aerospace

Article

# From Single-Parameter Reinforcement Learning to Integrated Multi-Parameter Optimization: A Data-Driven Design Framework for Airship Aerodynamics

**Qian Zhao [1,*], Yue Yu [2,3] and Carlo E. D. Riboldi [1]**

[1] Department of Aerospace Science and Technology, Politecnico di Milano, Via La Masa 34, 20156 Milan, Italy; carlo.riboldi@polimi.it
[2] Department of Mechanical Engineering, Politecnico di Milano, Via La Masa 1, 20156 Milan, Italy; yue.yu@polimi.it
[3] State Key Laboratory of Tribology, Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China
* Correspondence: qian.zhao@polimi.it

## Abstract

This study presents a reinforcement learning (RL)-based framework for the aerodynamic optimization of the Lotte airship, combining mid-fidelity dynamic simulations with adaptive learning strategies. To address the complex nonlinear coupling between the hull shape and tail configuration, a staged, data-driven optimization strategy is developed. In the first stage, single-parameter RL experiments are conducted to independently analyze the aerodynamic sensitivity of key design variables. This conceptual stage isolates pure aerodynamic potential, focusing on the unconstrained optimization of the hull's Bézier parameterized profile, alongside the individual sensitivities of empennage area, longitudinal shift, lift slope factor, and efficiency. These experiments yield a comprehensive sensitivity map, clarifying each parameter's independent influence on the average lift-to-drag ratio ($\overline{L/D}$) of the airship. In the second stage, the obtained sensitivities are utilized to structure an integrated multi-parameter optimization scenario. Crucially, this unified environment integrates the hull and tail while enforcing rigorous longitudinal trim constraints via a dynamic bisection search. This forces the RL agent to balance system-level aerodynamic recovery against inevitable trim drag penalties. The proposed framework is implemented in MATLAB R2023b using the SILCROAD airship dynamics environment and trained by the Deep Deterministic Policy Gradient (DDPG) algorithm. Results demonstrate that the initial single-parameter sensitivity extraction not only accelerates algorithmic convergence but also significantly improves the interpretability and physical validity of the final trimmed full airship configuration. This hierarchical approach establishes a systematic path from isolated parameter understanding to holistic, physics-informed aerodynamic design, offering a transferable methodology for future autonomous airship optimization.

**Keywords:** airship aerodynamics; reinforcement learning; data-driven optimization; tail configuration; lift-to-drag ratio; dynamic simulation; SILCROAD

Check for updates

Academic Editor: Bosko Rasuo

Received: 20 April 2026
Revised: 16 May 2026
Accepted: 27 May 2026
Published: 28 May 2026

## 1. Introduction

Airships have attracted renewed interest as sustainable and cost-efficient aerial platforms for applications such as environmental monitoring, border surveillance, and communication relays [1,2]. This is thanks to their inherently superior endurance, especially

https://doi.org/10.3390/aerospace13060504

in conjunction with electric propulsion, compared to fixed-wing or rotorcraft solutions with comparable mass and payload. Their large payload capacity and long endurance make them suitable for missions that demand persistent flight at low energy cost. Where buoyancy is the major weight-sustaining effect, aerodynamic force still retains a significant relevance, often covering the gap between buoyancy and weight (lift), and being associated with a large drag component [3]. Hence, the aerodynamic efficiency of an airship, typically represented by the average lift-to-drag ratio $(\overline{L/D})$, is an interesting performance index, highly sensitive to both hull geometry and empennage configuration [4].

Traditional aerodynamic design practices relying on computational fluid dynamics (CFD) optimization are computationally expensive and typically require an in-depth knowledge of the configuration and geometry of the airship [5]. Such numerical approaches demand relevant engineering judgment and human intervention to set up an accurate mesh for reliable computations [6]. In between simple lumped parameters models and costly CFD simulations, mid-fidelity aerodynamic models are typically a balanced compromise between accuracy and ease of use [7–10]. Considering slender bodies with tails in sub-sonic flight, a model more suitable for the design phase is represented by Munk-Jones-DeLaurier theory [11]. This approach mixes potential flow and basic viscous effects to estimate the aerodynamics of the hull and empennages by relying on a geometric description of the hull and treating the sizing of the empennages through lumped lift, drag, and moment coefficients.

Recent advances in reinforcement learning (RL) have enabled data-driven optimization of nonlinear systems under complex constraints [12]. By directly interacting with a simulated environment, RL agents can autonomously learn optimal policies without the need for explicit gradient information. This capability is particularly suitable for aerodynamic shape optimization, where the design space is high-dimensional and physical coupling exists between multiple aerodynamic surfaces. Nevertheless, applying RL to airship optimization presents three main challenges: the strong aerodynamic coupling between the hull and tail, the increased sample complexity and reduced interpretability associated with high-dimensional coupled design spaces, and the potential loss of physical transparency in purely data-driven frameworks. It should be noted that such high-dimensional action spaces are not intrinsically intractable for modern continuous action RL algorithms. The sequential strategy adopted in this work is therefore not intended as a proof that simultaneous optimization is impossible, but rather as a physically interpretable decomposition of the coupled design problem.

To address these challenges, this work develops a hierarchical RL-based framework for the aerodynamic optimization of the Lotte airship, whose baseline geometry is shown in Figure 1. The framework combines a mid-fidelity dynamic simulation incorporating the Munk-Jones-DeLaurier model with an adaptive RL pipeline [11]. It follows a Two-Stage Aerodynamic Design Paradigm. In the first stage, conceptual hull exploration is conducted via single-parameter RL experiments to independently analyze the aerodynamic sensitivity of variables such as hull shape and empennage characteristics, yielding a quantitative understanding of each parameter's contribution to $\overline{L/D}$. In the second stage, these sensitivities inform an integrated multi-parameter optimization environment. Here, the RL agent optimizes the full configuration under rigorous longitudinal trim constraints, balancing system-level aerodynamic recovery against the inevitable trim drag penalties associated with moment equilibrium.To provide a clearer overview of the complete methodology, the overall workflow of the proposed staged RL-based optimization framework is summarized in Figure 2. The framework begins with the baseline Lotte airship definition and SILCROAD-based aerodynamic evaluation, then proceeds through single-parameter RL studies, sensitivity-based weighting integration, and multi-parameter RL optimization.

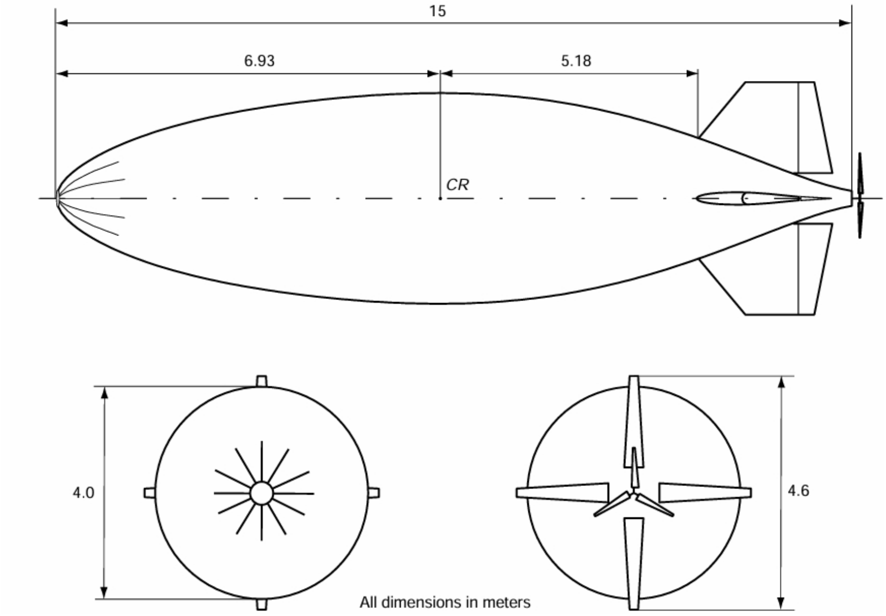


**Figure 1.** Three-view of the Lotte airship with major dimensions shown [13].

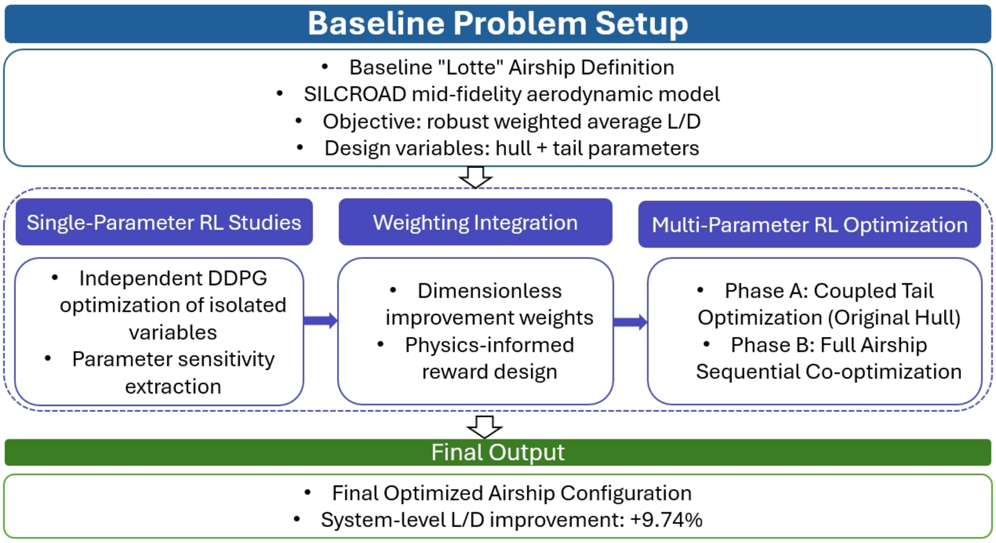


**Figure 2.** Overall workflow of the proposed reinforcement-learning-based airship aerodynamic optimization framework.

The remainder of this paper is organized as follows. Section 2 reviews prior research on airship aerodynamic optimization. Section 3 formulates the optimization problem and introduces the relevant design parameters. Section 4 details the RL framework and the DDPG algorithm. Section 5 describes the simulation setup and the baseline Lotte configuration. Section 6 presents the results of single-parameter RL studies and establishes the unified sensitivity framework. Section 7 extends the analysis to the coupled optimization of the tail assembly using a sensitivity weighted reward design. Section 8 integrates the previously optimized hull with retrained tail parameters to realize a full-airship co-optimization under trim constraints. Finally, Section 9 summarizes the main findings and discusses the physical implications of the learned design policies.

## 2. Related Work and Background

### 2.1. Aerodynamic Optimization of Airships

The aerodynamic design of airships has long been recognized as a multi-objective problem involving trade-offs among drag reduction, static stability, and payload volume. Early studies primarily relied on analytical models or empirical formulations derived from wind-tunnel testing [6]. While these approaches provided valuable insight into the general influence of fineness ratio and tail geometry, they were limited by simplified assumptions and lacked the flexibility to address nonlinear aerodynamic coupling.

With the emergence of computational fluid dynamics (CFD), airship shape optimization has progressively transitioned toward numerical simulation-based frameworks [14]. CFD-based analyses enable detailed flow-field visualization and accurate prediction of pressure distribution and viscous drag. However, their computational cost remains prohibitive, particularly for iterative optimization and design exploration. Consequently, most CFD-based optimization efforts are restricted to low-dimensional parameter spaces or rely on surrogate models and response surfaces to accelerate convergence [14]. Notwithstanding the advances achieved thus far, such methodologies remain susceptible to the effects of turbulence model selection and frequently necessitate extensive recalibration when geometry or Reynolds number undergoes variation.

In between simple, lumped parameter models, mid-fidelity aerodynamic models are typically a balanced compromise between accuracy on the one hand, and ease of use and reduced number of setting parameters on the other. Considering airships (more generally, slender bodies with tails in a largely sub-sonic flight), such models have been elaborated from Munk's non-viscous theory, adding friction and cross-flow wake effects (Jones-DeLaurier) [6,11]. Compared to wind-tunnel data, these models have proven reliable at least in static conditions.

### 2.2. Learning-Based and Data-Driven Optimization

In recent years, data-driven methodologies have emerged as powerful alternatives for aerodynamic and structural design. Machine learning (ML) and reinforcement learning (RL) have been applied to optimize airfoil geometries, turbine blades, and wing configurations, demonstrating the potential to reduce computational costs and discover non-intuitive designs [15–20]. For example, deep reinforcement learning has been used for airfoil shape optimization under unsteady flow [21], and Bayesian optimization has been integrated with CFD solvers for rapid exploration of high-dimensional design spaces [22].

Compared to conventional gradient-based or evolutionary algorithms [23–25], RL exhibits unique advantages for complex aerodynamic systems. It does not require explicit derivatives and can operate in noisy or partially observed environments, making it suitable for coupled aeroelastic or flight dynamics problems. Furthermore, RL agents can continuously adapt their strategies through trial-and-error interaction with simulation environments, a property that aligns naturally with the iterative nature of aerodynamic design.

Nevertheless, the majority of RL applications in airship aerodynamics remains focused on low-dimensional control or two-dimensional flow optimization. Few studies have addressed three-dimensional airship geometries or the coupled optimization of multiple aerodynamic components. This gap motivates the present study, which aims to integrate RL into the holistic aerodynamic design of airships, bridging physical modeling and data-driven decision-making.

### 2.3. SILCROAD Dynamics Library and Simulation Framework

The SILCROAD (Simulation Library for Craft Object Dynamics) framework has been widely used for research on airship modeling, stability analysis, and control synthesis [9,26].

It provides a modular structure for coupling aerodynamic forces, buoyancy effects, and control input models under realistic flight conditions. For airships, a mid-fidelity model is implemented by incorporating potential flow with friction and wake-induced effects derived from semi-empirical models. Through this approach, based on Munk-Jones-DeLaunier theory, SILCROAD achieves a balance between computational efficiency and physical fidelity [7–10,13,27].

In this work, SILCROAD serves as the simulation environment within which RL agents interact to evaluate aerodynamic performance. Compared with purely analytical or CFD-based approaches, this hybrid simulation framework offers three main computational benefits: (i) Rapid evaluation of design variations with consistent dynamic modeling; (ii) Computation of realistic stability derivatives essential for tail-surface optimization; (iii) Compatibility with MATLAB-based RL toolboxes for seamless policy training and testing. These characteristics make SILCROAD particularly suited for iterative, learning-driven aerodynamic optimization of airship configurations.It should be emphasized that SILCROAD is not introduced in this work as a new or unvalidated aerodynamic solver. The aerodynamic and dynamic routines employed here have already been validated and extensively used in previous airship modeling, stability analysis, and control studies. The present contribution, therefore, focuses on the reinforcement-learning-based design workflow built on top of this established simulation environment, rather than on a new validation of the SILCROAD code itself.

### 2.4. Summary of Research Gap

In summary, prior work has established solid foundations for airship aerodynamic modeling and local optimization. However, existing methods typically handle only isolated design parameters, lack adaptability across configurations, and do not exploit learning-based sensitivity analysis to inform global design strategies. The present study addresses these limitations by proposing a reinforcement learning framework that systematically bridges single-parameter understanding and integrated multi-parameter optimization within a unified aerodynamic simulation environment.

## 3. Problem Formulation and Mathematical Framework

### 3.1. Optimization Objective

The primary goal of the present work is the maximization of the aerodynamic performance of the Lotte airship, a metric quantified by the weighted average lift-to-drag ratio ($\overline{L/D}$). This aggregate value represents an interesting performance index for evaluating the aerodynamic quality of a design across its entire operational envelope. The quantity is defined over a discrete spectrum of angles of attack $\mathcal{A} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ according to the following expression:

$$\overline{L/D} = \sum_{\alpha_i \in \mathcal{A}} w(\alpha_i) \frac{L(\alpha_i)}{D(\alpha_i)}, \qquad \sum_{\alpha_i} w(\alpha_i) = 1, \tag{1}$$

where $L(\alpha_i)$ and $D(\alpha_i)$ denote the lift and drag forces predicted by the mid-fidelity simulation environment. The weighting coefficient $w(\alpha_i)$ reflects the aerodynamic relevance of each specific angle of attack. In alignment with the proposed methodology, these coefficients are hypothesized based on a preliminary knowledge of the intended airship mission. This ensures that the optimization process prioritizes flight conditions, such as steady cruise at low angles of attack, which are most critical to the vehicle's energy endurance.

To bridge the gap between conceptual exploration and system-level integration, this study adopts a Two-Stage Aerodynamic Design Paradigm. In the initial stage, the framework targets the pure aerodynamic ceiling of the bare hull without trim constraints. Subsequently, the second stage incorporates the empennage and enforces rigorous longitudinal moment equilibrium. Within this framework, the global optimization problem is formulated as the maximization of a multi-objective reward function $J(\mathbf{x})$

$$\max_{\mathbf{x}}; J(\mathbf{x}) = \overline{L/D}(\mathbf{x}) - \lambda_v P_{\text{vol}}(\mathbf{x}) - \lambda_s P_{\text{smooth}}(\mathbf{x}), \tag{2}$$

The decision vector $\mathbf{x}$ encompasses the geometric and configuration parameters defined within a longitudinal aerodynamics framework. The functions $P_{\text{vol}}$ and $P_{\text{smooth}}$ represent penalty terms designed to enforce physical feasibility and geometric continuity. Specifically, the volume penalty ensures strict volume conservation relative to the baseline $V_0$, which is an essential constraint to maintain the original aerostatic buoyancy capacity of the vehicle. The geometric smoothness penalty, $P_{\text{smooth}}$, regularizes the optimized hull profile by penalizing excessive local curvature or oscillatory variations of the Bézier radius distribution. This term prevents the RL agent from exploiting numerically favorable but physically unrealistic wavy hull shapes, and helps preserve a manufacturable and aerodynamically meaningful envelope. The terms $\lambda_v$ and $\lambda_s$ are weighting factors that regulate the relative priority of these constraints compared to the primary aerodynamic efficiency objective.

