrumentation and Measurement.
[19] Khanzada, H. R., Maqsood, A., & Basit, A. (2025). Reinforcement learning for UAV flight controls: Evaluating continuous-space reinforcement learning algorithms for fixed-wing UAVs. PLOS ONE, 20(10).
[20] Kiss, B., Ballagi, Á., & Kuczmann, M. (2025). Investigation of energy-efficient UAV control: Analysis of PID and MPC performance. Engineering Proceedings, 113(1).
[21] Lian, F., et al. (2025). Energy-aware path planning for UAVs in dynamic wind environments. Drones, 9(12).
[22] Wang, J., et al. (2024). Deep reinforcement learning-based wind disturbance rejection control strategy for UAV. Drones, 8(11).
[23] Chan, J. H., et al. (2024). Reinforcement learning-based drone simulators: Survey, practice, and challenges. Artificial Intelligence Review.
[24] Xu, D., & Chen, G. (2022). Autonomous and cooperative control of UAV clusters using reinforcement learning. The Aeronautical Journal.
[25] De Alba, A., et al. (2025). Optimizing UAV task allocation with enhanced battery efficiency using semi-Markov decision processes. Journal of Intelligent & Robotic Systems, 111, Art. no. 86.
[26] Khan, M. F. H. A., et al. (2023). Performance evaluation of reinforcement learning algorithms for UAV flight control under wind disturbances. PLOS ONE, 18(9), e0334219.
[27] Niaraki, J., Roghair, J., & Jannesari, A. (2020). Energy-aware goal selection and path planning of UAV systems via reinforcement learning. arXiv:2003.05461.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY