



City Research Online

City St George's, University of London

Citation: Moslem, F., Jebelli, M., Masdari, M., Askari, R. & Ebrahimi, A. (2025). Deep Reinforcement Learning for Active Flow Control in Bluff Bodies: A State-of-the-Art Review. *Ocean Engineering*, 327, 120989. doi: 10.1016/j.oceaneng.2025.120989

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://city-ref-test.eprints-hosting.org/id/eprint/34814/>

Link to published version: <https://doi.org/10.1016/j.oceaneng.2025.120989>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Deep Reinforcement Learning for Active Flow Control in Bluff Bodies: A State-of-the-Art Review

Foad Moslem ^a, Mohammad Jebelli ^b, Mehran Masdari ^{c*}, Rasoul Askari ^d, Abbas Ebrahimi ^b

^a University of Tehran, Tehran, Iran

^b Sharif University of Technology, Tehran, Iran

^c City St George's, University of London, London UK

^d College of Interdisciplinary Science and Technology, University of Tehran, Tehran, Iran

Corresponding author: Mehran Masdari

Mailing address:

City St George's, University of London, London UK

E-mail address: mehran.masdari@citystgeorges.ac.uk

1 **Abstract**

2 Active flow control (AFC) of bluff bodies is a critical concept in a variety of engineering
3 applications, ranging from structural safety to clean energy harvesting. The current study
4 reviews the application of deep reinforcement learning (DRL) in the AFC of bluff bodies with a
5 focus on key studies between 2017 and 2024. DRL has improved AFC capabilities by enabling
6 real-time optimization and adaptive control. The review showed a focus on a single circular
7 cylinder, generally simulated two-dimensionally at a low Reynolds number of 100, using CFD
8 tools like OpenFOAM and FEniCS. Synthetic jets are favored actuators and Proximal Policy
9 Optimization (PPO), paired with deep neural networks, is the most widely used DRL algorithm.
10 Despite significant advances, key challenges still remain in the field. Scalability to higher
11 Reynolds numbers and turbulent flows, high computational cost, and sensor optimization
12 issues are found as the main challenges. The review also identifies inconsistent reporting of
13 hyperparameters and varied training procedures, which may hinder reproducibility of works
14 for further steps. Future research should address these gaps to extend the applicability of DRL
15 in practical AFC systems. This review also proposes a comprehensive framework to connect
16 initial simulation results with real-world applications.

17 **Keywords:**

18 Active flow control, Machine Learning, Deep reinforcement learning, Bluff Body

1. Introduction

The Control of the flow field has a long history dating back to the early steps in fluid mechanics. Various methods and approaches have been utilized in this field. Machine learning (ML) has opened new horizons recently. One important area of interest is flow control around bluff bodies; as the interaction between these bodies and the flow can result in different phenomena (Derakhshandeh and Alam, 2019). Flow-induced vibration (FIV) refers to a range of phenomena in which flow exerts fluctuating forces on the structure, leading to significant dynamic responses. A key subset, vortex-induced vibration (VIV), arises when flow around a bluff body generates alternating vortices and lift forces, resulting in vibrations. This phenomenon is crucial in engineering applications for both the safety of the structure and as a new source of clean energy (Francis and Swain, 2024; Gad-el-Hak, 2000; Jebelli and Masdari, 2022; Wang et al., 2020; Zhao, 2023).

Flow control, either passive or active, with various mechanisms (Figure 1), plays a key role in managing fluid dynamics and optimizing energy efficiency (Wang and Feng, 2018). Active and passive control methods are employed to control the bluff bodies like circular cylinders and wake flows (Chen et al., 2022). While active methods are generally more effective, they also require energy input. In contrast, passive methods are simpler but less adaptable (Rashidi et al., 2016) and show even contradictory behavior when faced with different conditions (Jebelli et al., 2024). The challenge is to find the most effective control strategy that balances performance and cost while minimizing environmental impact (Gad-el-Hak, 2000; Wang and Feng, 2018). However, active flow control (AFC) mechanisms are essential for cylindrical structures, such as platforms, pipelines, or riser pipes, mounted in water flow. Even minor vibrations caused by VIV may lead to a catastrophe (Zhao, 2023).

42 A typical AFC system, widely applied in fields such as aerospace, automotive engineering, and
43 energy production, includes a fluid system, an actuator, a sensor, and a control algorithm
44 (Figure 2). The actuator excites the flow, the sensor measures variables from the flow, and the
45 algorithm adjusts the actuator based on sensor feedback to optimize control (Wang and Feng,
46 2018).

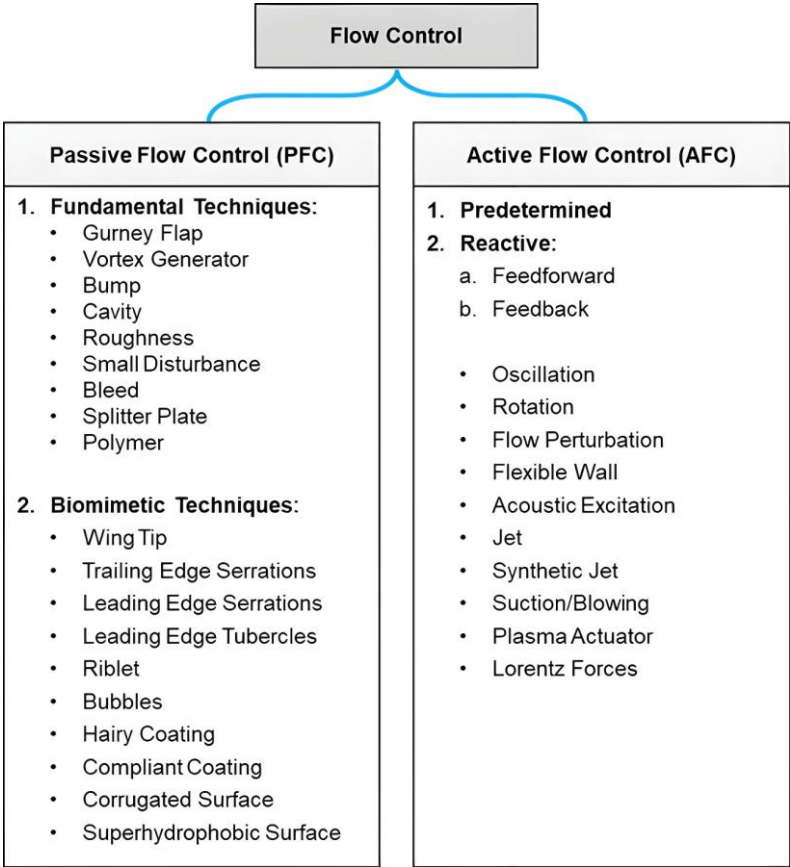


Figure 1) Classification of Passive and Active Flow Control Strategies

47 Flow control has gone through a few distinct eras: empirical before 1900, scientific between
48 1900 and 1940, WWII and Cold War with military objectives between 1940 and 1970, energy
49 crisis between 1970 and 1990, and climate change starting in the 1990s (Gad-el-Hak, 2000).
50 With the incorporation of ML, flow control has entered a new age, with the possibility for real-
51 time optimization, adaptive control, and novel control algorithms (Ren et al., 2021a).

In conclusion, flow control has evolved to meet global challenges like energy efficiency and sustainability. Now, ML integration may offer new possibilities to tackle these issues. While significant research has focused on controlling the flow around bluff bodies, there has been relatively little work on integrating Reinforcement Learning (RL), a subset of ML, into AFC for bluff bodies.

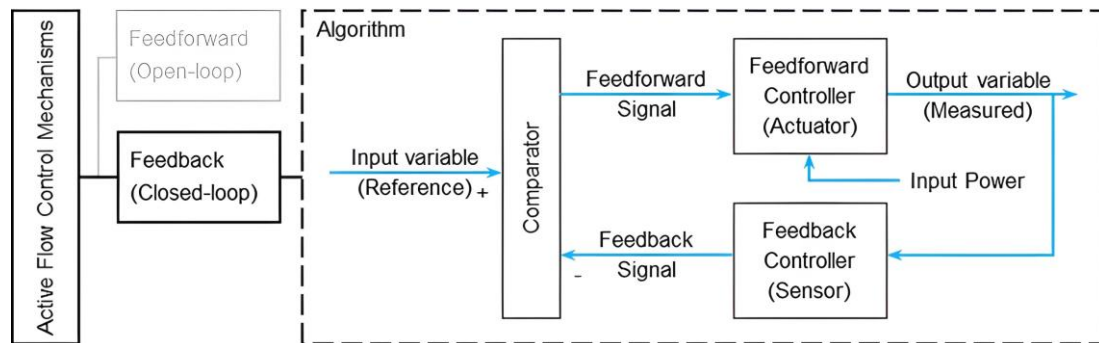


Figure 2) A typical active flow control system

This review aims to present the current state of the art researches on using RL for AFC of bluff bodies. To achieve this, Section 2 provides an overview of key reinforcement learning concepts and governing mechanisms. Section 3 then reviews recent studies employing DRL for AFC in bluff bodies. Section 4 categorizes relevant technical parameters, discusses existing research gaps and proposes a comprehensive framework for DRL-based AFC cases. Finally, concluded remarks are presented in Section 5.

2. Deep Reinforcement Learning for Active Flow Control

DRL has gained wide attention by breakthrough advances in AI and control. Among the breakthrough events in development of DRL is its beating humankind at the game of Go(Silver et al., 2016), implying an ability of DRL to come up with novel strategies that transcend human intuition and demonstrate its viability in complex decision-making tasks. Further, DRL has applied itself to high-dimensional control problems, such as the magnetic control of tokamak

69 plasmas, where it executed successful management of nonlinear dynamics (Degraeve et al.,
70 2022). All these are thereby indicative of the adaptability of the processes in DRL, thus making
71 their deployment for active flow control possible, as it has similar optimization and real-time
72 challenges.

73 In fluid mechanics, two fields are mainly studied for applying RL: shape optimization and flow
74 control (Vignon et al., 2023b). Within the former, a body is immersed in a flow, and the RL
75 algorithm modifies its shape to optimize some characteristics due to the dynamic of the fluid.
76 Flow control aims at actively influencing the behavior of a flow, such as controlling separation,
77 turbulence, or heat transfers, to reach a more desirable state.

78 DRL has emerged as a transformative approach for AFC, combining the strengths of advanced
79 RL techniques, artificial neural networks, and the complexities of fluid dynamics. Unlike
80 traditional control methods, DRL enables the development of intelligent agents capable of
81 learning and adapting to dynamic, nonlinear flow environments. Based on numerical
82 simulations and experimental data, DRL can optimize control strategies in real time for a
83 variety of objectives, including drag reduction, lift enhancement, and improved energy
84 efficiency. This section explores the principles, methodologies, and applications of DRL in AFC,
85 highlighting its potential to revolutionize flow control across diverse engineering domains.

86 2.1. Reinforcement Learning as a Control Method

87 Recently, AI has made significant advancements in fluid dynamics research, enabling more
88 advanced, adaptable, and user-friendly solutions (Frank et al., 2020). Among AI approaches,
89 ML uses statistical methods to develop data-driven algorithms, and RL, which is a subset of
90 ML, trains algorithms to make decisions that optimize specific objectives through iterative
91 interaction with their environment (“Part 2: Kinds of RL Algorithms — Spinning Up

documentation,” n.d.). The foundational work on RL, which combines trial-and-error methods with optimal control, began in the late 1980s (Sutton, 1988).