### *3.2. Design Variables for Optimization*

The array of design parameters $\mathbf{x}$ encompasses both hull and tail parameters that dominate the aerodynamic performance of an airship

$$\mathbf{x} = [x_{\text{hull}}, s_t, i_t, a_t, \eta_t, \Delta x_t, \Delta z_t, l_t]^T. \tag{3}$$

In Equation (3), each parameter represents a geometric feature affecting airship aerodynamics:

- $x_{\mathbf{hull}}$: the hull-shape design vector. In this work, the hull is represented by Bézier control pairs $(\xi_i, r_i)$, where the normalized longitudinal coordinates $\xi_i$ are fixed and the radial ordinates $r_i$ define the axisymmetric hull radius distribution $R(\xi)$. The RL agent modifies the intermediate radial ordinates, while the first and last radii are fixed to zero to preserve nose and tail closure.
- $s_t$: the horizontal tail area scaling factor such that $S_h = s_t \cdot S_{h_0}$, where $S_{h_0}$ is the baseline horizontal tail area, and $S_h$ is the current horizontal tail area.
- $i_t$: The horizontal tail incidence angle defined in degrees.
- $a_t$: the lift slope amplification factor controlling tail-lift sensitivity such that $a^t = a_t \cdot a_0^t$, where $a_0^t$ is the baseline lift slope, and $a^t$ is the current one within the optimization procedure.
- $\eta_t$: the tail hull interference efficiency coefficient accounting for induced losses.
- $\Delta x_t, \Delta z_t$: the longitudinal and vertical shifts of the aerodynamic center of the tail.
- $l_t$: the horizontal tail longitudinal arm.

Each variable is normalized within a feasible range to maintain numerical stability during RL training.

It should be noted that the parameters in Equation (3) are naturally defined within a longitudinal aerodynamics framework. However, the generically axis-symmetric shape of the hull produces a spillover of these longitudinal parameters into the lateral-directional domain. Therefore, the optimization at hand, while primarily targeting longitudinal performance, may also have an effect on the aerodynamic behavior of the airship in general. As an additional remark, it should be noted that $\eta_t$ is strictly speaking not a design parameter, being the result of a complex interplay between the wake of the hull and the tail, hence a result of a choice in the configuration. At this level, however, since many geometrical parameters defining the tail do not appear explicitly in the model, and they would affect the value of $\eta_t$, the latter has been employed as a lumped parameter to be changed. An a posteriori verification of the resulting values is therefore required.

### 3.3. Constraints

The optimization is subject to a set of physical and geometric constraints that ensure feasible and realistic airship configurations. To provide a clear mathematical representation, each constraint is explained one by one.

First, the volume conservation constraint is defined as

$$g_1(\mathbf{x}) : \left| \frac{V(\mathbf{x}) - V_0}{V_0} \right| \leq \varepsilon_V, \tag{4}$$

where $V(\mathbf{x})$ is the current hull volume and $V_0$ represents the baseline volume of the Lotte airship (110.30 $\text{m}^3$). The primary intent of this constraint is to keep the internal volume strictly constant, ensuring that the aerodynamic shape modification does not alter the baseline aerostatic buoyancy capacity of the vehicle.

Second, the geometric smoothness constraint is formulated as

$$g_2(\mathbf{x}) : \kappa(\mathbf{x}) \leq \kappa_{\max}, \tag{5}$$

where $\kappa(\mathbf{x})$ denotes the local curvature of the hull profile. This value is computed numerically along the 14 points of the Bézier parameterized curve by evaluating the second spatial derivative of the radius distribution. This condition is required to prevent unphysical surface waviness and to ensure manufacturing feasibility.

Third, the longitudinal stability constraint is given by

$$g_3(\mathbf{x}, \alpha) : |C_m(\mathbf{x}, \alpha)| \leq C_{m,lim}, \quad \forall \alpha \in \mathcal{A}. \tag{6}$$

It is critical to note that the pitching moment coefficient $C_m$ is a function of both the design vector $\mathbf{x}$ and the operational angle of attack $\alpha$. The actual meaning of this formulation is to enforce longitudinal trim equilibrium across the entire evaluated flight envelope. It ensures that the airship can maintain steady flight at varying incidence angles without requiring excessive control surface deflections that would induce prohibitive trim drag.

The tolerance thresholds $\varepsilon_V$, $\kappa_{\max}$, and $C_{m,lim}$ are empirically determined from baseline simulations. To integrate these physical limitations into the reinforcement learning reward, the penalty terms $P_{\text{vol}}$ and $P_{\text{smooth}}$ introduced in Equation (2) are formulated through a linear law. Specifically, these mathematical representations apply a linear penalty proportional to the magnitude of the constraint violation once the respective thresholds are exceeded.

### 3.4. Nature of the Optimization Problem

The aerodynamic design problem considered in this work should be regarded as a constrained, black-box, and non-convex optimization problem. The objective function is not available in closed analytical form, but is obtained from SILCROAD aerodynamic and dynamic evaluations for each candidate geometry or tail configuration. As a result, the mapping from the design vector $\mathbf{x}$ to the averaged lift-to-drag ratio $\overline{L/D}$ is governed by nonlinear aerodynamic force and moment calculations, geometric constraints, trim-related evaluations, and reward penalties.

Several features prevent the problem from being treated as a classical convex optimization problem. First, the lift-to-drag ratio itself is a nonlinear ratio between aerodynamic forces. Second, the hull-shape parameterization, volume preservation, curvature regularization, and parameter bounds introduce nonlinear constraints. Third, in the tail and full-airship optimization stages, the elevator-trim procedure and pitching-moment penalties create additional nonlinear and potentially non-smooth dependence on the design parameters. Therefore, neither convexity nor uniqueness of the optimum can be assumed.

The design landscape may contain multiple local optima or broad near-optimal regions, especially in the coupled hull–tail optimization stages, where different parameter combinations can compensate for one another. For example, tail area, tail arm, lift-slope factor, and tail efficiency may produce similar trim authority through different aerodynamic mechanisms. This coupling makes the optimization potentially multimodal and sensitive to the selected initialization and exploration strategy.

To reduce the influence of initialization and stochastic training variability, all reported RL results are averaged over three independent random seeds. Moreover, the staged workflow adopted in this study first analyzes single-parameter trends and then uses the resulting physical sensitivities to guide the coupled multi-parameter search. This strategy does not provide a mathematical proof of global optimality; rather, it improves interpretability and practical robustness when optimizing a nonlinear airship aerodynamic design space. Accordingly, the optimized configurations reported in this work should be interpreted as high-performing solutions found by the proposed RL workflow within the prescribed design bounds, rather than as guaranteed global optima of a convex problem.

### 3.5. Formulation as a Reinforcement Learning Problem

Within the RL framework, the optimization problem defined in Equations (2)–(6) is reformulated as a continuous Markov decision process (MDP) [28]. The mathematical formulation relies on the standard reinforcement learning paradigm, where an autonomous agent learns optimal design strategies through trial-and-error interaction with the simulated environment [29]. At each discrete time step $t$, the agent observes the environment state $s_t$, executes an action vector $\mathbf{a}_t$ corresponding to a specific incremental update of the design parameters $\mathbf{x}_t$, and receives a scalar reward $R_t$ defined by the current aerodynamic performance. It is important to note that the bolded action vector $\mathbf{a}_t$ defines the perturbation applied by the policy and is distinct from the scalar lift-slope parameter introduced previously.

The current scalar reward $R_t$ is computed as

$$R_t = w_1 \overline{L/D}(\mathbf{x}_t) - w_2 P_{\mathrm{vol}}(\mathbf{x}_t) - w_3 P_{\mathrm{smooth}}(\mathbf{x}_t). \tag{7}$$

In this expression, the coefficients $w_1$, $w_2$, and $w_3$ are strictly positive weighting parameters that scale the relative importance of the competing objectives. Specifically, $w_1$ regulates the primary aerodynamic efficiency metric, while $w_2$ and $w_3$ govern the strictness of the volume conservation and geometric smoothness penalties, respectively.

The policy $\pi_\theta(\mathbf{a}_t|s_t)$, parameterized by deep neural network weights $\theta$, is optimized to maximize the expected cumulative reward $\mathbb{E}\left[\sum_t \gamma^t R_t\right]$. The parameter $\gamma$ is the discount factor bounded between $0 < \gamma < 1$, which determines the present value of future rewards and ensures the convergence of the infinite horizon sum. This robust MDP formulation empowers the agent to explore high-dimensional, nonlinear design spaces and autonomously converge toward configurations that yield globally improved aerodynamic performance.

### 3.6. Discussion on Physical Interpretability

Although reinforcement learning provides an adaptive optimization mechanism, physical interpretability remains essential for engineering applications. In the present framework, the single-parameter training phase explicitly isolates the aerodynamic effect of each variable. This conceptual exploration produces a sensitivity vector **S** that quantifies the local gradient of the objective function (see Section 6 for detailed results)

$$S_i = \frac{\partial \overline{L/D}}{\partial x_i}, \quad i = 1, \ldots, 8. \tag{8}$$

This sensitivity analysis is subsequently used to support parameter ranking and reward interpretation in the integrated optimization stage. Since the design variables have different physical units, the coupled reward does not directly combine the dimensional absolute sensitivities. Instead, the later multi-parameter stage uses dimensionless improvement-based weights as a physics-informed heuristic, as detailed in Section 7.

## 4. Reinforcement Learning Framework and Methodology

### *4.1. Environment Description*

The reinforcement learning (RL) environment integrates the SILCROAD airship dynamics simulator with the RL toolbox of MATLAB. At each episode, the agent proposes a set of design variables $\mathbf{x}$, which are subsequently transferred to the simulation core to compute aerodynamic forces, moments, and the resulting $\overline{L/D}$ value. The simulation returns the averaged aerodynamic performance as the environment feedback (see Section 3.1).

The environment state $s_t$ encodes normalized aerodynamic and geometric information, including current design variables, the previously averaged lift-to-drag ratio, and constraint residuals. The action vector $\mathbf{a}_t$ represents incremental updates to the design vector, yielding the following progression:

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{a}_t, \tag{9}$$

where this update is strictly subject to parameter limits and normalization boundaries.

### *4.2. Reward Function Design*

The scalar reward guides the agent toward aerodynamic configurations that achieve high efficiency while maintaining practical feasibility. With respect to the baseline representation introduced in Equations (7), the comprehensive reward function explicitly incorporates stability and robustness terms to form a generalized architecture

$$R_t = w_1 \overline{L/D}(\mathbf{x}_t) - w_2 P_{\text{vol}}(\mathbf{x}_t) - w_3 P_{\text{smooth}}(\mathbf{x}_t) - w_4 P_{\text{stab}}(\mathbf{x}_t) + w_5 R_{\text{robust}}(\mathbf{x}_t). \tag{10}$$

Here, $P_{\text{stab}}$ penalizes large deviations of the pitching moment coefficient $C_m$, and $R_{\text{robust}}$ rewards consistent aerodynamic performance across all sampled angles of attack. All coefficients $w_i$ are normalized so that their sum equals unity. It should be explicitly noted that penalty terms, such as $P_{\text{vol}}$ and $P_{\text{smooth}}$, will be specified later on a case-by-case basis depending on the specific optimization stage. This hierarchical formulation ensures that the reward logic adapts seamlessly from single-parameter explorations to fully integrated multi-parameter environments.

### *4.3. Learning Algorithm: DDPG Implementation*

The learning algorithm adopted in this study is the Deep Deterministic Policy Gradient (DDPG), a model-free, off-policy actor-critic method suitable for continuous action spaces.

DDPG was selected because the present airship aerodynamic design problem is naturally formulated as a bounded continuous-action optimization problem. The design variables, including Bézier hull radii, tail-area scaling, tail-arm shift, lift-slope multiplier, and tail-efficiency factor, are continuous quantities rather than discrete choices. Value-based algorithms such as DQN would therefore require an artificial discretization of the design space, which may either reduce resolution or greatly increase the number of actions.

By contrast, DDPG employs a deterministic actor that directly outputs continuous design updates, making it well aligned with the parameterization adopted in this work.

In addition, the aerodynamic evaluation is performed through SILCROAD simulations, for which each function call has a non-negligible computational cost. The off-policy actor–critic structure of DDPG, together with experience replay, allows previously evaluated design samples to be reused during training, improving sample efficiency compared with purely on-policy policy-gradient methods. The deterministic nature of the policy is also appropriate for the present steady aerodynamic design problem, where the environment response is repeatable for a given configuration, and the final objective is to identify a high-performing deterministic geometry rather than a stochastic control law.

More recent continuous-control algorithms, such as TD3 or SAC, could also be considered as promising alternatives [30,31]. However, DDPG was adopted here as a well-established and computationally simple baseline for continuous aerodynamic design optimization, particularly suited to the low-to-moderate-dimensional action spaces investigated in this study. The objective of the present work is therefore not to prove the superiority of DDPG over all possible optimizers, but to demonstrate a physically interpretable RL-based design workflow for airship aerodynamic optimization. To provide a clear and mathematically rigorous description of the optimization mechanics, the algorithm relies on two deep neural networks. The first is an actor network $\pi_\theta(s_t)$ that outputs deterministic continuous action vectors $\mathbf{a}_t$. The second is a critic network $Q_\phi(s_t, \mathbf{a}_t)$ designed to estimate the state action value (Q-value).

The overarching objective of the policy optimization is to maximize the expected discounted return, which aggregates the generalized aerodynamic reward $R_t$ defined previously. This objective function is mathematically expressed as

$$J(\theta) = \mathbb{E}_{s_t, \mathbf{a}_t \sim \pi_\theta}\left[\sum_{t=0}^{T} \gamma^t R_t\right], \tag{11}$$

where $\gamma \in (0, 1)$ is the discount factor ensuring the convergence of future rewards. To optimize the actor network, the gradient of this objective with respect to the policy parameters $\theta$, known as the deterministic policy gradient, is approximated by applying the chain rule

$$\nabla_\theta J(\theta) \approx \mathbb{E}_{s_t}\left[\nabla_\theta \pi_\theta(s_t) \nabla_{\mathbf{a}} Q_\phi(s_t, \mathbf{a})\big|_{\mathbf{a}=\pi_\theta(s_t)}\right]. \tag{12}$$

This formulation dictates that the actor network is updated in the direction that maximizes the Q-value predicted by the critic.

Concurrently, the critic parameters $\phi$ are optimized by minimizing the Bellman error. This loss function measures the mean squared difference between the current Q-value estimate and the target Q-value

$$L_Q(\phi) = \mathbb{E}_{s_t, \mathbf{a}_t, R_t, s_{t+1}}\left[\left(Q_\phi(s_t, \mathbf{a}_t) - y_t\right)^2\right], \tag{13}$$

where the target $y_t$ is computed as $y_t = R_t + \gamma Q_{\phi^-}(s_{t+1}, \pi_{\theta^-}(s_{t+1}))$. To stabilize the learning process and prevent divergence, $y_t$ is evaluated using separate, slowly updating target networks denoted by parameters $(\theta^-, \phi^-)$. These target networks track the learned networks via a soft update strategy

$$\theta^- \leftarrow \tau\theta + (1-\tau)\theta^-, \tag{14}$$

$$\phi^- \leftarrow \tau\phi + (1-\tau)\phi^-, \tag{15}$$

where the parameter $\tau \ll 1$ controls the update rate. Finally, to promote adequate exploration of the continuous aerodynamic design space, temporally correlated exploration noise is introduced into the actor's decisions through an Ornstein–Uhlenbeck (OU) process.

### 4.4. Training Workflow

As explicitly stated in the Introduction (Section 1), the training process proceeds in three hierarchical stages that are strictly consistent with the problem formulation:

1. **Stage I (RL-based environment calibration and single-parameter sensitivity screening):** Each design parameter is varied independently while the others remain fixed. For a purely scalar optimization problem, deterministic approaches such as grid search, bisection, or gradient-based search would indeed provide faster and more direct convergence. The use of DDPG in this stage is therefore methodological rather than computational: it keeps the environment, reward logic, action normalization, and SILCROAD interface consistent with the subsequent multi-parameter RL stages. This isolated exploration also provides a physically interpretable sensitivity map for parameter ranking and reward design in the coupled optimization stage.
2. **Stage II (weighting integration):** The normalized sensitivities $\omega_i$ are utilized to define the relative aerodynamic importance for the subsequent joint optimization.
3. **Stage III (multi-parameter RL optimization):** All parameters are trained simultaneously using the sensitivity weighted reward architecture, which enables the agent to autonomously balance aerodynamic efficiency and physical constraint satisfaction.

### 4.5. Computational Considerations

Each RL episode consists of $N_{\text{step}}$ interactions with the SILCROAD simulator. To improve computational efficiency, aerodynamic coefficients are cached and reused whenever geometric shape variations fall below a prescribed tolerance threshold. Furthermore, an experience replay buffer comprising $10^5$ samples is employed to decorrelate sequential data and stabilize the neural network learning process. The training process was observed to converge consistently across repeated runs, thereby demonstrating stable learning behavior under the proposed hierarchical framework.