Unlike traditional control strategies, such as Proportional Integral Derivative (PID) controllers and Linear Quadratic Regulators (LQR), which rely on assumptions like linearity and local system analysis, RL aims to accomplish control without imposing constraints. This flexibility helps RL to adapt to complex, nonlinear, and dynamic systems (Rabault and Kuhnle, 2023). Table 1 compares RL with conventional control methods, showing its distinct advantages in terms of adaptability, scalability, and energy efficiency. For instance, while PID and LQR rely on predefined models or rules, RL dynamically learns optimal control policies through continuous interaction with the system. This interaction-based learning enables RL to excel in real-time adaptability and cumulative performance optimization, outperforming conventional approaches in many scenarios.

Table 1) RL vs Conventional Control Approaches. Based on (Rabault and Kuhnle, 2023)

Control Approach	PID	LQR	RL
Underlying Assumption	System linearity	Local system analysis	No assumptions about system linearity
Decision Making	Predefined rules (Feedback control)	Predefined optimization	Learns optimal policy through interaction
Adaptability	Low	Medium	High
Interaction Model	Fixed analytical model	Analytical model	Agent-environment interaction, black-box approach
Energy Efficiency	Moderate	Moderate	High (optimizes for cumulative reward)
Real-time Adaptability	Low	Low	High
Scalability	Moderate	Moderate	High

By addressing the limitations of traditional control methods, RL creates new opportunities for solving complex, real-world problems, particularly in fluid dynamics where adaptability and scalability are critical. It is worth mentioning that among non-RL machine learning methods, genetic programming (GP) (Koza, 1994), inspired from natural selection and evolutionary principles, has shown competitive performance in fluid dynamic problems. However, these methods fall outside the scope of this article.

2.2. Reinforcement Learning, Basic Concepts

In RL as an AFC method, an agent learns to make change the flow field through controlled interventions. In this process, the agent interacts with the flow environment by observing states, like velocity or pressure gradients, and makes adjustments to achieve desired outcomes, like reducing drag or enhancing lift.

The RL is based on Markov Decision Process (MDP) (Bellman, 1957), a mathematical framework used to model decision-making situations where outcomes are partly random and under the control of a decision-maker called agent. A MDP has four key components including: S , A , R and a transition probability function. The first two ones are sets of all valid states and actions of the system and R denotes to the set of all possible rewards. Finally, the transition probability function, $P(s' | s, a)$, is the probability of transitioning into the state s' if starting in state s and take action a . Then a policy for RL can be defined as:

$$\begin{aligned} \pi_{\theta}: S &\rightarrow A \\ s &\rightarrow \pi_{\theta}(s) \end{aligned} \tag{1}$$

Note that $\pi_{\theta}(s)$ returns the action probability distribution when the current state is s . The action is then chosen by the agent according to $\pi_{\theta}(s)$, where θ represents the set of

parameters that characterizes π (“Part 2: Kinds of RL Algorithms — Spinning Up documentation,” n.d.; Vignon et al., 2023b).

During RL process, an agent interacts with an environment in three steps. First, the environment provides a partial observation of the agent's state (s), which is necessarily incomplete. To address this, the state at time t (s_t) is defined as a series of past observations and actions: $s_t = (s_0, a_1, s_1, a_2, \dots, s_{t-1}, a_t)$. Second, the agent takes an action (a) based on the observation (o) of state (s) following a policy (π), which changes the environment, resulting in a new state (s_{t+1}). Finally, the agent receives a reward (r) based on the transition from the previous state to the new state (Vignon et al., 2023b) (Figure 3).

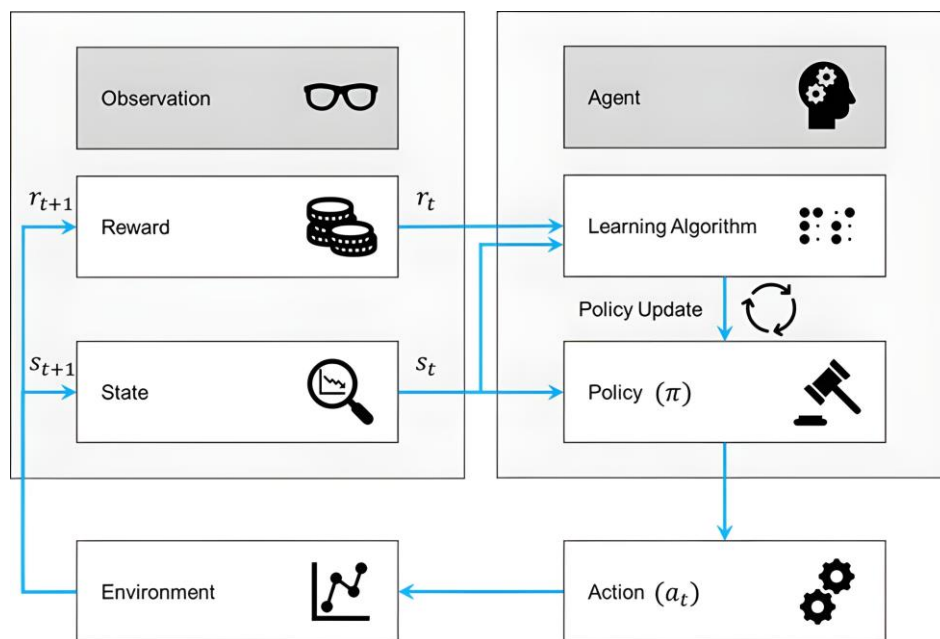


Figure 3) Reinforcement Learning Process Loop

The RL algorithm collects information and gradually improves the reward via the three steps interactions, which occur at each time step of a global process. To this purpose, the reward specifies the feature the user wishes to enhance. The RL algorithm typically maximizes a cumulative reward, represented by R (Eq.(2) rather than an instantaneous reward r_t at time t . T is the length of an episode, and γ is the discount factor (often, $\gamma \gtrsim 0.99$).

$$R = \sum_{t=0}^T \gamma^t r_t \quad (2)$$

The three-step process repeats at each time step t until t reaches T , marking the end of an episode. The cumulative reward R is calculated at the end of each episode, and the policy is adjusted to maximize R . Each episode includes $[T/t]$ actions. The discount factor γ^t is used to prioritize short-term rewards over long-term ones, allowing R to converge even with multiple actions. Instead of calculating the cumulative reward over the entire trajectory, it is often done starting from a specific time step t_0 , replacing (Eq.(1) with (Eq.(3).

$$R = \sum_{t=t_0}^T \gamma^{t-t_0} r_t \quad (3)$$

In summary, RL is able to provide a robust framework by enabling agents to interact with and adapt flow environments through iterative learning processes. By leveraging the MDP structure, RL algorithms optimize policies to maximize cumulative rewards, effectively addressing complex objectives.