## 5. Simulation Setup and Baseline Configuration

### 5.1. Baseline Airship Model

The baseline configuration adopted in this study is that of the Lotte airship [13], whose geometric and inertial characteristics are defined in the specific initialization routine within the SILCROAD framework. The model represents a single-thruster, four-fin semi-rigid airship featuring a prolate spheroidal hull. The hull length is 16 m with a maximum diameter of 4 m, resulting in a fineness ratio of 4. The envelope volume is 110.3 $\text{m}^3$, and the total mass is 136.8 kg. The center of gravity (CG) lies slightly below the center of buoyancy (CB), providing inherent pitch stability in an ideal stand-still buoyant condition.

A single longitudinal thruster is located near the aft end of the hull at $\mathbf{r}_T = (-9.07, 0, 0)$ m from the center of buoyancy, which is taken as the reference point. The propulsion system provides a nominal maximum thrust of 500 N. Four identical empennages, comprising two horizontal and two vertical fins, form a cruciform layout. Each fin has an area of 4.63 $\text{m}^2$, giving a total lifting surface area of approximately 18.5 $\text{m}^2$. Aerodynamic interaction effects are modeled using the hull fin interference factors $\eta_{k,\text{hor/ver}} = 1.00$, and the fin efficiency coefficient $\eta_f = 0.29$.

The hull aerodynamic characteristics follow the Munk-type formulation. In terms of apparent mass and inertia, the coefficients are $(k_1, k_2, k_3, k_3') = (0.08, 0.86, 0.86, 0.69)$, and the corresponding inertia tensor is $(J_x, J_y, J_z) = (213.96, 3310, 3211)\ \text{kg} \cdot \text{m}^2$. These parameters reproduce the dynamic response observed in previous Lotte flight tests and constitute the reference model for all subsequent RL-based optimization stages. Table 1 summarizes the principal geometric and inertial properties supported by the established literature [13].

**Table 1.** Baseline geometric and inertial properties of the Lotte airship (from `SetUpAirshipOriginalLotte.m`) [13].

| Parameter | Symbol | Value |
|---|---|---|
| Overall hull length | $L_r$ | 16.00 m |
| Maximum hull diameter | $D_{\max}$ | 4.00 m |
| Fineness ratio | $L_r/D_{\max}$ | 4.00 |
| Envelope volume | $V_0$ | 110.30 $\text{m}^3$ |
| Mass | $m$ | 136.8 kg |
| Moments of inertia | $(J_x, J_y, J_z)$ | (213.96, 3310, 3211) $\text{kg}\cdot\text{m}^2$ |
| Center of gravity | $r_{\text{CG}}$ | $(-0.06, 0, 0.455)$ m |
| Center of buoyancy | $r_{\text{CB}}$ | (0, 0, 0) m |
| Number of thrusters | $N_t$ | 1 (single aft thruster) |
| Maximum thrust | | 500 N |
| Thruster position | $r_T$ | $(-9.07, 0, 0)$ m |
| Number of tail surfaces | $N_e$ | 4 (2 horizontal + 2 vertical) |
| Fin area (each) | $S_{\text{fin}}$ | 4.63 $\text{m}^2$ |
| Total fin area | $S_{\text{fin,tot}}$ | 18.5 $\text{m}^2$ |
| Fin efficiency factor | $\eta_f$ | 0.29 |
| Hull–fin interference factor | $\eta_{k,\text{hor/ver}}$ | 1.00 |
| Hull Munk coefficients | $(k_1, k_2, k_3, k_3')$ | (0.08, 0.86, 0.86, 0.69) |

### 5.2. Aerodynamic Simulation Parameters

Aerodynamic forces and moments are evaluated across an angle of attack range of $\alpha \in [0^\circ, 30^\circ]$. The reference dynamic pressure is calculated based on a free stream velocity of $U_\infty = 10\ \text{m/s}$ and a standard air density of $\rho = 1.225\ \text{kg/m}^3$. Furthermore, the weighting function $w(\alpha)$ introduced in Equation (1) is implemented as a normalized piecewise weighting rule

$$\tilde{w}(\alpha_i) = \begin{cases} 1, & 0^\circ \le \alpha_i \le 15^\circ, \\ 0.1, & 15^\circ < \alpha_i \le 30^\circ, \end{cases} \qquad w(\alpha_i) = \frac{\tilde{w}(\alpha_i)}{\sum_{j=1}^{n} \tilde{w}(\alpha_j)}. \tag{16}$$

This mathematical choice explicitly emphasizes the low-to-moderate angle-of-attack range, which is more representative of typical cruise operations of the airship, while retaining a reduced contribution from higher off-design incidence conditions. In the present study, the term "evaluated angle-of-attack range" refers to the full prescribed grid $0^\circ \le \alpha \le 30^\circ$, whereas the operationally emphasized range refers to $0^\circ \le \alpha \le 15^\circ$. The interval $15^\circ < \alpha \le 30^\circ$ is retained only as a down-weighted off-design robustness check.

### 5.3. Reinforcement Learning Hyperparameters

The reinforcement learning agent employs the DDPG architecture described in Section 4. All network and training settings were empirically tuned to achieve stable convergence. Table 2 lists the principal hyperparameters used for both single and multi parameter optimization stages.

**Table 2.** Principal reinforcement-learning hyperparameters.

| Parameter | Symbol/Setting | Value |
|---|---|---|
| Algorithm | | DDPG (Actor–Critic) |
| Discount factor | $\gamma$ | 0.99 |
| Actor learning rate | $\alpha_\theta$ | $1 \times 10^{-4}$ |
| Critic learning rate | $\alpha_\phi$ | $1 \times 10^{-3}$ |
| Soft-update coefficient | $\tau$ | 0.005 |
| Replay buffer size | $N_{\text{buffer}}$ | $1 \times 10^5$ |
| Mini-batch size | $N_{\text{batch}}$ | 256 |
| Exploration noise (OU process) | $\sigma_{\text{OU}}$ | 0.2 |
| Episode length | $N_{\text{step}}$ | 200 |
| Number of episodes (per stage) | | 800–1000 |
| Activation function | | ReLU |
| Optimizer | | Adam |

### *5.4. Evaluation Metrics*

The main performance indicator is the averaged lift-to-drag ratio $\overline{L/D}$ defined in Equation (1). Secondary metrics include the convergence trend of cumulative rewards and the geometric smoothness of the optimized hull. All results reported in subsequent sections represent averages of at least three independent training runs to mitigate stochastic variability.

## 6. Single-Parameter Reinforcement Learning Studies

This section presents a set of single-parameter RL-based exploratory studies, in which individual airship design variables are varied while all remaining configuration parameters are kept at their baseline values. It should be emphasized that these scalar studies are not proposed as computationally superior alternatives to classical one-dimensional optimization methods. Rather, their purpose is to test the RL environment in a controlled setting, verify the behavior of the reward functions and numerical safeguards, and extract parameter sensitivities under the same simulation interface later used for the coupled multi-parameter optimization. The primary objectives are to (i) characterize the learning dynamics of the agent; (ii) quantify the aerodynamic benefits achieved at convergence; and (iii) assess the local sensitivities of the mean lift-to-drag ratio, $J = \overline{L/D}$, under a unified mathematical convention. As recommended for analytical clarity, expanding upon this foundation is essential to provide the clear definitions and explanations required to interpret the subsequent multi-parameter analysis. Unless noted otherwise, all results are reported as the mean ± one standard deviation evaluated over three independent random seeds. It should be emphasized that the single-parameter studies in this section are not intended to demonstrate that DDPG is the most efficient optimizer for one-dimensional scalar problems. Instead, they serve as controlled diagnostic tests of the proposed RL environment. The resulting sensitivity curves also act as deterministic post-training checks of the RL-converged trends, allowing the monotonicity, local slope, and physical plausibility of each parameter response to be inspected before constructing the coupled multi-parameter reward.

### *6.1. Sensitivity Methodology and Notation*

Sensitivity is reported under a common convention across all isolated experiments. Let $J = \overline{L/D}$ denote the mean lift-to-drag ratio evaluated over the fixed angle of attack grid and weighting distribution (which are identical across all experiments, as detailed in Section 5.2). For a generic scalar design variable $x_i$ (selected from the design vector $\mathbf{x}$ to avoid symbolic conflict with the angle of attack $\alpha$) with a baseline value $x_{i,0}$, and a symmetric perturbation step $h$, the absolute sensitivity is computed via central differencing

$$S_{\mathrm{abs}} = \frac{\partial J}{\partial x_i} \approx \frac{J(x_{i,0}+h) - J(x_{i,0}-h)}{2h}. \tag{17}$$

The non-dimensional elastic sensitivity, which represents the percentage change in the aerodynamic objective per unit percentage change in the design variable, is defined as

$$S_{\mathrm{elas}} = \frac{x_{i,0}}{J_0} S_{\mathrm{abs}}, \quad J_0 := J(x_{i,0}). \tag{18}$$

When the baseline parameter value is strictly zero (e.g., a zero baseline incidence angle, $x_{i,0} = 0^\circ$), the relative percentage variation diverges, rendering the elastic sensitivity mathematically undefined. To circumvent this singularity, we additionally report a scaled sensitivity

$$S_{\mathrm{scaled}} = \frac{x_{\mathrm{ref}}}{J_0} S_{\mathrm{abs}}, \tag{19}$$

where $x_{\mathrm{ref}}$ acts as a unit scale (1° for angular parameters, 1 m for lengths). Furthermore, the local linearity of the objective landscape is optionally indicated by its second-order finite difference. To prevent any symbolic confusion with the geometric hull curvature $\kappa$ introduced previously in Equation (5), the mathematical curvature of the objective function is denoted herein as $J''$

$$J'' = \frac{J(x_{i,0}+h) - 2J_0 + J(x_{i,0}-h)}{h^2}. \tag{20}$$

### *6.2. Hull Optimization via Reinforcement Learning*

#### 6.2.1. Experimental Setup

We parameterize the Lotte hull radius distribution $R(\xi)$ with a 14 points Bézier control vector ($n_{\mathrm{ctrl}} = 14$); the first and last radii are fixed to zero to ensure smooth geometric closure [32]. A DDPG agent updates the intermediate radii to maximize the aerodynamic objective subject to shape smoothness and constant volume constraints. The actor and critic networks consist of two fully connected layers (64 neurons, ReLU activation). Each episode performs one geometry update followed by an aerodynamic evaluation with SILCROAD.

To achieve a rigorous decoupling of the airship aerodynamics, this stage adopts a "virtual wind tunnel" approach on a bare hull configuration (i.e., empennages are completely removed). Consequently, the aerodynamic evaluation is conducted without enforcing longitudinal trim. The fundamental rationale is to isolate the pure aerodynamic efficiency of the evolving geometry. Enforcing a trim condition at this stage would necessitate artificial tail deflections, thereby polluting the reward signal with varying trim drag and obscuring the intrinsic geometric benefits of the hull. Furthermore, to prevent non-physical performance plateaus at high angles of attack, all aerodynamic forces are rigorously projected from the body frame to the wind axes before computing the lift-to-drag ratio.

##### Reward Function

The environment implements a robustness-oriented reward with curvature smoothing and invalid-state protection. Let $n$ be the number of control radii, $\Delta r_i$ the difference between adjacent control radii, and $J = \overline{L/D}$ the mean lift-to-drag ratio returned by the solver

$$\mathcal{R}_{\mathrm{hull}} = \begin{cases} -10^3, & \text{if } J \text{ is non-finite;} \\ J \;-\; 0.1\,\dfrac{1}{n}\displaystyle\sum_{i=1}^{n-1} |\Delta r_i|, & \text{otherwise.} \end{cases} \tag{21}$$

The hull volume is strictly preserved by scaling the Bézier radius field to match the original volume before evaluation. This is a relevant action, implemented so as to keep the buoyancy property of the airship constant, thus ensuring that aerodynamic optimization does not induce changes in aerostatic capacity.

### 6.2.2. Learning Dynamics

The training reward rises steadily and saturates after a phase of active exploration, indicating stable, oscillation-free convergence across all evaluated random seeds; see Figure 3.

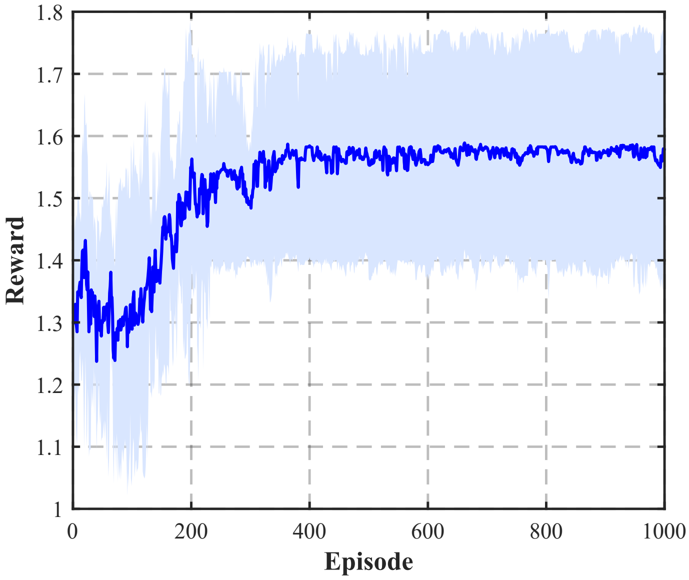


**Figure 3.** Hull RL—training reward (mean ± std over three seeds).

### 6.2.3. Aerodynamic Response

Relative to the baseline, the optimized radius distribution exhibits a slightly fuller mid-body and a smoother aft taper. Thanks to the bare hull formulation, these modifications purely reflect enhanced pressure recovery and reduced form drag. Consequently, the $L/D$–AoA curve shifts upward, with pronounced gains at moderate-to-high AoA (Figure 4).

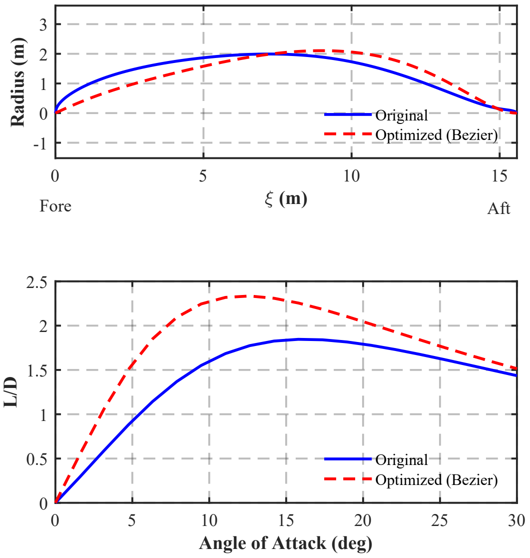


**Figure 4.** Hull RL—(**top**) baseline vs. optimized radii; (**bottom**) $L/D$–AoA curves.

The hull-optimization results reported in this section should be interpreted within the validated operating scope of the SILCROAD mid-fidelity aerodynamic model. The purpose of this study is not to revalidate SILCROAD, which has already been employed in previous airship studies, but to use it as an established simulation environment for RL-based design exploration. At the same time, as with any mid-fidelity model based on semi-empirical aerodynamic corrections, the optimized Bézier hull should be regarded as a model-predicted improvement rather than as a final high-fidelity-certified aerodynamic optimum. Smooth Bézier parameterization, fixed nose and tail closure, volume preservation, and curvature regularization were used to reduce the possibility of non-physical shape exploitation. Further high-fidelity CFD assessment could be useful in future work if the optimized geometry is to be advanced toward detailed aerodynamic design.

### 6.2.4. Sensitivity

Centered radius scaling of ±5–10% is applied to the optimized shape. It must be explicitly clarified that this global scaling evaluation intentionally relaxes the volume conservation constraint to isolate the purely geometric effect of hull slenderness. Because the RL optimization converged to a constrained optimum on the constant volume manifold, the unconstrained gradient with respect to global thickness scaling is naturally non-zero.

With the converged optimal objective $J_{\text{opt}} = 1.7558$ and a baseline scaling factor $x_{i,0} = 1.0$, the absolute sensitivity evaluates to $S_{\text{abs}} \approx 0.4143$ per unit scale. Consequently, the elastic sensitivity is $S_{\text{elas}} = 0.236$ (per 1% radius scaling). The response trend exhibits near linear behavior within the probed unconstrained neighborhood (Figure 5).

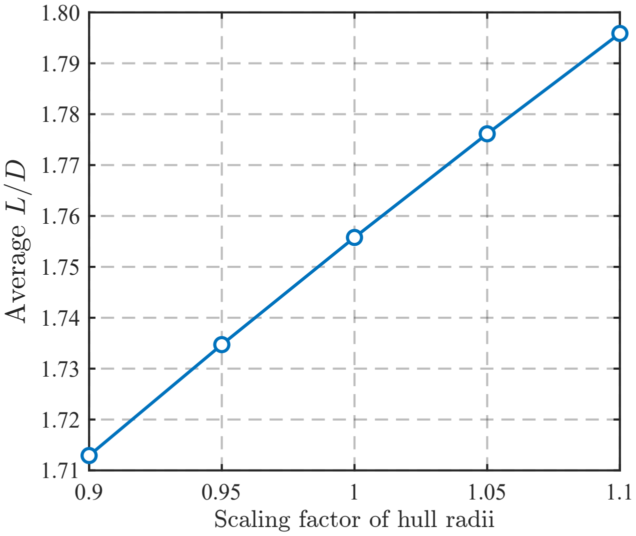


**Figure 5.** Hull RL—sensitivity of $J$ to global radius scaling (centered differences near the optimum).

The corresponding statistical summary of the hull optimization results and sensitivity metrics is reported in Table 3.

**Table 3.** Hull RL—statistical summary (three seeds).

| Metric | Baseline | Optimized (Mean ± Std) | Improvement (%) |
|---|---|---|---|
| Average $\overline{L/D}$ | 1.3949 | 1.6635 ± 0.1889 | +19.26 |
| Absolute sensitivity $S_{\text{abs}}$ | 0.4143 (per unit radius scale) | | — |
| Elastic sensitivity $S_{\text{elas}}$ | 0.236 (per 1% radius) | | — |

### *6.3. Horizontal Tail Area Optimization via Reinforcement Learning*

#### 6.3.1. Experimental Setup

Following the bare hull optimization in Section 6.2, the design framework progresses to the second stage of the aerodynamic paradigm: empennage integration and longitudinal trim. The horizontal tail area scale $k_A = S_t/S_0$ is optimized while the previously optimized hull geometry is retained. The feasible exploration range is $k_A \in [0.5, 1.8]$.

Unlike the virtual wind tunnel approach applied to the bare hull, evaluating the tail area necessitates enforcing a strict longitudinal trim condition. The DDPG agent must now navigate the delicate trade-off between the aerodynamic stabilizing benefits of the tail surface and the severe penalty of trim drag. The environment relies on a 32–32 ReLU network architecture (actor learning rate $10^{-4}$, critic $10^{-3}$). Each episode triggers one trimmed aerodynamic evaluation, and all results are averaged over three independent random seeds.

##### Reward Function

Consistent with the environment implementation, the reward uses the unnormalized mean lift-to-drag ratio penalized by the absolute deviation of the area scale from unity

$$\mathcal{R}_{\text{area}} = \begin{cases} -10^3, & \text{if } J \text{ is non-finite;} \\ J \; - \; \lambda_A \left| k_A - 1.0 \right|, & \text{otherwise,} \end{cases} \qquad \lambda_A = 0.1, \quad J = \overline{L/D}. \tag{22}$$

#### 6.3.2. Learning Dynamics

Unlike the pronounced learning curve observed in the 14-dimensional hull optimization, the training reward for the tail area exhibits a remarkably rapid, nearly instantaneous stabilization (Figure 6). This flat convergence profile is highly characteristic of low-dimensional RL tasks and is attributed to two primary factors. First, the single-parameter action space allows the agent's initial exploration noise to sample the optimal region almost immediately. Second, the baseline Lotte empennage is already a highly refined engineering design; thus, the theoretical margin for improvement (4.60%) is fundamentally constrained. Consequently, rather than executing a steep initial ascent, the DDPG agent demonstrates a rapid lock onto the local aerodynamic optimum, maintaining stable and bounded oscillations around this peak without divergence.

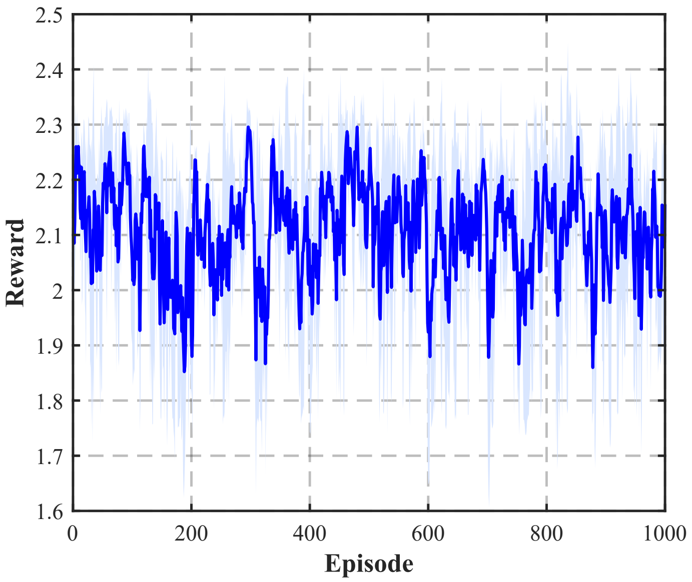


**Figure 6.** Tail-area RL—training reward (mean ± std over three seeds).

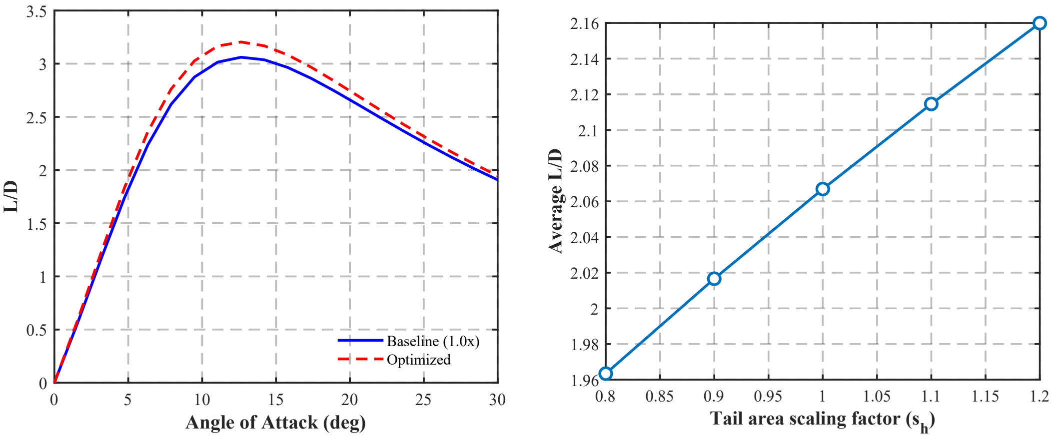
### 6.3.3. Outcome: Aerodynamic Configuration

The optimal solution corresponds to an approximately 1.205× enlargement of the horizontal tail ($k_A = 1.205$). By enhancing the trim authority and the overall lifting surface, this configuration delivers consistent $L/D$ gains across the entire incidence range $0^\circ \leq \alpha \leq 30^\circ$. The average lift-to-drag ratio improves from the baseline of 2.0669 to an optimized value of $2.1620 \pm 0.0047$, yielding a solid 4.60% aerodynamic enhancement.

### 6.3.4. Sensitivity

The sensitivity analysis confirms that the optimized tail area resides at a true local aerodynamic optimum, where the local gradient of the objective function naturally approaches zero. This physical behavior directly demonstrates the precise balance achieved by the RL agent: the initial increase in tail area provides significant lift-to-drag benefits by improving the overall lifting efficiency and trim authority; however, further area expansion introduces excessive wetted surface area, leading to a rise in parasitic friction drag that directly overrides any additional aerodynamic gains.

Centered ±10% perturbations in $k_A$ about the optimum give an absolute sensitivity of $S_{\text{abs}} \approx 0.4900$ per unit $k_A$ (evaluated with $J_{\text{opt}} = 2.1620$ and $p_0 = 1.0$). Consequently, the elastic sensitivity is $S_{\text{elas}} = 0.227$ (per 1% area). The response trend demonstrates the diminishing returns of tail enlargement in the neighborhood of the optimum, as shown in Figure 7.



(**a**) (**b**)

**Figure 7.** Tail-area RL—(**a**) $L/D$–AoA comparison between baseline and optimized configurations; (**b**) sensitivity of $J$ to $k_A$ (centered differences near the optimum).

The corresponding statistical summary of the tail-area optimization results and sensitivity metrics is reported in Table 4.

**Table 4.** Tail-area RL—statistical summary (three seeds).

| Metric | Baseline | Optimized (Mean ± Std) | Improvement (%) |
|---|---|---|---|
| Average $L/D$ | 2.0669 | $2.1620 \pm 0.0047$ | +4.60 |
| Optimal $k_A$ | 1.0000 | 1.205 | — |
| Absolute sensitivity $S_{\text{abs}}$ | ≈0.4900 (per unit $k_A$) | | — |
| Elastic sensitivity $S_{\text{elas}}$ | 0.227 (per 1% area) | | — |

## *6.4. Tail Incidence Angle Optimization via Reinforcement Learning*

### 6.4.1. Experimental Setup

The incidence angle $i_t \in [-3^\circ, 3^\circ]$ is optimized as a single continuous action within the same DDPG framework (32–32, ReLU). Building upon the trimmed evaluation paradigm established, the core objective of this stage is to perform high-precision aerodynamic fine-tuning. By adjusting the zero-lift line of the empennage, the agent seeks to minimize the residual pitching moment and the associated trim drag without inducing premature flow separation. The aerodynamic evaluator returns the mean lift-to-drag ratio and the pitching-moment history. Three seeds with 1000 episodes are used.

#### Reward Function

The environment uses a mean-$L/D$ objective with mild regularization on extreme incidence and pitch moment magnitude, plus a numerical warning penalty

$$\mathcal{R}_{\text{inc}}(i_t) = \begin{cases} -10^3, & \text{if } J \text{ is non-finite;} \\ J - \lambda_i \left| \dfrac{i_t}{3^\circ} \right| - \lambda_M \dfrac{\overline{|M_y|}}{100} - 0.01\,\texttt{warnCount}, & \text{otherwise,} \end{cases} \tag{23}$$

where $J = \overline{L/D}$ and $\overline{|M_y|}$ is the mean absolute pitching moment from the solver. The hyperparameters $\lambda_i = 0.05$ and $\lambda_M = 0.01$ act as soft physics-informed constraints. Specifically, the incidence penalty weight ($\lambda_i$) is scaled to gently discourage extreme deflections that risk premature flow separation and excessive structural loads. Concurrently, the moment penalty weight ($\lambda_M$) penalizes severe aerodynamic imbalance, implicitly guiding the agent toward naturally trimmed states to minimize subsequent elevator-induced trim drag.

### 6.4.2. Learning Dynamics

The reward trajectory stabilizes rapidly, reflecting a mathematically shallow objective landscape with respect to $i_t$ (Figure 8). Because the baseline configuration (0°) is already a mature engineering estimate close to the theoretical optimum, the DDPG agent effectively operates in an exploitation-dominant regime, meticulously narrowing down the policy variance to lock onto a fractional incidence angle.

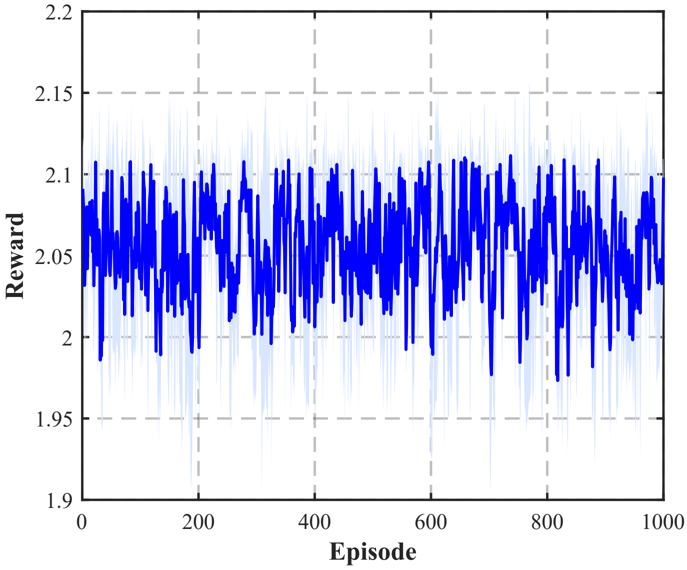


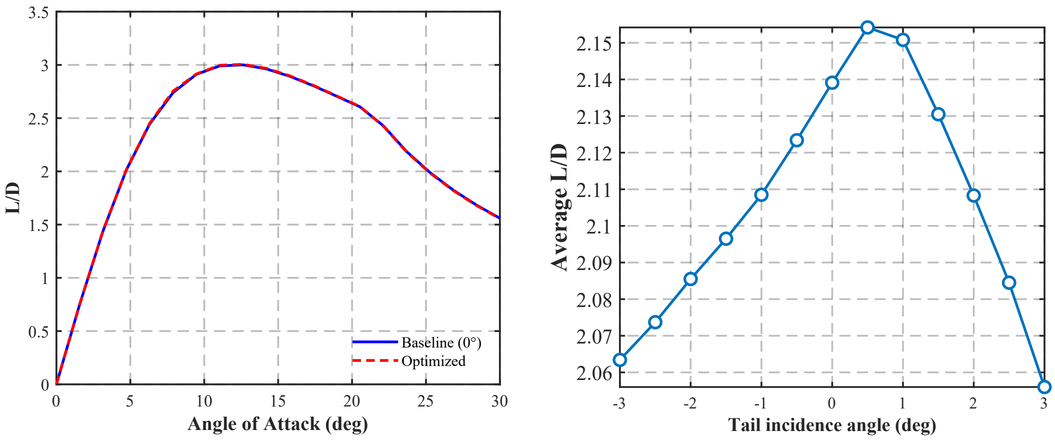
**Figure 8.** Tail-incidence RL—training reward (mean ± std over three seeds).

### 6.4.3. Outcome: Aerodynamic Configuration

The cross seed mean optimal angle converges robustly to $i_t^* = 0.132°$. This optimized micro deflection yields an average lift-to-drag ratio of 2.1426 ± 0.0012, representing a +0.17% improvement over the 2.1391 baseline. While the absolute aerodynamic gain is naturally marginal, this result is physically significant: it demonstrates the high-fidelity capability of the RL framework to discover nuanced, fractional degree structural adjustments that successfully mitigate trim drag penalties that analytical methods might overlook.

### 6.4.4. Sensitivity

To evaluate the local sensitivity, centered finite differences of ±1° are applied around the baseline. Because the baseline incidence is exactly zero ($p_0 = 0°$), the classical elastic sensitivity diverges and is mathematically undefined. Addressing this singularity via the scaled convention established in Section 6, with a reference perturbation $p_{\text{ref}} = 1°$ and $J_0 = 2.1391$, the absolute sensitivity evaluates to $S_{\text{abs}} = 0.0212$ (L/D per degree). Consequently, the well-defined scaled sensitivity is $S_{\text{scaled}} = 0.0099$. The smooth, shallow curvature near 0° confirms that the airship operates at a refined, well-balanced longitudinal state, as shown in Figure 9.



(**a**) (**b**)

**Figure 9.** Tail-incidence RL—(**a**) baseline (0°) vs. evaluation at $i_t^* = 0.132°$; (**b**) sensitivity of $J$ to $i_t$ about 0° (centered differences).

The corresponding statistical summary of the tail-incidence optimization results and sensitivity metrics is reported in Table 5.

**Table 5.** Tail-incidence RL—statistical summary (three seeds).

| **Metric** | **Baseline** | **Optimized (Mean ± Std)** | **Change (%)** |
|---|---|---|---|
| Average $L/D$ | 2.1391 | $2.1426 \pm 0.0012$ | +0.17 |
| Mean optimal $i_t^*$ (deg) | | 0.132 | |
| Absolute sensitivity $S_{\text{abs}}$ | | 0.0212 (L/D per degree) | |
| Scaled sensitivity $S_{\text{scaled}}$ | | $9.9 \times 10^{-3}$ (with $p_{\text{ref}} = 1°$) | |

## *6.5. Tail-Arm Length Optimization via Reinforcement Learning*

### 6.5.1. Experimental Setup

The tail-arm shift, denoted as $\Delta x_t \in [-3, 3]$ m (positive for aft shifts), defines the longitudinal translation of the horizontal tail's aerodynamic center relative to the baseline station. Similar to the previous stage, the DDPG agent fine-tunes this continuous spatial parameter under trimmed conditions. The underlying physical trade-off is fundamental: extending the tail arm enhances the longitudinal restoring moment and trim authority, but excessively aft placements induce severe structural weight penalties, potential aeroelastic flutter risks, and elevated local friction drag.

### Reward Function

To enforce these realistic engineering constraints, the agent is guided by a trim-aware reward function that balances pure aerodynamic efficiency against structural and balance penalties

$$\mathcal{R}_{\text{arm}}(\Delta x_t) = \begin{cases} -10^3, & \text{if } J \text{ is non-finite;} \\ J \;-\; \lambda_l\,|\Delta x_t| \;-\; \lambda_M\,\dfrac{\overline{|M_y|}}{100} \;-\; 0.01\,\texttt{warnCount}, & \text{otherwise,} \end{cases} \tag{24}$$

where $J = \overline{L/D}$. The penalty weights ($\lambda_l = 0.005$ and $\lambda_M = 0.01$) are implemented as soft physics-informed regularizations. The arm length penalty ($\lambda_l$) deliberately discourages unrealistic, needle-like aft extensions that violate structural mass limits. Meanwhile, the pitching moment penalty ($\lambda_M$) ensures the airship avoids regimes where the elevator struggles to maintain a clean trim.

### 6.5.2. Learning Dynamics

Learning proceeds smoothly and monotonically, achieving stable saturation without divergence (Figure 10). This robust convergence highlights the agent's capability to balance the linearly increasing spatial penalty against the aerodynamic benefits of the extended moment arm.