2.3. Reinforcement Learning, Algorithms

Reinforcement learning algorithms can be categorized into two primary branches: "Model-Based" and "Model-Free" algorithms (Figure 4) ("Part 2: Kinds of RL Algorithms — Spinning Up documentation," n.d.; Rabault and Kuhnle, 2023). In model-based RL, the agent interacts with a model of the environment. Historically, linear models were employed in flow control due to their simplicity and computational efficiency in stabilizing unstable flows (Barbagallo et al., 2012; Dahan et al., 2012). However, their applicability in High Re numbers and to highly turbulent regimes remains limited because of the nonlinear nature of such flows. To address this challenge, nonlinear stochastic models have been developed, demonstrating improved

performance in stabilizing wakes and reducing drag in three-dimensional turbulent flows(Brackston et al., 2016). While model-based methods offer computational efficiency by simplifying the environment, they often yield sub-optimal results at high Re numbers due to incomplete turbulence modeling.

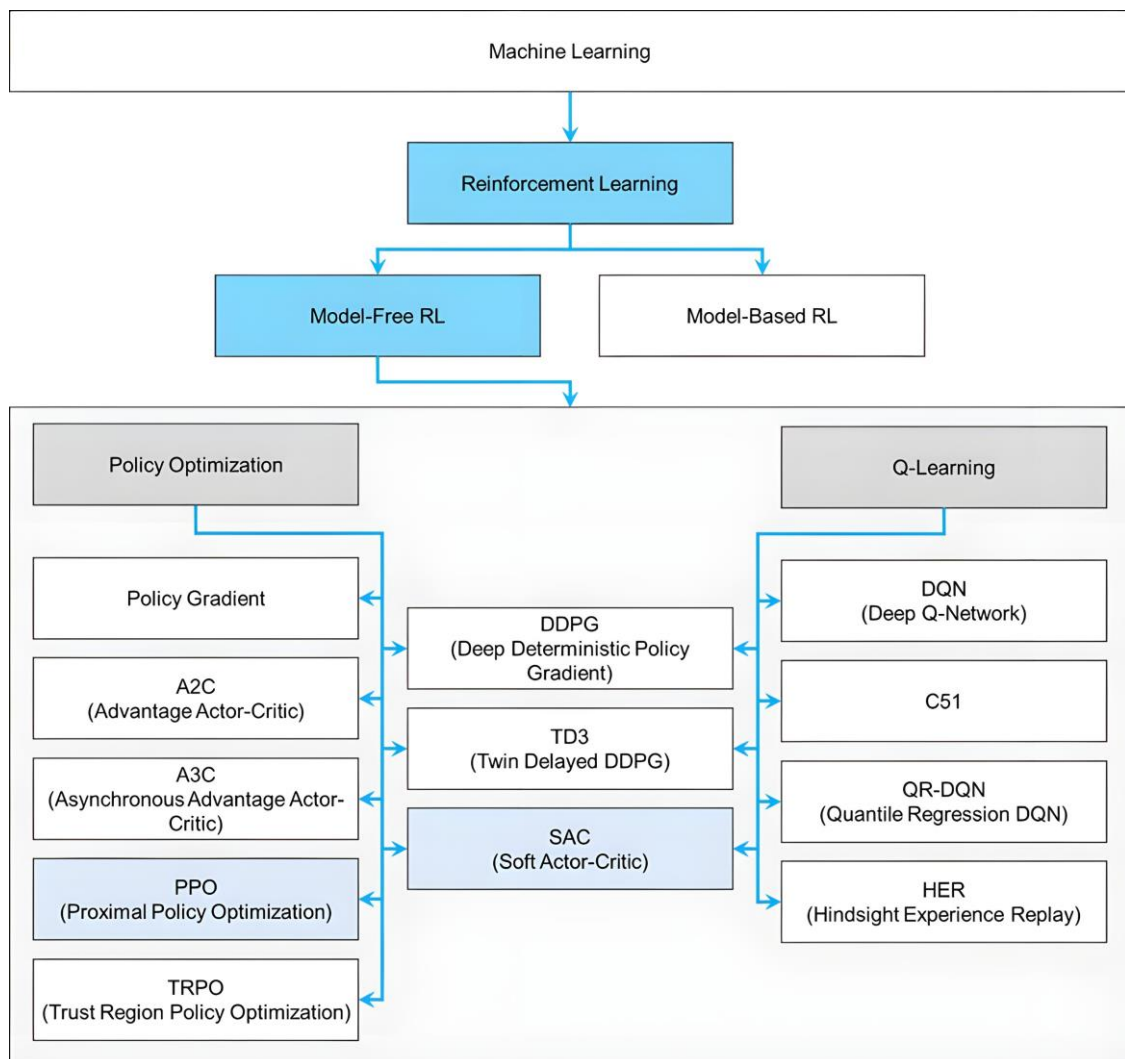


Figure 4) Overview of RL algorithms ("Part 2: Kinds of RL Algorithms — Spinning Up documentation," n.d.): divided into two primary branches: Model-Based RL and Model-Free RL

In contrast, model-free RL methods enable the agent to directly interact with the environment, receiving incomplete observations as inputs and producing actions in response. These methods are particularly suitable for turbulent or unsteady flows, which are highly nonlinear and challenging to model accurately. As model-free RL does not require an analytical description of the environment, it is often referred to as a "black-box" approach (Richard S.