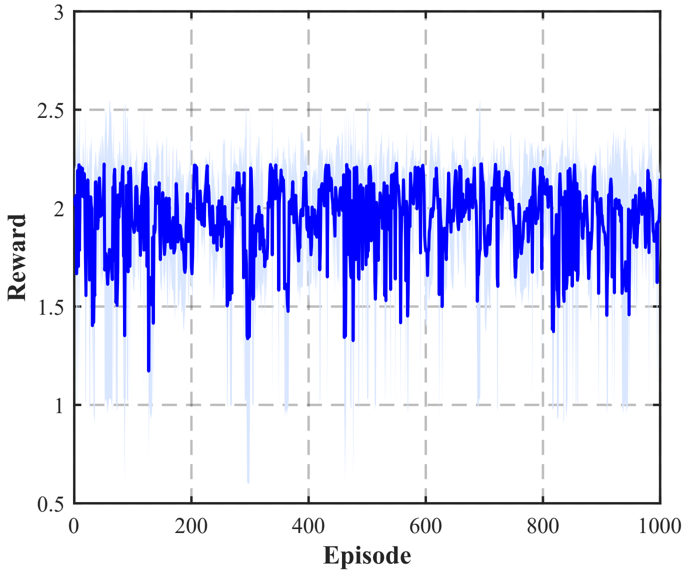


**Figure 10.** Tail-arm RL—training reward (mean ± std over three seeds).

### 6.5.3. Outcome: Aerodynamic Configuration

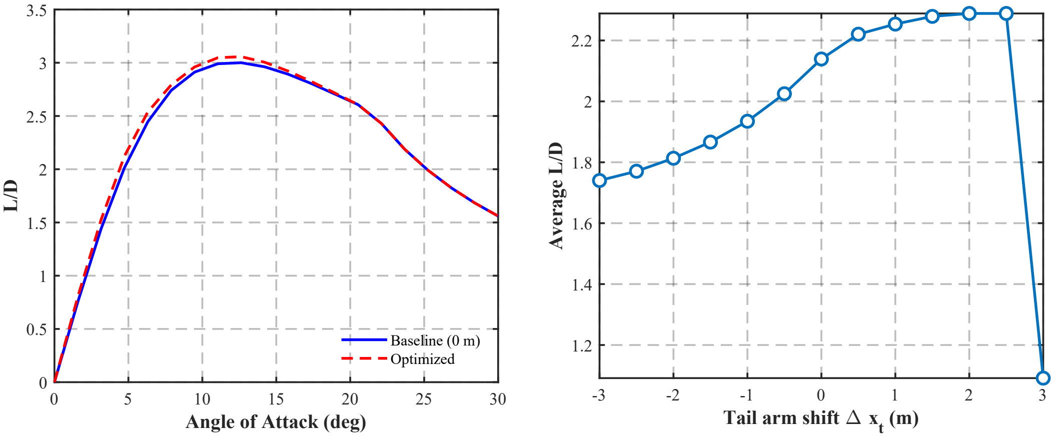
The optimized tail arm converges precisely to an aft shift of $\Delta x_t^* = +0.242$ m, elevating the average $L/D$ from 2.1391 to a stabilized 2.1924 ± 0.0133 (+2.49% improvement).

However, a critical system-level discrepancy is observed when comparing this RL optimum against the pure aerodynamic sensitivity. The sensitivity analysis reveals that the unpenalized $L/D$ curve monotonically increases well beyond the RL optimum, only reaching a saturation plateau between +2.0 m and +2.5 m. This pure aerodynamic trend is physically sound: an extended tail arm amplifies the restoring moment, reducing the elevator deflection required for trim and consequently minimizing trim drag.

### 6.5.4. Sensitivity and the Operational Frontier

Because the baseline shift is exactly zero ($p_0 = 0\,\text{m}$), the scaled sensitivity formulation is utilized. With a reference perturbation $p_{\text{ref}} = 1\,\text{m}$ and evaluated via centered $\pm 0.5\,\text{m}$ finite differences, the absolute sensitivity is $S_{\text{abs}} = 0.1955$ (L/D per metre), yielding a scaled sensitivity of $S_{\text{scaled}} = 0.0914$.

The strictly positive gradient directly answers why the RL agent halted at $+0.242\,\text{m}$ instead of advancing to the aerodynamic peak at $+2.5\,\text{m}$. While shifting the tail beyond $+0.242\,\text{m}$ yields marginal aerodynamic gains, it triggers severe multi disciplinary penalties in the reward function. Excessive aft placements inherently violate structural mass limits and push the empennage into complex wake regions where the solver fails to maintain a clean trim, generating residual pitching moments and numerical warnings. Therefore, the RL convergence at $+0.242\,\text{m}$ is not a failure to find the aerodynamic peak, but rather the successful identification of the critical operational frontier—the maximum aerodynamic gain achievable before structural and trim-failure penalties overwhelm the system, as illustrated in Figure 11.



(**a**) (**b**)

**Figure 11.** Tail-arm RL—(**a**) $L/D$–AoA: baseline vs. evaluation at $\Delta x_t^* = +0.242$ m; (**b**) sensitivity of $J$ to $\Delta x_t$ about the baseline.

The corresponding statistical summary of the tail-arm optimization results is reported in Table 6.

**Table 6.** Tail-arm RL—statistical summary (three seeds).

| Metric | Baseline | Optimized (Mean ± Std) | Improvement (%) |
|---|---|---|---|
| Average $L/D$ | 2.1391 | $2.1924 \pm 0.0133$ | +2.49 |
| Mean optimal $\Delta x_t^*$ (m) | | +0.242 | |
| Absolute sensitivity $S_{\text{abs}}$ | | 0.1955 (L/D per metre) | |
| Scaled sensitivity $S_{\text{scaled}}$ | | 0.0914 (with $p_{\text{ref}} = 1$ m) | |

## *6.6. Tail Efficiency Optimization via Reinforcement Learning*

### 6.6.1. Experimental Setup and Physical Justification

The tail efficiency factor, denoted as $\eta_f \in [0.2, 0.6]$, represents the aerodynamic effectiveness of the horizontal tail operating within the retarded wake flow of the massive hull. From a rigorous engineering standpoint, $\eta_f$ is a physical consequence of wake interference rather than a direct geometric parameter. However, optimizing $\eta_f$ as a continuous action serves as a vital surrogate representation for profound structural modifications. In reality, enhancing $\eta_f$ necessitates moving the empennage out of the boundary layer (e.g., employing T-tail configurations or extending the root chord), which inevitably incurs severe structural complexity and parasitic drag penalties.

A single action DDPG agent controls this surrogate variable $\eta_f$ under strictly trimmed conditions. The objective is to identify the critical threshold where the aerodynamic benefit of an uncluttered tail flow is perfectly balanced by the simulated structural cost. Three random seeds and 1000 episodes are employed.

Reward Function

To mathematically emulate the aforementioned physical and structural trade-offs, the reward function maximizes the mean lift-to-drag ratio while deliberately penalizing any deviation from the historical Lotte baseline ($\eta_{f,0} = 0.29$):

$$\mathcal{R}_{\text{etaf}}(\eta_f) = \begin{cases} -10^3, & \text{if } J \text{ is non-finite;} \\ J \; - \; \lambda_\eta \, |\eta_f - 0.29| \; - \; \lambda_M \, \dfrac{\overline{|M_y|}}{100} \; - \; 0.01\, \texttt{warnCount}, & \text{otherwise,} \end{cases} \tag{25}$$

where $J = \overline{L/D}$. The efficiency penalty weight ($\lambda_\eta = 0.005$) acts as a physics-informed surrogate for the structural and drag penalties incurred by aggressively improving wake conditions. The moment penalty ($\lambda_M = 0.01$) ensures robust pitch trim.

### 6.6.2. Learning Dynamics

Learning exhibits a monotonic increase and converges smoothly, stabilizing reliably across all random seeds (Figure 12). This confirms that the DDPG agent has successfully deciphered the surrogate cost function, effectively balancing the enhanced trim authority against the imposed structural penalty.

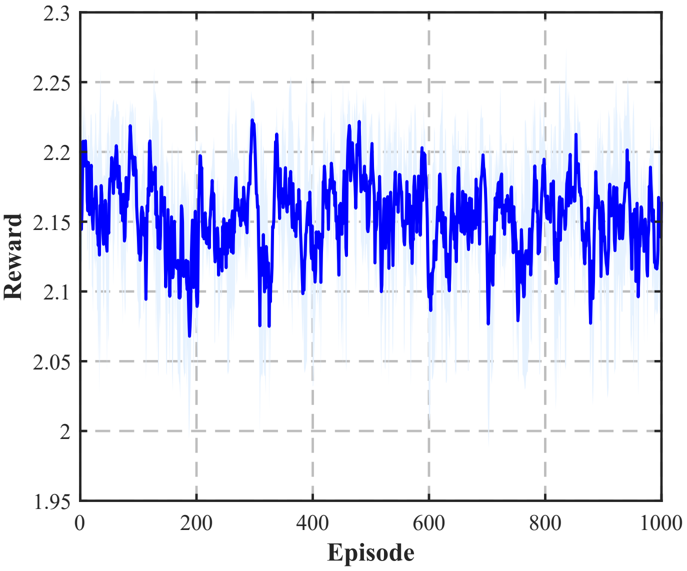


**Figure 12.** Tail-efficiency RL—training reward (mean ± std over three seeds).

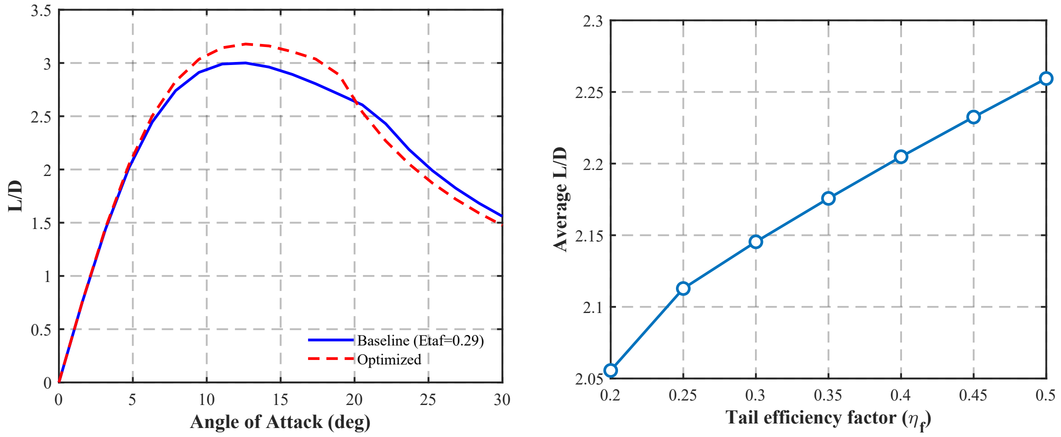
### 6.6.3. Outcome: Aerodynamic Configuration

The learned system-level optimal efficiency is $\eta_f^* = 0.411$ (mean over three seeds). Reevaluation at $\eta_f^*$ shows that the $L/D$–AoA curve rises consistently over the evaluated incidence range, improving the average lift-to-drag ratio from the 2.1391 baseline to $2.2116 \pm 0.0031$, which constitutes a solid $+3.39\%$ aerodynamic improvement. This demonstrates that elevating the tail efficiency significantly mitigates the elevator deflections required for trim, thereby minimizing overall trim drag.

### 6.6.4. Sensitivity and the System-Level Optimum

The sensitivity of the mean lift-to-drag ratio with respect to $\eta_f$ elucidates the multidisciplinary trade-off. Evaluated via centered finite differences about the optimum, the absolute sensitivity is $S_{\text{abs}} \approx 0.32$ (L/D per unit $\eta_f$). With $J_0 = 2.1391$ and a reference scale $p_{\text{ref}} = 0.1$, the scaled sensitivity is robustly defined.

Most crucially, the $J$–$\eta_f$ curve displays a clear convex shape, peaking precisely near the RL optimum of 0.411. This saturation reveals that beyond $\eta_f = 0.411$, the virtual structural penalty ($-\lambda_\eta|\eta_f - 0.29|$) strictly overrides the diminishing returns of trim drag reduction. The agent proves that blindly pursuing maximum theoretical tail efficiency ($\eta_f = 0.6$) is a suboptimal engineering choice when structural costs are accounted for, as shown in Figure 13.



(**a**) (**b**)

**Figure 13.** Tail-efficiency RL—(**a**) $L/D$–AoA: baseline ($\eta_f = 0.29$) vs. optimized ($\eta_f^* = 0.411$); (**b**) sensitivity of $J$ to $\eta_f$.

The corresponding statistical summary of the tail-efficiency optimization results and sensitivity metrics is reported in Table 7.

**Table 7.** Tail-efficiency RL—statistical summary (three seeds).

| Metric | Baseline | Optimized (Mean ± Std) | Improvement (%) |
|---|---|---|---|
| Average $L/D$ | 2.1391 | $2.2116 \pm 0.0031$ | +3.39 |
| Mean optimal $\eta_f^*$ | | 0.411 | |
| Absolute sensitivity $S_{\text{abs}}$ | | ≈0.32 (L/D per unit $\eta_f$) | |
| Scaled sensitivity $S_{\text{scaled}}$ | | 0.0150 (with $p_{\text{ref}} = 0.1$) | |

## *6.7. Tail-Lift Slope Optimization via Reinforcement Learning*

### 6.7.1. Experimental Setup

The tail-lift curve slope factor ($k_{L\alpha}$, or $C_{l\alpha,t}$ multiplier) determines the sensitivity of tail lift to the angle of attack. This parameter serves as a high-level surrogate for airfoil section refinement or planform adjustments (e.g., aspect ratio modifications). It is varied within [0.6, 1.4] to represent under- and over-responsive tail behavior. A single action DDPG agent controls this scaling factor under trimmed conditions, with aerodynamic evaluations provided by SILCROAD. Three independent seeds and 1000 episodes are employed.

Reward Function

The reward function is specifically designed to navigate the flat response surface of the lift slope parameter by applying a scaling factor and a nonlinear moment penalty

$$\mathcal{R}_{\text{slope}} = 10\left(\overline{L/D} - \lambda_M \ln\left(1 + \frac{\overline{|M_y|}}{1000}\right) - \lambda_W \,\texttt{warnCount}\right), \tag{26}$$

where $\lambda_M = \lambda_W = 0.005$. The scaling factor of 10 is implemented to amplify the reward gradient, facilitating the agent's convergence in a region of marginal returns. The logarithmic term $\ln(1 + \dots)$ introduces a soft saturated penalty for the pitching moment, providing a robust gradient for large imbalances while allowing for nuanced trade-offs near the trimmed equilibrium.

### 6.7.2. Learning Dynamics

The learning process converges smoothly around episode 900 (Figure 14). The stabilization of the mean reward confirms that the DDPG agent has successfully identified the subtle balance between enhanced lift generation and the associated trim penalties within the provided design bounds.

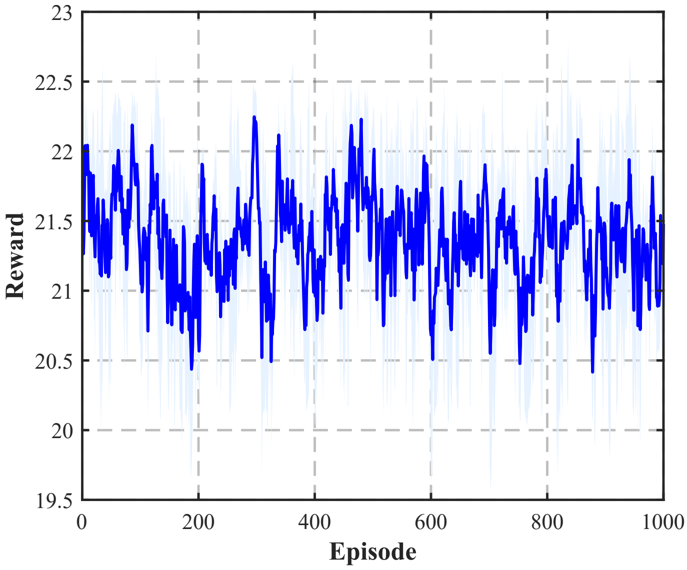


**Figure 14.** Tail-lift slope RL—training reward (mean ± std over three seeds).

### 6.7.3. Outcome: Aerodynamic Configuration

The optimized lift slope factor is $k^*_{L\alpha} = 1.014$ (mean over three seeds), representing a minor correction toward a more responsive tail configuration. Re-evaluation yields an average $L/D$ of $2.1424 \pm 0.0074$, achieving a fractional but consistent improvement of $+0.15\%$ over the 2.1391 baseline.

This marginal gain carries significant physical implications: it demonstrates that the original Lotte empennage's airfoil selection and planform efficiency are already operating near the theoretical aerodynamic limit for this specific hull wake environment. The RL agent's convergence near the baseline value ($k_{L\alpha} \approx 1.0$) validates the maturity of the original design while still capturing a residual 0.15% gain through high precision policy search.

### 6.7.4. Sensitivity

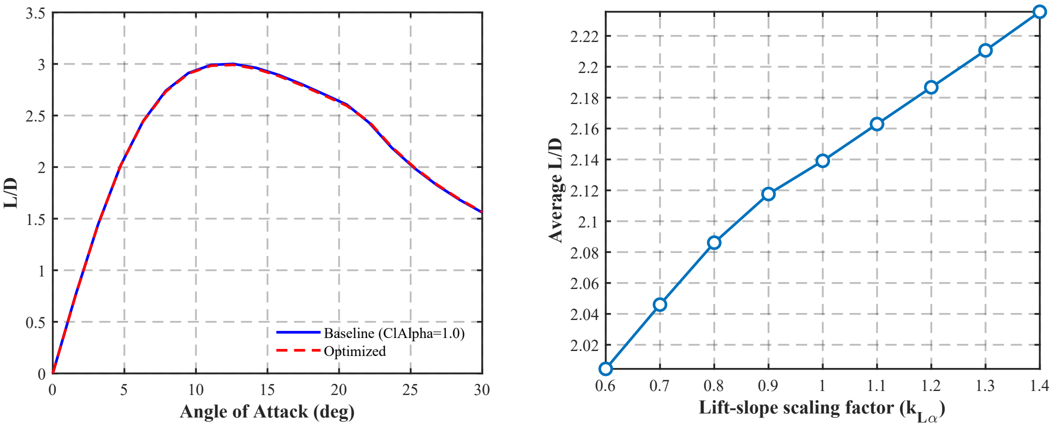
The response of mean $L/D$ to $k_{L\alpha}$ is nearly linear within the neighborhood of the baseline and saturates as the factor exceeds 1.1. Evaluated via centered finite differences ($\pm 0.1$), the absolute sensitivity is $S_{\text{abs}} = 0.225$ (L/D per unit multiplier). Given $J_0 = 2.1391$

and a reference scale $p_{\text{ref}} = 1$, the scaled sensitivity is $S_{\text{scaled}} = 0.105$. The smooth, weakly convex $J$–$k_{L\alpha}$ map confirms that while increasing the lift slope provides a direct benefit to trim authority, the aerodynamic advantage diminishes as the tail's contribution to total parasitic drag begins to offset the reduced trim drag.The corresponding statistical summary of the tail-lift-slope optimization results and sensitivity metrics is reported in Table 8.

**Table 8.** Tail-lift slope RL—statistical summary (three seeds).

| Metric | Baseline | Optimized (Mean ± Std) | Improvement (%) |
|---|---|---|---|
| Average $L/D$ | 2.1391 | 2.1424 ± 0.0074 | +0.15 |
| Mean optimal $k^*_{L\alpha}$ | | 1.014 | |
| Absolute sensitivity $S_{\text{abs}}$ | | 0.225 (L/D per $k_{L\alpha}$) | |
| Scaled sensitivity $S_{\text{scaled}}$ | | 0.105 (with $p_{\text{ref}} = 1$) | |

The corresponding $L/D$–AoA comparison and sensitivity curve are shown in Figure 15.



(**a**) (**b**)

**Figure 15.** Tail-lift slope RL—(**a**) $L/D$–AoA: baseline vs. optimized ($k^*_{L\alpha} = 1.014$); (**b**) sensitivity of $J$ to $k_{L\alpha}$.

### *6.8. Summary and Discussion*

Tables 9 and 10 summarize the results of all single-parameter reinforcement learning optimizations performed in Section 6.1 through Section 6.6. Table 9 reports the baseline and optimized performance indicators, together with the corresponding aerodynamic improvement percentages, while Table 10 separately reports the sensitivity information and the associated physical interpretation.

**Table 9.** Summary of single-parameter RL optimization outcomes (three-seed averages).

| Parameter | Symbol/Range | Baseline $J_0$ | Optimized $\bar{J}$ | Imp. (%) |
|---|---|---|---|---|
| Hull shape | 14D Bézier control points | 1.3949 | 1.6635 ± 0.1889 | +19.26 |
| Horizontal tail area | $k_A \in [0.5, 1.8]$ | 2.0669 | 2.1620 ± 0.0047 | +4.60 |
| Tail efficiency | $\eta_f \in [0.2, 0.6]$ | 2.1391 | 2.2116 ± 0.0031 | +3.39 |
| Tail-arm length | $\Delta x_t \in [-3, 3]$ m | 2.1391 | 2.1924 ± 0.0133 | +2.49 |
| Tail incidence angle | $i_t \in [-3°, 3°]$ | 2.1391 | 2.1426 ± 0.0012 | +0.17 |
| Tail-lift slope factor | $k_{L\alpha} \in [0.6, 1.4]$ | 2.1391 | 2.1424 ± 0.0074 | +0.15 |

**Table 10.** Sensitivity interpretation of the single-parameter RL studies.

| **Parameter** | **Sensitivity Measure** | **Physical Interpretation** |
|---|---|---|
| Hull shape | Dominant improvement contribution | Primary driver of global aerodynamic efficiency; controls pressure recovery and form drag. |
| Horizontal tail area | $S_{\text{abs}} \approx 0.4900$ per unit $k_A$ | Strong positive effect until saturation; further enlargement increases parasitic friction drag. |
| Tail efficiency | $S_{\text{abs}} \approx 0.32$ per unit $\eta_f$ | Improves effective tail authority and reduces trim drag, but represents a surrogate structural/wake-interference variable. |
| Tail-arm length | $S_{\text{abs}} = 0.1955$ L/D per metre | Aft shifts increase moment leverage, but excessive shifts introduce structural and trim-related penalties. |
| Tail incidence angle | $S_{\text{abs}} = 0.0212$ L/D per degree | Fine-tuning parameter; the baseline is already close to the local optimum. |
| Tail-lift slope factor | $S_{\text{abs}} = 0.225$ per unit multiplier | Minor correction to tail responsiveness; aerodynamic benefit is limited near the baseline. |

It should be noted that the hull-shape optimization and the tail-parameter optimizations are evaluated under different aerodynamic settings. The hull-shape study is performed in a bare-hull virtual wind tunnel configuration without longitudinal trim, in order to isolate the intrinsic aerodynamic contribution of the envelope geometry. In contrast, the tail-related parameters are evaluated with the full airship configuration using the longitudinal trim-related aerodynamic evaluation introduced in the corresponding subsections. Therefore, the absolute values of $\overline{L/D}$ should not be used for direct comparison between the hull and tail experiments. Instead, the percentage improvements are used as dimensionless indicators of the relative aerodynamic potential of each design parameter.

It is crucial to note that $S_{\text{abs}}$ inherently carries the physical units of the associated parameter (e.g., $L/D$ per metre, per degree, or per scaling factor).

### 6.8.1. Dimensionless Ranking and Global Trends

Because the absolute sensitivities ($S_{\text{abs}}$) possess distinct physical dimensions, direct numerical comparison across different parameters is physically ill-posed. Instead, Figure 16 evaluates the unified aerodynamic importance of each parameter by ranking their maximum achievable percentage improvements.

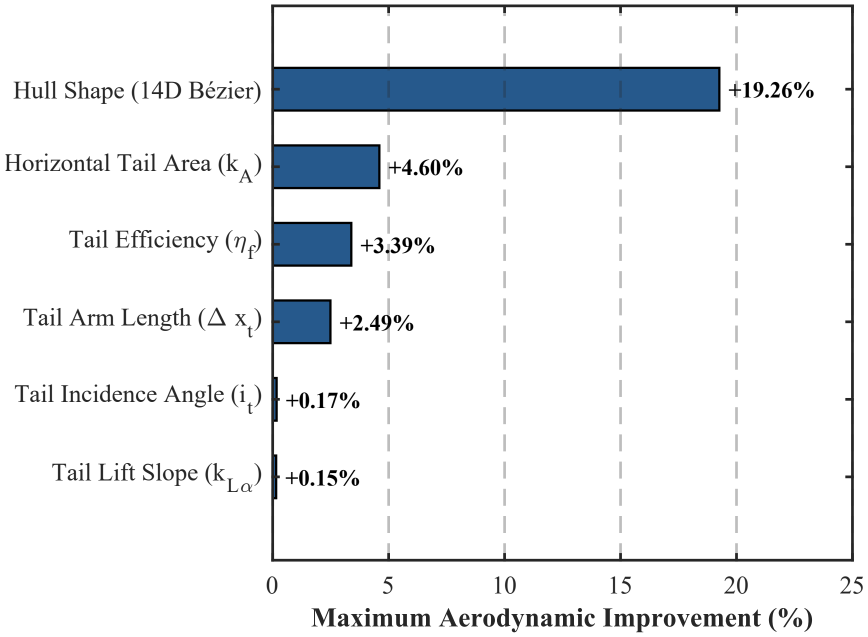


**Figure 16.** Unified aerodynamic improvement ranking across all RL-optimized parameters (dimensionless percentage).

Across all isolated experiments, the hull geometry unequivocally emerges as the dominant contributor to aerodynamic efficiency, dictating the primary optimization frontier with a near 20% potential gain. In contrast, empennage-related parameters yield comparatively constrained but physically critical modifications. Enlarging the horizontal tail area and manipulating the tail-arm length/efficiency factor provide modest $L/D$ gains (roughly 2% to 5%) by reducing the required trim drag. Meanwhile, the incidence angle and lift slope operate deep within the fine-tuning regime (<0.2% gain), confirming that the initial engineering baseline for these variables was already situated near the local aerodynamic optima.

### 6.8.2. Neutral or Non-Influential Parameters

Throughout the exploratory phase, it was also determined that vertical tail area scaling and minor vertical/longitudinal empennage displacements have no measurable positive effect on the steady longitudinal $L/D$. The vertical tail acts primarily as a parasitic drag source under symmetric zero sideslip flight; thus, its isolated optimization yields an immediate boundary collapse (minimization) or near-zero gradients. Such parameters were deliberately excluded from the optimization ranking to maintain focus on longitudinally active variables.

### 6.8.3. Discussion and Transition

The comprehensive single-parameter analyses yield three pivotal insights:

1. **Hull dominance vs. tail regulation:** The hull shape dictates the absolute baseline of aerodynamic efficiency, whereas the tail configuration acts as an active regulator, balancing lift recovery against trim-induced drag.
2. **Trade-off saturation:** Parameters such as tail-arm length and efficiency do not exhibit monotonic unbounded growth; they saturate rapidly due to the coupled penalties of structural limits, moment imbalances, and excessive wetted area.
3. **The necessity of integration:** The single-parameter RL agents repeatedly halted at local operational frontiers. Because aerodynamic forces (hull lift) and trim requirements (tail moments) are fundamentally coupled, isolated tuning has reached its theoretical limit.

These findings directly motivate the transition to the integrated methodology presented in Section 7. By coupling the hull geometry and empennage configuration into a unified, multi-dimensional action space, the subsequent framework will exploit the synergistic trade offs that isolated optimizations inherently overlook.

## 7. Multi-Parameter Reinforcement Learning Optimization of the Airship Tail Assembly

### *7.1. Motivation and Link to Section 6*

Building upon the single-parameter sensitivity studies presented in Section 6, this section investigates the *coupled optimization* of the tail subsystem through multi-action reinforcement learning (RL). The preceding analyses revealed that while the airship's hull geometry dominates aerodynamic performance, several tail parameters—notably the horizontal-tail area scaling factor $k_A$, tail-arm shift $l_t^{\text{shift}}$, efficiency factor $\eta_f$, and tail lift-curve slope multiplier $k_{L\alpha}$—exhibit meaningful, albeit secondary, optimization potentials (see Table 9).

However, those experiments treated each variable independently, inherently neglecting nonlinear interactions. In reality, these four parameters jointly govern static stability, lift recovery, and induced drag through strongly coupled fluid dynamic mechanisms. Hence, their isolated optimization cannot fully exploit synergistic benefits. This physical reality motivates the present multi-parameter RL framework, in which all four design variables are optimized *simultaneously* under a unified, physics-informed reward formulation derived from the dimensionless sensitivity ranking established in Section 6.

## *7.2. Environment and Reinforcement Learning Setup*

### 7.2.1. Design Variables and Selection Rationale

The DDPG agent controls four continuous variables simultaneously,

$$\boldsymbol{a} = \left[l_t^{\text{shift}}, k_A, k_{L\alpha}, \eta_f\right]^\top,$$

representing the longitudinal tail-arm shift, horizontal tail area ratio, tail lift curve slope scaling, and tail efficiency factor, respectively.

The strategic selection of these specific four parameters is deeply grounded in flight dynamics. While the tail incidence angle ($i_t$) yielded a marginally higher isolated gain than the lift slope ($k_{L\alpha}$) in Section 6, it was deliberately excluded from the coupled action space. Physically, $i_t$ acts merely as a linear pitch trim offset that is highly redundant with the elevator deflection under trimmed conditions. In contrast, $k_{L\alpha}$ acts as a surrogate for the fundamental aerodynamic hardware (e.g., airfoil profile and effective aspect ratio). It establishes a multiplicative synergy with the tail area ($k_A$), arm length ($l_t^{\text{shift}}$), and wake efficiency ($\eta_f$) to nonlinearly reshape the longitudinal stability margin. Furthermore, the vertical tail area was excluded since it serves purely as a parasitic drag contributor under the symmetric zero sideslip assumption.

The normalized action space $a_i \in [-1, 1]$ is linearly mapped to the physical intervals defined in Section 6

$$l_t^{\text{shift}} \in [-3, +3]\ \text{m}, \qquad k_A \in [0.5, 1.8],$$
$$k_{L\alpha} \in [0.6, 1.4], \qquad \eta_f \in [0.2, 0.6].$$

### 7.2.2. Aerodynamic Evaluation

Each episode performs a single trimmed evaluation via the aerodynamic solver. The environment returns the mean lift-to-drag ratio $\overline{L/D}$, the pitching-moment history $M_y(\alpha)$, and the number of numerical warnings $N_{\text{warn}}$.

### 7.2.3. Sensitivity-Weighted Reward Function

To mathematically embed the physical insights from Section 6 into the RL agent's objective, the reward is structured as a combination of a **sensitivity weighted performance term** and a set of uniform stability penalties

$$r = W_{\text{sens}} \cdot \overline{L/D} - W_{\text{dev}} \cdot P_{\text{param}} - W_{M_y} \cdot \frac{\overline{|M_y|}}{100\ \text{N} \cdot \text{m}} - W_{\text{warn}} \cdot N_{\text{warn}}, \tag{27}$$

where the fixed penalty weights are established as $W_{\text{dev}} = 0.02$, $W_{M_y} = 0.02$, and $W_{\text{warn}} = 0.01$.

(1) Penalty-weight selection.

The penalty weights in Equation (27) were selected through preliminary reward-scaling calibration rather than through an analytical derivation. The objective of this calibration was to keep the mean lift-to-drag ratio as the dominant term in the reward while preventing the agent from exploiting unrealistic parameter excursions, large residual pitching moments, or numerically unreliable evaluations. The parameter-deviation penalty $P_{\text{param}}$ is normalized by the admissible range of each design variable; therefore, $W_{\text{dev}} = 0.02$ provides only a mild regularization that discourages premature convergence to the parameter bounds without suppressing exploration. Similarly, the pitching-moment term is scaled by 100 N · m before multiplication by $W_{M_y} = 0.02$, so that it penalizes severe trim imbalance while remaining secondary to the aerodynamic performance term. Finally, $W_{\text{warn}} = 0.01$ assigns a small but nonzero cost to numerical warnings, discouraging the agent from repeatedly selecting configurations associated with unreliable solver behavior. These values were retained because preliminary training runs showed stable convergence, bounded parameter evolution, and no collapse toward either excessive constraint violation or overly conservative exploration.

(2) Physics-informed weighting heuristic.

The relative weights used in this stage are derived from the dimensionless maximum aerodynamic improvements (Δ%) observed in the single-parameter studies of Section 6. This choice avoids the direct summation of absolute sensitivities, which would be dimensionally inconsistent because the corresponding parameters have different physical units. Based on the isolated gains ($\Delta_{l_t} = 2.49\%$, $\Delta_{k_A} = 4.60\%$, $\Delta_{k_{L\alpha}} = 0.15\%$, $\Delta_{\eta_f} = 3.39\%$), the normalized heuristic weights are defined as

$$\begin{aligned} w_{l_t} &= \tfrac{2.49}{2.49+4.60+0.15+3.39} \approx 0.234, \\ w_{k_A} &= \tfrac{4.60}{2.49+4.60+0.15+3.39} \approx 0.433, \\ w_{k_{L\alpha}} &= \tfrac{0.15}{2.49+4.60+0.15+3.39} \approx 0.014, \\ w_{\eta_f} &= \tfrac{3.39}{2.49+4.60+0.15+3.39} \approx 0.319. \end{aligned}$$

These weights should be interpreted as a physics-informed heuristic for encoding the isolated aerodynamic relevance of each parameter, rather than as a mathematically rigorous measure of its importance in the fully coupled nonlinear design space. In particular, a large single-parameter improvement does not necessarily imply an equally dominant role once aerodynamic coupling, trim penalties, and parameter interactions are introduced. Therefore, the weights are used only to guide the reward scaling and to support the physical interpretation of the multi-parameter search, not to prove optimality of the selected weighting strategy.

(3) Uniform deviation penalty.

To prevent premature convergence towards unrealistic extrema, a uniform parameter-deviation penalty promotes balanced multidimensional exploration

$$P_{\text{param}} = \frac{1}{4}\sum_{i=1}^{4} \frac{|a_i - a_{i,0}|}{a_{i,\max} - a_{i,\min}},$$

where $a_{i,0}$ denotes the baseline values $(0, 1.0, 1.0, 0.29)$.

(4) Stability constraints.

The remaining terms ($W_{M_y}$ and $W_{\text{warn}}$) penalize severe aerodynamic imbalance and flow solver convergence failures, implicitly constraining the agent to naturally trimmable states.

(5) Sensitivity to the weighting choice.

The proposed weighting strategy should be regarded as a physics-informed initialization of the multi-parameter reward rather than as a unique optimal weighting law. Since the dominant objective remains the evaluated mean lift-to-drag ratio, alternative choices such as uniform weights would also constitute a valid reward design. Based on the single-parameter trends reported in Section 6, a uniform weighting scheme would be expected to preserve the same qualitative tendency toward increasing tail area and improving tail efficiency, although the exact converged parameter values may differ because of nonlinear aerodynamic coupling and trim-related penalties. Therefore, the present weighting strategy is used primarily to improve interpretability and guide exploration, while a systematic ablation study comparing sensitivity-based, uniform, and alternative weighting schemes is left for future work.

#### 7.2.4. Learning Framework

The DDPG agent utilizes twin 64-neuron hidden layers for both the actor and critic networks. Learning rates are set to $10^{-4}$ (actor) and $10^{-3}$ (critic) with a discount factor $\gamma = 0.99$. Training runs comprise 1500 episodes, and all reported metrics are averaged over three distinct random seeds.