Sutton, 2018). This characteristic makes it a preferred choice for AFC problems, where the flow behavior is inherently complex.

Model-free RL methods can be broadly divided into two main categories: Policy Optimization and Q-Learning. Policy Optimization focuses on directly optimizing a policy, which is a function mapping states to actions. The goal is to adjust the policy's parameters in a way that maximizes cumulative rewards over time. This approach is particularly effective in environments with continuous action spaces (like flow field), where discretizing the action space can lead to suboptimal control, precision loss, and lower efficiency. Algorithms such as Policy Gradient, A2C/A3C (Mnih, 2016), PPO (Schulman et al., 2017), and TRPO (Schulman, 2015) fall under this category. Among these, PPO is often preferred in AFC applications due to its balance between stability and adaptability. PPO maintains a stable training process while allowing sufficient flexibility for the policy to improve efficiently. On the other hand, Q-Learning focuses on estimating the optimal action-value function (Q-function). This function represents the expected return (or cumulative reward) from taking a specific action in a given state and then following the optimal policy thereafter. Q-Learning algorithms, such as DQN (Mnih, 2013), C51 (Bellemare et al., 2017), QR-DQN (Dabney et al., 2018), and HER (Andrychowicz et al., 2017) are primarily designed for discrete action spaces leading to overshooting or undershooting desired control targets and suboptimal flow behavior.

Actor-Critic methods aim to combine the strengths of both Q-Learning and Policy Optimization. These methods use two main components: an "actor", responsible for optimizing the policy, and a "critic", responsible for estimating the value function (how good a state or action is). This dual-structure approach enables Actor-Critic methods to handle complex tasks that involve continuous control, while also ensuring effective exploration and

stability during training. Prominent algorithms in this category include DDPG (Lillicrap, 2015), TD3 (Fujimoto et al., 2018), and SAC (Fujimoto et al., 2018). While DDPG and TD3 are effective in continuous control tasks, but they can suffer from issues like overestimation bias and sensitivity to hyperparameters. In contrast, SAC uses entropy regularization to encourage exploration through controlled randomness in actions. This improves sample efficiency and enhances learning stability. As a result, SAC is highly effective in AFC applications where precision and adaptability are essential.

Overall, while model-based RL offers computational efficiency and simplified representations, model-free RL, through Policy Optimization, Q-Learning, and Actor-Critic methods, provides the flexibility and adaptability required for controlling highly nonlinear flow systems effectively. A detailed discussion on the strengths and limitations of algorithms in the context of fluid mechanics and AFC can be found in the review articles, which is highly recommended for interested readers (Garnier et al., 2021; Vignon et al., 2023b).

2.4. Reinforcement Learning and Artificial Neural Networks

Traditional RL algorithms relied on simple function approximators such as matrices, polynomials, or basic mathematical functions to represent value or policy functions. While these methods were effective for problems with small state and action spaces, they faced scalability challenges when applied to more complex, high-dimensional problems. As the state-action space grows, these simplistic representations become inefficient and computationally impractical.

The concept of Artificial Neural Networks (ANNs) dates back to middle of 20th (McCulloch and Pitts, 1943), with the introduction of the McCulloch-Pitts neuron, a mathematical model, inspired by the structure and function of biological neural networks in the human brain. A

213 neuron is an entity characterized by three features: a set of weights corresponding to the
214 number of inputs connected to the neuron, a bias factor and the activation function. A neuron
215 takes a vector input and returns a scalar output (Figure 5).

216 ANNs consist of interconnected neurons, organized into layers (input, hidden, and output
217 layers), and are designed to process and learn from data by adjusting the weights of
218 connections between neurons. They are known to be universal function approximators
219 (Hornik et al., 1989), able to treat high-dimensional and non-linear problems, and thus to
220 overcome the limitations set by the simple function approximators available in traditional RL.

221 Different types of ANNs are used in fluid for tasks like state representation, policy
222 approximation, value estimation, and environment modeling. Convolutional Neural Networks
223 (CNNs) are commonly applied to extract spatial features from flow fields, such as velocity and
224 pressure maps, making them ideal for analyzing flow structures and guiding control actions
225 (Rabault et al., 2019; Zhang et al., 2024) which can also be beneficial in accelerating AFC
226 problems (Ye et al., 2024).

227 For time-dependent flow data, Recurrent Neural Networks (RNNs), including Long Short-Term
228 Memory (LSTM) networks, are effective at capturing patterns over time, especially from
229 sensor readings (Paris et al., 2021). To handle large and noisy CFD datasets, Autoencoders
230 reduce the data size while preserving key features, improving efficiency during training
231 (Morton et al., 2018). For more complex, irregular flow structures, Graph Neural Networks
232 (GNNs) are useful in modeling relationships between vortices and flow patterns (Chen et al.,
233 2021).

234 When neurons are organized in successive layers, calling Deep Neural Network (DNN), each
235 neuron of a layer is connected to one or more neurons of the precedent and next layers. When

each neuron of each layer is connected to all the neurons of the precedent and of the next layer, the network is said fully connected or FCNN (Figure 5). Over the last decade, RL algorithms have been enhanced by integrating DNNs with conventional RL approaches, evolving into DRL. This advancement overcomes the limitations of simple function approximators in traditional RL (Vignon et al., 2023b) and enables more precise changes in the flow environment(F. Xie et al., 2023).