### *7.3. Results and Convergence Behavior*

#### Reward Convergence

Figure 17 illustrates the smoothed reward evolution. A monotonic ascent is observed, with the learning process achieving a stable plateau. No oscillatory collapse occurs, confirming that the physics-informed reward landscape successfully guides the agent through the coupled parameter space.

#### Optimized Parameters and Aerodynamic Improvement

The multi-action RL optimization converges to the following configuration:

$$[l_t^{\text{shift}}, k_A, k_{L\alpha}, \eta_f] = [-0.390\ \text{m}, 1.177, 1.048, 0.422].$$

This configuration features a forward shift of the tail arm, an expansion in tail area, a minor increase in lift slope, and enhanced tail efficiency. Collectively, these coupled modifications elevate the average lift-to-drag ratio from the 2.1478 baseline to **2.3191** ± **0.0029**. This constitutes a substantial **+7.98% improvement**.

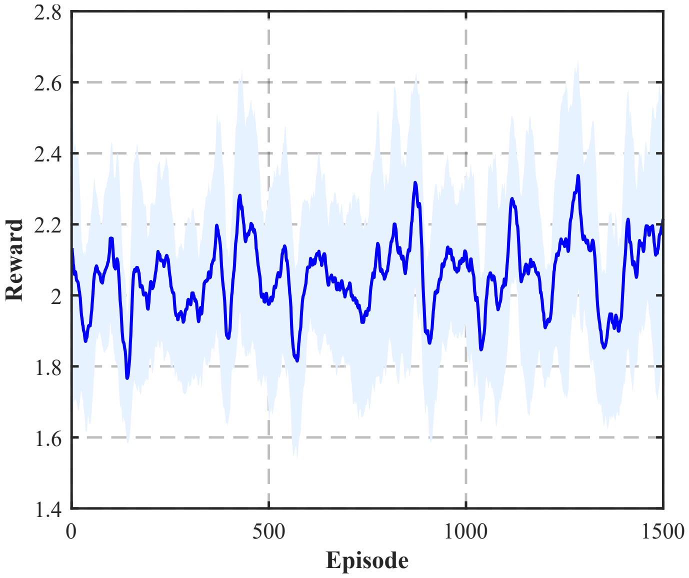


**Figure 17.** Reward convergence of the multi-parameter tail optimization (mean ± std).

As depicted in Figure 18, the optimized configuration provides consistent $L/D$ improvement over the prescribed evaluated angle-of-attack grid, particularly within the operationally emphasized range $0^\circ \leq \alpha \leq 15^\circ$.

Figure 19 visualizes the relative parameter variations from their respective baselines.

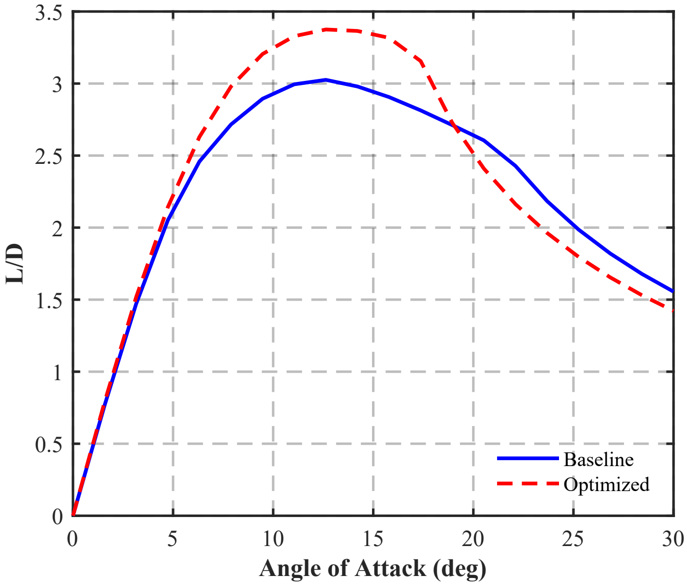


**Figure 18.** Baseline vs. optimized lift-to-drag ratio curves for the coupled tail optimization.

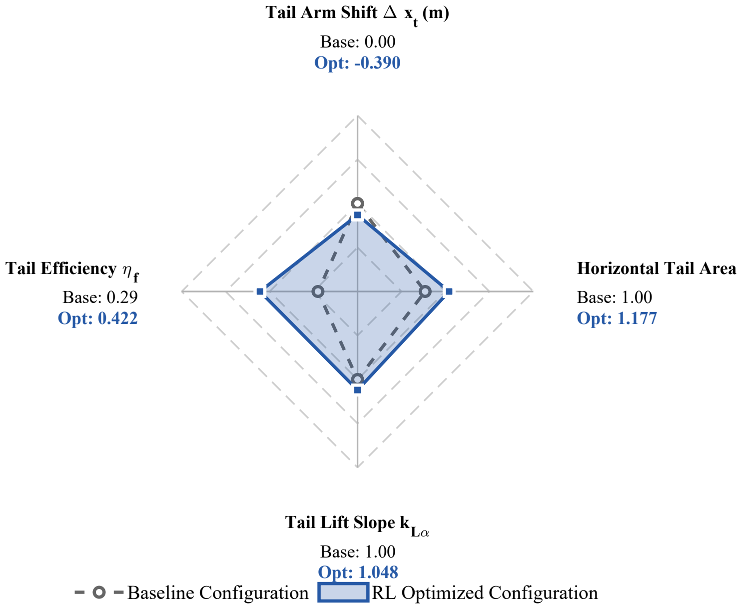


**Figure 19.** Optimized tail parameter distribution relative to bounds.

## *7.4. Discussion*

### 7.4.1. Synergistic Aerodynamic Effects

The integrated optimization yields a $+7.98\%$ improvement, which proves the presence of nonlinear interactions. For instance, the agent elected to shift the tail arm forward ($-0.390$ m), which structurally shortens the moment arm. To dynamically compensate and maintain trim authority, the agent simultaneously increased the tail area ($k_A = 1.177$) and enhanced wake efficiency ($\eta_f = 0.422$). This specific synergistic combination—a shorter but far more aerodynamically potent empennage—minimizes the massive wetted-area drag penalty associated with extending a long tail boom, a trade-off invisible during single-parameter sweeps.

### 7.4.2. Implications for Full System Integration

The success of this four parameter coupling validates the sensitivity-weighted reward formulation as a robust strategy for mapping isolated aerodynamic knowledge into a multi dimensional RL space. This provides the methodological foundation for the full airship optimization pursued in Section 8, where the optimized 14 parameters hull geometry from Section 6 will be coupled with these tail parameters to unlock the system's global aerodynamic optimum.

## 8. Integrated Reinforcement Learning Optimization of the Full Airship Configuration

### *8.1. Motivation and Problem Statement*

The preceding sections established a progressive reinforcement learning (RL) framework for aerodynamic optimization of the Lotte airship family. Section 6 demonstrated that the hull geometry dominates overall aerodynamic performance, achieving a significant lift-to-drag improvement through Bézier parameterized shape optimization. Section 7 subsequently optimized four key tail parameters on the original hull configuration, confirming the presence of synergistic nonlinear interactions within the empennage.

However, these efforts were performed sequentially on separated subsystems. In practice, hull and tail aerodynamics are strongly coupled: the heavily modified pressure field and retarded wake distribution induced by the newly optimized hull directly alter the tail's effective angle of attack, operational dynamic pressure, and stabilizing leverage. Consequently, a physically complete design strategy must evaluate and re-adapt the tail configuration to match the new fluid environment dictated by the optimized hull. This motivates the present section, which integrates the previously optimized hull geometry from Section 6 as the new baseline for retraining the four tail parameters—thereby realizing a true system-level sequential co-optimization of the full airship.

### *8.2. Experimental Setup*

### 8.2.1. Baseline and Design Variables

The aerodynamic baseline for this stage consists of the Bézier-optimized hull from Section 6 combined with the default tail configuration ($l_t^{\text{shift}} = 0\,\text{m}$, $k_A = 1.0$, $k_{L\alpha} = 1.0$, $\eta_f = 0.29$). Due to the drastically altered hull wake, placing the original tail on the new hull results in a highly suboptimal fluid-structure interaction, degrading the initial steady-state longitudinal efficiency to $\overline{L/D}_{\text{initial}} = 2.1300$.

To recover and maximize system efficiency, the agent controls the same four continuous tail parameters as defined in Section 7

$$\boldsymbol{a} = [l_t^{\text{shift}}, k_A, k_{L\alpha}, \eta_f]^\top,$$

with identical physical operational ranges constraints.

### 8.2.2. Environment and Reward Formulation

The aerodynamic evaluator is utilized to assess the full airship. The reward structure retains the exact sensitivity weighted formulation established in Equation (27) of Section 7. This deliberate consistency ensures that the RL algorithm operates under the same robust, physics-informed penalty landscape, allowing for a fair evaluation of how the optimal policy shifts in response to the new hull geometry. The DDPG hyperparameter configuration remains identical to the previous section.

### *8.3. Results and Analysis*

#### 8.3.1. Learning Convergence

The reward evolution in Figure 20 shows a clear improvement trend during training, indicating that the agent progressively overcomes the initial aerodynamic mismatch between the optimized hull and the baseline empennage. However, unlike the lower-dimensional tail-only optimization cases, the full-airship case does not reach a perfectly flat reward plateau within the prescribed 1500 episodes. Instead, the curve approaches a slowly varying quasi-converged regime, with residual fluctuations and a mild upward trend still visible near the end of training.

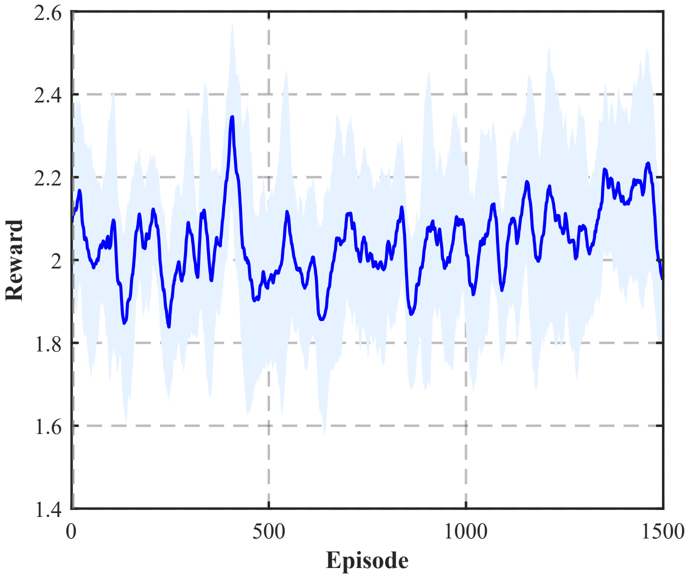


**Figure 20.** Reward evolution of the full airship (hull+tail) RL optimization (mean ± std).

This behavior is consistent with the higher-dimensional and more strongly coupled nature of the full-airship optimization problem. In this stage, the tail parameters are not optimized in isolation but must adapt to the modified wake and pressure field generated by the previously optimized hull. As a result, the reward landscape is more complex and contains coupled aerodynamic and trim-related penalties, making convergence slower than in the single-parameter or tail-only cases. Therefore, the final configuration reported here should be interpreted as a high-performing solution obtained within the selected training budget, rather than as evidence of complete asymptotic convergence.

#### 8.3.2. Optimized Configuration

The converged optimal tail parameters adapted for the new hull are

$$[l_t^{\text{shift}}, k_A, k_{L\alpha}, \eta_f] = [-0.436\ \text{m}, 1.124, 1.006, 0.420].$$

Compared to the tail optimization on the original hull in Section 7, this new configuration dictates an even more aggressive forward shift of the tail arm (−0.436 m) but requires a comparatively smaller increase in tail area (1.124). This indicates that the optimized hull geometry produces a cleaner, higher-energy wake, allowing the empennage to generate sufficient restoring moments without demanding excessive wetted area, thereby minimizing parasitic drag. The corresponding optimized parameter distribution is shown in Figure 21.

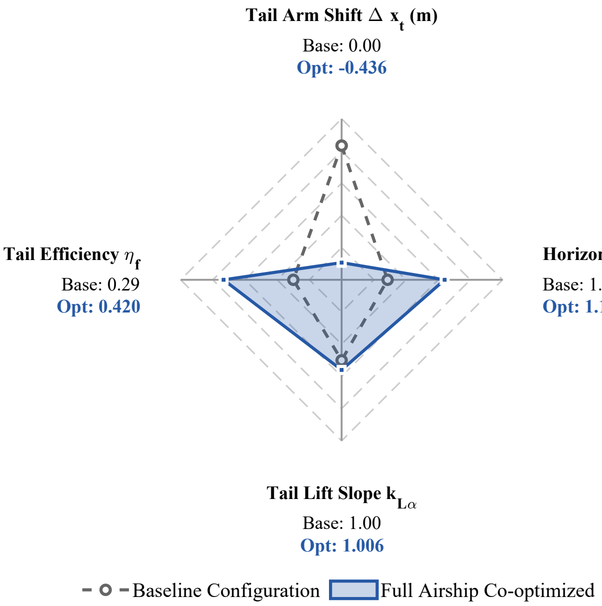


**Figure 21.** Optimized tail parameters for the full-airship RL co-optimization.

### 8.3.3. Aerodynamic Improvement and System-Level Gain

Using the optimized full airship configuration, the solver returns a stabilized efficiency of

$$\overline{L/D}_{\text{optimized}} = 2.2682 \pm 0.0068.$$

Relative to the mismatched initial state of this specific training run (2.1300), the tail re-adaptation achieves a solid **+6.49% improvement**. Figure 22 illustrates this consistent superiority over the prescribed evaluated angle-of-attack grid, with the main operationally emphasized range corresponding to $0^\circ \leq \alpha \leq 15^\circ$.

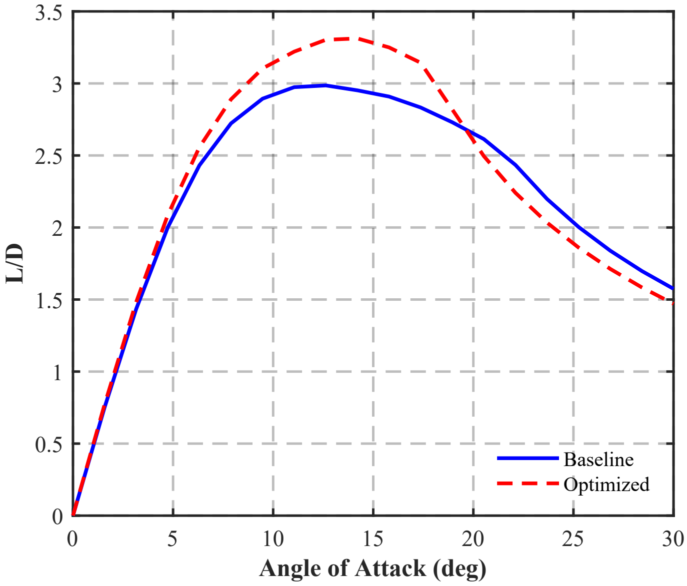


**Figure 22.** Comparison of baseline and optimized $L/D$–AoA curves for the full airship optimization.

More importantly, the true value of this framework is revealed when calculating the **global system-level aerodynamic gain**. By comparing the absolute performance of the ultimate fully-optimized airship against the original Lotte baseline from Section 6 ($\overline{L/D}_{\text{Lotte,orig}} = 2.0669$), the total integrated improvement is definitively established

$$\Delta J_{\text{total}} = \frac{2.2682 - 2.0669}{2.0669} = 0.0974. \tag{28}$$

The full system sequential RL optimization successfully extracts a **+9.74% absolute aerodynamic gain** over the mature, flight tested baseline design.

### *8.4. Discussion*

#### 8.4.1. Physical Coupling and Wake Interaction

The results decisively prove that tail optimization cannot be treated independently of the hull shape. The new optimal parameters reflect a deep physical adaptation: the RL agent recognized that the aerodynamically sleek hull (Section 6) altered the downwash and dynamic pressure at the tail station. It subsequently redesigned the tail to be more compact ($l_t < 0$, smaller $k_A$ relative to Section 7) but highly efficient ($\eta_f \approx 0.42$), perfectly matching the new flow topology.

#### 8.4.2. Implications for Design Methodology

The integrated reinforcement-learning framework demonstrated here provides a computationally tractable and physically interpretable workflow for the aerodynamic design of lighter-than-air vehicles. In the present implementation, the design problem is decomposed into a hull-shaping stage followed by a tail re-adaptation stage. This sequential strategy reduces the practical sampling burden of the RL training and, more importantly, allows the aerodynamic role of the hull and the compensating function of the empennage to be interpreted separately.

However, this sequential procedure should not be interpreted as evidence that an 18-dimensional simultaneous RL optimization is intractable. Modern continuous-action RL algorithms can in principle handle action spaces of comparable or even higher dimensionality. The present decomposition is adopted as a pragmatic and physics-informed design strategy, rather than as a mathematical proof of global optimality.

#### 8.4.3. Summary of Findings

1. Placing the original tail on the optimized hull creates an aerodynamic mismatch, underscoring the necessity of integrated co-optimization.
2. Re-optimizing the tail parameters for the new hull geometry recovers and enhances trim efficiency, providing an additional +6.49% gain over the mismatched state.
3. The ultimate, fully coupled airship configuration achieves a **+9.74% global aerodynamic improvement** over the original Lotte baseline (2.0669 → 2.2682).
4. The final parameters physically reflect a tailored adaptation to the improved hull wake, yielding a more compact and synergistically efficient empennage.