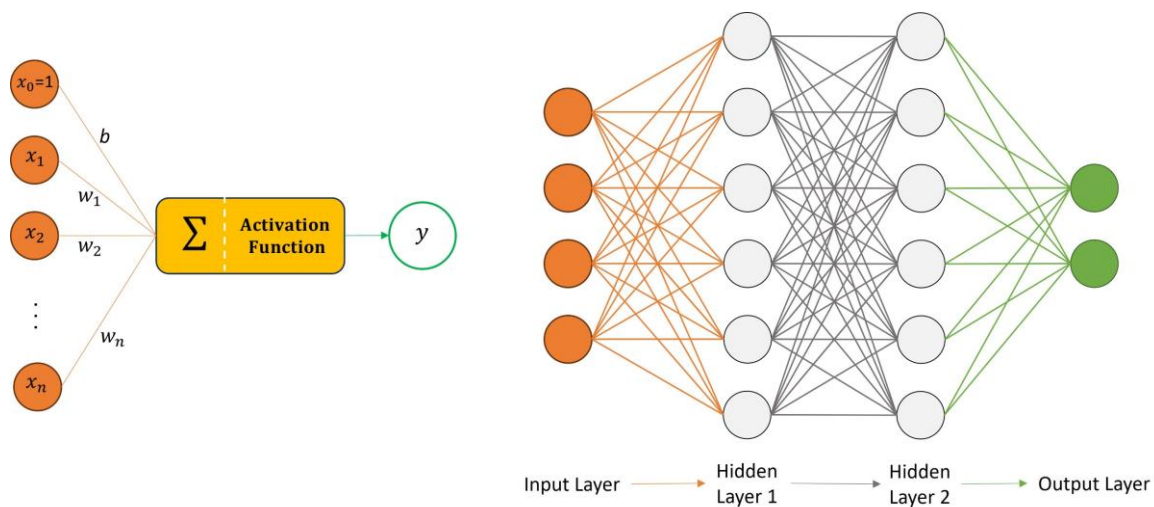


Figure 5) Schematic view of a single artificial neuron (left) and a fully connected artificial neural network with two hidden layers (right)

3. Combination of DRL and AFC of Bluff Bodies

Initial works on using reinforcement learning to control the flow around bluff bodies date back to the late 2010s. Pivot et al.(Pivot et al., 2017) presented the application of a continuous RL framework to control the drag force of a circular cylinder. Another notable early study by Koizumi et al.(Koizumi et al., 2018) employed DRL with Deep Policy Gradient to control vortex shedding from a circular cylinder. The work of Rabault et al. is the most well-known study on DRL for AFC around bluff bodies, reporting the first application of an artificial neural network trained with a DRL agent for AFC. In their study, which included a 2D simulation of a circular cylinder at $Re = 100$, the ANN effectively learned a control strategy by manipulating the mass

flow rates of two jets positioned on either side of the cylinder. By interacting with the unsteady wake, the agent successfully stabilized the vortex street and achieved an approximate 8% reduction in drag with minimal actuation effort (Rabault et al., 2019). Following the path, studies explored diverse approaches (Figure 6).

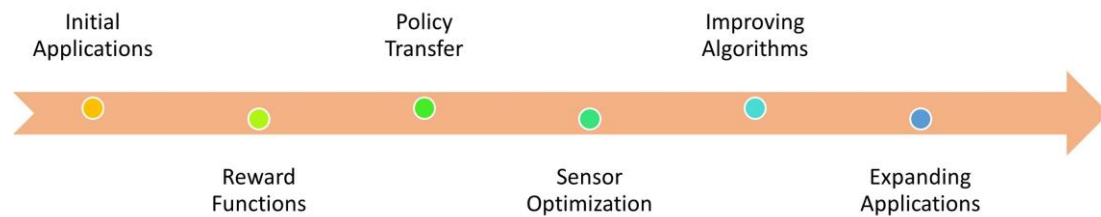


Figure 6) Application of DRL-based AFC for bluff bodies

Flow control and stabilization are considered the most attractive fields for researchers. To achieve this goal, changes in learning algorithms and reward functions have been more extensively investigated. In early studies, the reward functions were defined as a simple combination of lift and drag coefficient with the aim of minimizing the drag and controlling the lift fluctuations. Qin et al. proposed a reward function based on dynamic mode decomposition (DMD) applied to flow field measurements from simulations. They gave the zeroth mode of dynamic modes a larger value as the corresponding mode for flow stabilization. Using this data-driven reward, the DRL model learns an AFC policy that controls the synthetic jets' mass flow rate, resulting in a lower drag and larger recirculation zone (Qin et al., 2024). In another study, Jiang and Cao defined a reward function to decrease perturbation kinetic energy in the near wake and consequently reduce drag force. This reward function, accompanied by manipulating the cylinder oscillation in the transverse direction, changed the wake flow and resulted in a lower drag coefficient (Jiang and Cao, 2023). Focusing on the recirculation zone in the reward function is another approach to improve flow control

269 around a cylinder. Xu et al. defined their reward function based on drag fluctuations and
270 recirculation area resulting in effectively stabilized periodic vortex shedding (Xu et al., 2020).

271 While most of the studies use synthetic jets as the actuators of the control system, some
272 researchers tried other actuation systems. Tokarev et al. applied DRL to control the flow over
273 a cylinder and set its angular velocity as the control parameter. The neural network develops
274 a strategy based on low-frequency harmonic oscillations with additional modulations to
275 stabilize the vortex street, resulting in an acceptable drag reduction (Tokarev et al., 2020).

276 Tandem arrangement of cylinders can lead to significantly larger lift fluctuations. Applying DRL
277 to reduce lift forces on the aft cylinder by self-rotation shows even in the most challenging
278 configurations, the DRL agent devises an effective strategy to reduce the lift fluctuations (Zhao
279 et al., 2024). The rotation of the object as an actuation method is also used by Xie et al. to
280 control the flow around a downstream cylinder in tandem arrangement (Z. Xie et al., 2023).

281 Mounting two rotating control rods in the rear side of the cylinder is another approach used
282 to control the flow (Fan et al., 2020; Xu et al., 2020).

283 Yousif et al. focused on optimizing plasma actuators for drag reduction. They employed the
284 PPO algorithm to control plasma actuators based on the Dielectric-Barrier-Discharge principle.
285 The results showed significant drag reduction across various configurations, particularly in the
286 setup with three pairs of actuators on the rear and top-bottom surfaces of the cylinder (Yousif
287 et al., 2023).

288 Training a DRL controller for AFC is time-consuming, prompting researchers to explore policy
289 transfer for efficient performance in different scenarios. Wang et al. showed that agents
290 trained at higher Re numbers are still capable of controlling the flow at lower Re numbers.
291 Interestingly, they also found that utilizing a trained agent to learn in different Re numbers

dramatically enhances learning efficiency by reducing the number of required learning episodes. This approach also leads to even better flow control outcomes (Y.-Z. Wang et al., 2022). He et al. explored policy transfer for AFC over a cylinder. In their study, the control model, trained in 2D environments at $Re = 100, 200$, and 300 , successfully transferred to 3D environments, maintaining significant drag reduction even with varying spanwise lengths. Additionally, they showed that models trained at higher Re perform well in lower Re environments, but not vice versa. Successful policy transfer from 2D to 3D reduces the extensive computational time and resources typically required for 3D flow control, enhancing its practical applications (He et al., 2023). Utilizing transfer learning has improved the DRL to be capable of controlling the flow in 3D objects in turbulent flow. Yan et al. used transfer learning for flow control on a square cylinder at high Re number 22000 . The initial training of the agent was done in a 2D domain at $Re=1000$, and they developed the SDTL-SAC algorithm to tackle the issue of mismatch in the state dimension for transfer learning from 2D to 3D. They reduced training costs to about fifty percent with considerable drag reduction and suppression of lift fluctuations (Yan et al., 2024a).

The coupling between the DRL algorithm and flow solver is also a crucial part that can influence training time. Some researchers have focused on this topic, resulting in frameworks like DRLinFluids (Q. Wang et al., 2022), DRLFluent (Mao et al., 2023), and Gym-preCICE (Shams and Elsheikh, 2023).

Researchers also explored reducing sensors and moving them closer to or onto the object's surface. While this enhances practicality, it may weaken flow control effectiveness. Xia et al. tried to mount the sensors on the body surface of a square cylinder rather than its wake while still maintaining effective flow control by reducing drag force. Mounting sensors on the body

315 initially resulted in a 50% less performance. To address this, they proposed a method to train
316 a dynamic neural network controller using an approximation of a finite-history sufficient
317 statistic. This new dynamic controller successfully improved drag reduction performance,
318 achieving near-optimal results (Xia et al., 2024). Paris et al. introduced a novel algorithm, S-
319 PPO-CMA, to optimize the sensors' layout. The algorithm showed drag reduction with
320 robustness across Reynolds numbers (100–216) and high noise levels. Sensor optimization
321 reduces the layout to five sensors without performance loss (Paris et al., 2021). Wang et al.
322 introduced DF-DRL for controlling the wake behind a cylinder. DF-DRL improves traditional
323 approaches by converting limited sensor data into dynamic features that predict future flow
324 behavior. DF-DRL with just one surface pressure sensor achieves the same results as methods
325 that rely on many sensors, making it more practical and efficient (Wang et al., 2024). These
326 studies show that the negative effects of changing sensor positions and reducing their number
327 can be mitigated by enhancing the learning algorithm.

328 Enhancements and modifications to the learning algorithm offer significantly broader
329 applications and benefits. Rabault and Kuhnle explored accelerating DRL using a multi-
330 environment approach as AFC faces challenges in algorithm efficiency and simulation speed.
331 The authors significantly enhance training speed by parallelizing data collection from multiple
332 independent simulations while maintaining learning quality (Rabault and Kuhnle, 2019). Their
333 Results demonstrate perfect speedups with optimal parallelization and acceptable
334 performance with over-parallelization. Zheng et al. improved AFC strategies for VIV control by
335 enhancing RL with expert demonstrations to avoid the time-consuming process of extensive
336 trial and error in the learning process. To accelerate learning, they demonstrated examples of
337 successful control strategies developed through techniques like pole-placement and surrogate
338 model optimization. This approach enabled rapid learning of AFC strategies through a

339 combination of prior demonstration data and online experience., making the approach faster
340 and more practical for real-world applications (Zheng et al., 2022). Zheng et al. studied AFC
341 strategies for VIV of a cylinder with two different approaches: active learning and
342 reinforcement learning. In the active learning framework, a Gaussian process regression
343 model and Bayesian optimization predict vibration amplitude with fewer simulations. The RL
344 approach employs SAC algorithm for real-time control, effectively lowering vibration
345 amplitude up to 82.7%, much larger than active learning method (Zheng et al., 2021). Li and
346 Zhang tried to use and embed physical information of the flow. By embedding flow stability
347 insights directly into the reward function, the strategy suppressed vortex shedding more
348 effectively (Li and Zhang, 2022). Mao et al. used a hybrid DRL method combining a Markov
349 decision process with time delays and an autoregressive policy to stabilize vortex shedding
350 and reduce force fluctuations. This method achieves a 90% reduction in lift and drag
351 fluctuations compared to standard DRL, while maintaining similar drag reduction. The
352 approach highlights the importance of incorporating physics-informed delays and historical
353 data for robust AFC in turbulent conditions (Mao et al., 2022). Castellanos et al. further
354 explored the potential of DRL by comparing it to Linear Genetic Programming Control
355 (Castellanos et al., 2022). Both techniques proved effective in minimizing drag, but DRL
356 exhibited greater resilience to initial conditions and noise and faster convergence with
357 reduced fluctuations in lift coefficients. Ghraieb et al. focused on single-step PPO for open-
358 loop control problems. The approach eliminates the need for continuous environmental
359 feedback, simplifying control strategies. Testing on 2D, including NACA 0012 airfoil and side-
360 by-side cylinders, demonstrates its efficacy in optimizing lift and drag reduction (Ghraieb et
361 al., 2021). Tang et al. proposed a new interpolate function to help the PPO algorithm for better
362 convergence in the training process. They reported that the DRL controller is not only able to

reduce drag but also can effectively reduce drag for any previously unseen value of Re close to the training range (Tang et al., 2020). Exploring alternatives to DRL, Liu and MacArt compared deep learning PDE augmentation methods with DRL for active flow control. By focusing on energy minimization through deterministic control, the deep learning PDE augmentation method achieved lower drag coefficients (Liu and MacArt, 2024).