## 9. General Conclusions and Outlook

### *9.1. Synthesis of Research Findings*

This research has presented a complete reinforcement learning (RL) framework for the aerodynamic optimization of a lighter-than-air vehicle, progressively expanding from single-parameter exploratory studies to system-level co-optimization. By combining the Deep Deterministic Policy Gradient (DDPG) algorithm with steady-state aerodynamic evaluations via the SILCROAD solver, the study demonstrated that RL can autonomously discover physically consistent, high-efficiency configurations for complex airship geometries.

Specifically, the research proceeded through the following interconnected stages:

1. **Single-parameter Reinforcement Learning:** Individual aerodynamic parameters were optimized independently to quantify their isolated influence on the mean lift-to-drag ratio ($\overline{L/D}$). The hull shape, parameterized via a 14-degree Bézier curve, emerged as the dominant contributor, yielding a **+19.26%** improvement. Tail-related parameters (area, arm length, lift slope, and efficiency) provided secondary but physically interpretable improvements. Crucially, a dimensionless percentage improvement ranking was established to fairly evaluate parameter importance, avoiding the dimensionally ill-posed summation of absolute sensitivities.
2. **Multi Parameter Tail Optimization:** Guided by the established hierarchy, four key tail parameters ($l_t^{\text{shift}}, k_A, k_{L\alpha}, \eta_f$) were jointly optimized using a physics-informed, sensitivity-weighted reward. On the original hull geometry, this coupled approach achieved a **+7.98%** improvement, proving that the RL agent successfully exploited nonlinear aerodynamic synergies (e.g., compensating for a shorter tail arm with increased area and efficiency) that isolated tuning could not capture.
3. **Full Airship Integrated Co-optimization:** The optimized hull geometry from Section 6 was utilized as the new baseline. Recognizing the profound fluid dynamic coupling between the hull's wake and the empennage, the tail parameters were re-adapted to the new flow field. This sequential system-level strategy extracted an absolute aerodynamic gain of **+9.74%** over the mature, flight-tested Lotte baseline (2.0669 → 2.2682), delivering a globally optimized vehicle.

Overall, this work systematically applied the DDPG algorithm to airship aerodynamic optimization, covering the complete spectrum from single-variable analysis to coupled subsystem training and ultimately to the full-configuration reinforcement learning design. The sequential coupled strategy not only revealed the aerodynamic sensitivities of each design variable but also quantified their synergistic effects under a consistent physical framework.

### *9.2. Scientific and Methodological Contributions*

The key scientific and methodological contributions of this research can be summarized as follows:

- **Physics-informed RL Framework:** A novel RL environment was developed that directly interfaces with an aerodynamic solver. The reward functions uniquely incorporate structural deviation penalties, stability constraints, and sensitivity-derived normalizations, ensuring robust convergence and preventing unphysical geometric mutations.
- **Physics-informed Reward Weighting:** The research links single-parameter aerodynamic sensitivity studies with RL reward shaping by using dimensionless maximum improvement percentages ($\Delta\%$) rather than dimensionally heterogeneous absolute derivatives. This provides a physically interpretable heuristic for initializing and explaining the relative weighting of design variables in the coupled optimization stage. The resulting weights should not be interpreted as a rigorous proof of parameter importance in the fully nonlinear multi-parameter space, but as a practical mechanism for transferring isolated aerodynamic insight into the integrated RL framework.
- **Interpretable Sequential Co-design Workflow:** A two-stage sequential workflow is introduced, in which the hull geometry is first optimized, and the empennage parameters are subsequently re-adapted to the modified wake environment. This strategy provides a physically transparent way to analyze hull–tail coupling and reduces the practical sampling burden of the design study. However, it does not guarantee recovery of the global optimum of the fully coupled 18-parameter problem.

### 9.3. Physical Interpretation and Design Insights

The results demonstrate that RL, when strictly guided by physical constraints, does not merely operate as a mathematical "black box" but can autonomously rediscover and exploit complex fluid dynamic logic.

A profound design insight emerged during the transition from single to multi-parameter optimization. While isolated tuning suggested that lengthening the tail boom was necessary for stability, the multidimensional RL agent discovered an alternative, globally superior strategy: employing a shorter, more compact tail arm but drastically increasing the tail area and wake efficiency. This synergistic trade-off minimizes the massive wetted area drag associated with an extended tail boom while maintaining trim authority. Furthermore, the final system-level co-optimization proved that an aerodynamically sleek hull produces a higher energy wake, which in turn reduces the necessary size of the empennage required to achieve stability.

Although the sequential hull-then-tail optimization adopted in this work provides a physically interpretable and computationally efficient design workflow, it remains a greedy decomposition of the fully coupled problem. The method implicitly assumes that the hull geometry can be optimized first and that the tail can subsequently recover the major effects of the altered wake field. This assumption is useful for reducing the complexity of the design analysis, but it does not guarantee that the resulting configuration is the true global optimum of the simultaneous 18-parameter hull–tail optimization problem. In principle, a slightly less optimal hull in isolation could generate a wake environment that enables a more efficient empennage configuration, leading to a better total-system performance. Therefore, the final configuration reported here should be interpreted as the optimum of the proposed sequential co-design workflow, rather than as a mathematically proven global optimum of the full coupled design space. A direct comparison with a simultaneous 18-parameter RL optimization is left for future work.

### 9.4. Limitations and Assumptions

Although the proposed reinforcement-learning framework demonstrates a physically interpretable workflow for airship aerodynamic optimization, several limitations and assumptions should be explicitly acknowledged. First, the aerodynamic evaluations are performed using the mid-fidelity SILCROAD environment rather than high-fidelity Navier–Stokes CFD. Therefore, the optimized configurations should be interpreted as model-predicted improvements within the validity range of the underlying semi-empirical aerodynamic formulation, rather than as final high-fidelity-certified designs.

Second, the optimization is based on steady aerodynamic evaluations over a prescribed angle-of-attack grid. The bare-hull optimization stage adopts a virtual-wind-tunnel-type evaluation without enforcing longitudinal trim, in order to isolate the intrinsic aerodynamic effect of hull geometry. In contrast, the tail and full-airship stages employ an elevator-trimmed aerodynamic evaluation in which the elevator deflection is searched to reduce the pitching moment. This procedure enforces longitudinal pitch trim at each prescribed angle of attack, but it is not equivalent to a complete six-degree-of-freedom free-flight trim solution involving simultaneous force, moment, propulsion, and trajectory equilibrium.

Third, the performance metric used in this work is a robust averaged lift-to-drag indicator over a finite angle-of-attack range. It includes angle-of-attack weighting, numerical safeguards, and robustness-oriented penalty terms. Consequently, the reported improvements should be interpreted with respect to this specific objective definition, rather than as universal aerodynamic gains under all possible mission conditions. In addition, the lift-to-drag clipping used to avoid non-physical numerical spikes is a post-processing and reward-stabilization measure, not a physical aerodynamic limit.

Fourth, the design variables are reduced-order representations of the actual airship geometry and aerodynamic configuration. The Bézier hull parameterization, volume preservation, curvature regularization, tail-area scaling, lift-slope multiplier, and tail-efficiency factor are useful for systematic design exploration, but they do not capture all structural, manufacturing, aeroelastic, propulsion, and control-system constraints. In particular, the tail-efficiency factor is treated as a lumped surrogate for wake-interference and empennage effectiveness, rather than as an independently manufacturable design variable.

Finally, the proposed workflow follows a sequential hull-then-tail optimization strategy. This decomposition improves interpretability and reduces the practical sampling burden, but it does not guarantee recovery of the mathematical global optimum of the fully coupled hull–tail design space. The use of three random seeds mitigates stochastic variability, but a more extensive optimizer-benchmarking and ablation study would be needed to fully quantify sensitivity to initialization, reward weighting, and algorithm selection. These limitations define the scope of the present contribution and motivate the future extensions outlined below.

### *9.5. Future Research Directions*

Although the present framework successfully demonstrates the feasibility of RL-based steady-state aerodynamic optimization, several directions remain open for future investigation:

1. **High-fidelity aerodynamic assessment of optimized geometries:** Although SILCROAD has already been validated and widely employed in previous airship studies, future work may use high-fidelity Navier–Stokes CFD as an additional assessment tool for optimized geometries that are intended to progress toward detailed aerodynamic design. Such simulations could provide complementary information on surface pressure recovery, boundary-layer development, adverse pressure-gradient regions, and possible local separation effects around strongly modified hull shapes.
2. **Multi objective probabilistic optimization:** Future iterations could incorporate competing objectives (e.g., structural weight, internal gas volume, and manufacturing cost) into a Bayesian RL framework, ensuring design robustness under operational and manufacturing uncertainties.
3. **Active morphing control:** Transitioning from static geometric optimization to dynamic active control policies could enable the development of real-time morphing tail airships that autonomously reconfigure their shape based on varying flight phases.
4. **Reward weighting ablation studies:** A dedicated comparison among sensitivity-based, uniform, and alternative reward-weighting schemes would be valuable in future extensions of this work. Such an ablation would clarify the extent to which the final DDPG policy is influenced by the heuristic weights derived from the single-parameter studies, thereby providing a more complete assessment of the robustness of the coupled optimization framework.
5. **Ablation studies on staged learning strategy:** Another important extension concerns the role of the staged learning strategy itself. Future investigations should compare the proposed sequential workflow with direct multi-parameter optimization performed without prior sensitivity screening. To ensure a fair assessment, these comparisons should be conducted under identical training budgets, random seeds, design bounds, and convergence criteria, so that the effects on convergence stability, final aerodynamic performance, and physical interpretability can be quantified.

### *9.6. Concluding Remarks*

In conclusion, this research has firmly established that Deep Reinforcement Learning serves as a highly potent, physically aware engine for the system-level aerodynamic design of lighter-than-air vehicles. By systematically breaking down the design space—from isolated geometric sensitivities to coupled empennage dynamics, and finally to whole-vehicle fluid structure interaction—this work bridges classical aerodynamic theory with modern data-driven intelligence. The extraction of a verifiable **+9.74%** overall aerodynamic gain on an already mature engineering baseline represents a significant leap forward, offering a scalable methodology for the next generation of autonomous aerospace design.

**Author Contributions:** Conceptualization, Q.Z. and C.E.D.R.; methodology, Q.Z., C.E.D.R. and Y.Y.; software, Q.Z. and C.E.D.R.; validation, Q.Z. and C.E.D.R.; formal analysis, Q.Z. and C.E.D.R.; investigation Q.Z.; resources, Q.Z. and C.E.D.R.; data curation, Q.Z.; writing—original draft preparation, Q.Z.; writing—review and editing, Q.Z. and C.E.D.R.; visualization, Q.Z.; supervision, C.E.D.R.; project administration, C.E.D.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was partially supported by the China Scholarship Council (Support No. 202308100003), which provided the scholarship for the first Author.

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

## References

1. Riboldi, C.E.D.; Fanchini, L. Assessing the Technical–Economic Feasibility of Low-Altitude Unmanned Airships: Methodology and Comparative Case Studies. *Aerospace* **2025**, *12*, 244. [CrossRef]
2. Xie, J.; Huang, W.; Miao, J.; Li, J.; Cao, S. Off-Policy Deep Reinforcement Learning for Path Planning of Stratospheric Airship. *Drones* **2025**, *9*, 650. [CrossRef]
3. Wu, X. Modelling and Control of a Buoyancy Driven Airship. Ph.D. Thesis, École Centrale de Nantes, Nantes, France; South China University of Technology, Guangzhou, China, 2011.
4. Cimarelli, A.; Madonia, M.; Angeli, D.; Dumas, A. Aerodynamic Study of Advanced Airship Shapes. *J. Aerosp. Eng.* **2017**, *30*, 04016087. [CrossRef]
5. Stockbridge, C.; Ceruti, A.; Marzocca, P. Airship research and development in the areas of design, structures, dynamics and energy systems. *Int. J. Aeronaut. Space Sci.* **2012**, *13*, 170–187. [CrossRef]
6. Jones, R.T. *Classical Aerodynamic Theory*; NASA: Washington, DC, USA, 1979.
7. Riboldi, C.E.D.; Belan, M.; Cacciola, S.; Terenzi, R.; Trovato, S.; Usuelli, D.; Familiari, G. Preliminary sizing of a low-altitude airship including ion-plasma thrusters. In Proceedings of the 34th Congress of the International Council of the Aeronautical Sciences (ICAS 2024), Florence, Italy, 9–13 September 2024.
8. Riboldi, C.E.D.; Rolando, A. Layout analysis and optimization of airships with thrust-based stability augmentation. *Aerospace* **2022**, *9*, 393. [CrossRef]
9. Riboldi, C.E.D.; Rolando, A. Thrust-based stabilization and guidance for airships without thrust-vectoring. *Aerospace* **2023**, *10*, 344. [CrossRef]
10. Riboldi, C.E.D.; Rolando, A.; Regazzoni, G. On the feasibility of a launcher-deployable high-altitude airship: Effects of design constraints in an optimal sizing framework. *Aerospace* **2022**, *9*, 210. [CrossRef]
11. Munk, M.M. *The Aerodynamic Forces on Airship Hulls*; Technical Report Report No. 184; National Advisory Committee for Aeronautics (NACA): Washington, DC, USA, 1926.
12. Zhao, B.; Liu, D.; Luo, C. Reinforcement learning-based optimal stabilization for unknown nonlinear systems subject to inputs with uncertain constraints. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *31*, 4330–4340. [CrossRef] [PubMed]
13. Kornienko, A. System Identification Approach for Determining Flight Dynamical Characteristics of an Airship from Flight Data. Ph.D. Thesis, University of Stuttgart, Stuttgart, Germany, 2006.
14. Shields, K. CFD Applications in Airship Design. Ph.D. Thesis, West Virginia University Libraries, Beckley, WV, USA, 2010.
15. Li, R.; Zhang, Y.; Chen, H. Learning the Aerodynamic Design of Supercritical Airfoils Through Deep Reinforcement Learning. *AIAA J.* **2021**, *59*, 3988–4001. [CrossRef]
16. Kou, J.; Xiao, T. Artificial Intelligence and Machine Learning in Aerodynamics. *Metasci. Aerosp.* **2024**, *1*, 190–218. [CrossRef]
17. Ramos, D.; Lacasa, L.; Valero, E.; Rubio, G. Aerodynamic and structural airfoil shape optimisation via transfer learning-enhanced deep reinforcement learning. *arXiv* **2025**, arXiv:2505.02634.

18. Yu, Y.; Karimi, H.R.; Gelman, L.; Tian, J.; Mei, P. A novel multi-source sensor correlation adaptive fusion framework with uncertainty quantification for intelligent fault diagnosis. *Reliab. Eng. Syst. Saf.* **2025**, *267*, 111812. [CrossRef]
19. Yu, Y.; Karimi, H.R.; Shi, P.; Peng, R.; Zhao, S. A new multi-source information domain adaption network based on domain attributes and features transfer for cross-domain fault diagnosis. *Mech. Syst. Signal Process.* **2024**, *211*, 111194. [CrossRef]
20. Tian, J.; Yu, Y.; Karimi, H.R.; Gao, F.; Lin, J. A continual test-time domain adaptation method for online machinery fault diagnosis under dynamic operating conditions. *Neural Netw.* **2025**, *194*, 108192. [CrossRef]
21. Scavella, P.; Paolillo, G.; Greco, C. Deep reinforcement learning-based airfoil design and optimization: An aerodynamic analysis. *Aerosp. Sci. Technol.* **2025**, *167*, 110638. [CrossRef]
22. Perlini, A.; Abergo, L.; Gori, G. A Hybrid Bayesian-Adjoint Framework for Aerodynamic Shape Optimization. In Proceedings of the AIAA AVIATION 2024 FORUM, Las Vegas, NV, USA, 29 July–2 August 2024. [CrossRef]
23. Goldberg, D.E. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley: Boston, MA, USA, 1989.
24. Kennedy, J.; Eberhart, R. Particle swarm optimization. In *Proceedings of the ICNN'95—International Conference on Neural Networks*; IEEE: Piscataway, NJ, USA, 1995; Volume 4, pp. 1942–1948.
25. Jones, D.R.; Schonlau, M.; Welch, W.J. Efficient global optimization of expensive black-box functions. *J. Glob. Optim.* **1998**, *13*, 455–492. [CrossRef]
26. Riboldi, C.E.D.; Rolando, A. Autonomous Flight in Hover and Near-Hover for Thrust-Controlled Unmanned Airships. *Drones* **2023**, *7*, 545. [CrossRef]
27. Riboldi, C.E.D.; Belan, M.; Cacciola, S.; Terenzi, R.; Trovato, S.; Usuelli, D.; Familiari, G. Preliminary Sizing of High-Altitude Airships Featuring Atmospheric Ionic Thrusters: An Initial Feasibility Assessment. *Aerospace* **2024**, *11*, 590. [CrossRef]
28. Garcia, F.; Rachelson, E. Markov decision processes. In *Markov Decision Processes in Artificial Intelligence*; John Wiley & Sons: Hoboken, NJ, USA, 2013; pp. 1–38.
29. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous control with deep reinforcement learning. *arXiv* **2015**, arXiv:1509.02971.
30. Fujimoto, S.; Hoof, H.; Meger, D. Addressing function approximation error in actor-critic methods. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 1587–1596.
31. Haarnoja, T.; Zhou, A.; Abbeel, P.; Levine, S. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 1861–1870.
32. Alfeld, P.; Neamtu, M.; Schumaker, L.L. Circular Bernstein-Bezier polynomials. In *Mathematical Methods for Curves and Surfaces*; Vanderbilt University Press: Nashville, TN, USA, 1995; pp. 11–20.