Sakamoto and Okabayashi aimed at optimizing flow control laws by DRL while cutting down on computational costs. They used Dynamic Mode Decomposition with Control (DMDc) to develop a reduced-order model that represents flow dynamics in the DRL environment. The results show that the DRL in DMDc outperform conventional mathematical methods in terms of control performance, but the learning is still unstable. The DMDc method is able to provide control laws better than CFD in the same number of learning epochs and reduce computation time up to a factor of three (Sakamoto and Okabayashi, 2024).

Dong et al. used DRL to optimize suction flow on a circular cylinder within a wind tunnel. The main challenges are the lengthy training periods and low data utilization, which leads to high experimental costs. To overcome, they developed a surrogate model, SM-DRL, for optimal reward function and hyperparameter optimization to accelerate training. The DRL agent successfully adapts suction flow rates to incoming wind velocity, reducing lift fluctuations (Dong et al., 2023).

Building on the demonstrated ability of DRL-based AFC for a simple circular cylinder at low Re numbers in early studies, subsequent research has explored its capabilities under various conditions. These studies include diverse approaches, such as other geometries like square and rectangular with sharpened edges and different flow separation mechanisms, VIV control or

amplification for structural safety or clean energy harvesting, heat transfer management, and hydrodynamic stealth.

Study on a square cylinder with surface pressure sensors and multiple jets using SAC algorithm showed effectively minimize drag for different wind angles, even with those not encountered previously (Yan et al., 2024c). In another study, a rectangular cylinder with different ARs in $Re=1000$ also exhibited the capability of the DRL agent to control the flow through the use of multiple jet actuator (Yan et al., 2024b).

Ren et al. extended the application to an intermediate Re number of 1000, where weak turbulence presents challenges (Ren et al., 2021a). Despite requiring more training episodes, the DRL agent achieves a significant 30% drag reduction by elongating the recirculation bubble and reducing wake turbulence. The results mark the first successful use of DRL for AFC in weakly turbulent flows, paving the way for applications in more turbulent conditions. In another study, Ren et al. tried to enhance heat transfer around a heated circular cylinder undergoing VIV in the lock-in condition (Ren et al., 2024). Through a group of trainings, the RL agent eventually learned an effective control strategy that improves the average heat transfer rate by 76.7%. Another key study by Ren et al. (Ren et al., 2021b) explored how DRL-based AFC could enhance hydrodynamic stealth in bluff bodies. The focus was on reducing the velocity deficit in the cylinder's wake, a crucial factor in stealth applications. The DRL-controlled AFC strategy substantially reduced VIV and velocity deficits, adapting effectively to stationary and vibrating cylinder cases. Xie et al. explored DRL-based AFC on tandem circular cylinders to control the lift fluctuations of the downstream one. Actuations done via rotating the downstream cylinder resulted in a considerable reduction in lift fluctuations. Their strategy modifies the phase difference between the lift fluctuations of the two cylinders by

408 delaying the merging with the upstream cylinder wake and accelerating the formation of
409 recirculating bubbles after the vortex merging(Z. Xie et al., 2023). Noack tried to mitigate the
410 VIV of a square cylinder and tested three positions of jet actuators, including the front, middle,
411 and back of the cylinder. The DRL agent optimized the velocity of the jets to minimize drag
412 and lift coefficients, resulting in a reduction of up to 96% in vibration amplitude. The trained
413 agent is found to be robust under various reduced velocities (Noack, 2023). Conversely, Mei
414 et al. explored amplifying hydrodynamic forces around a circular cylinder to enhance energy
415 harvesting from VIV (Mei et al., 2021). With a reward function based on the mean and
416 instantaneous lift and drag forces to maximize force, the simulations showed a significant
417 increase in forces, resulting in higher energy harvesting potential.

418 In a study by Suárez et al., they applied DRL to control synthetic jets for drag reduction on a
419 3D cylinder at Re values between 100 and 400. Utilizing a multi-agent PPO setup, they
420 successfully minimized drag while maintaining acceptable lift limits, proving DRL's efficacy in
421 3D environments (Suárez et al., 2023). In another study, Amico et al. performed the first DRL
422 experimental application for a 3D bluff body at a high Reynolds number of $Re=10^5$. DRL
423 achieved up to 10% drag reduction when focusing solely on drag minimization and 5% with an
424 optimized power budget (Amico et al., 2022). Varela et al. also investigated DRL for minimizing
425 drag across a range of Reynolds numbers. Their research at Re values of 100, 1000, and 2000
426 revealed that the PPO agent adapted to different flow regimes, offering low-frequency control
427 at lower Re and high-frequency actuation at higher Re, resulting in consistent drag reduction
428 (Varela et al., 2022).

429 MARL, multi-agent reinforcement learning, is another approach which is used to enhance DRL
430 development when it comes to turbulent and flow control problems in three dimensions.

Since the flow becomes quite turbulent, a single DRL agent has massive adaptation problems. MARL allows multiple agents to interactively work at the same time, hence improving flow control efficiency. The study of Suárez et al. successfully overcoming flow control in 3D turbulent flows through a multi-stage training, drag reduction, and control performance improvement(Suárez et al., 2024).

The application of DRL for AFC around bluff bodies is an emerging and rapidly evolving field. Initial studies focused on controlling flow around a stationary circular cylinder at low Reynolds number of 100, demonstrating the potential of DRL-based strategies to stabilize vortex shedding and reduce drag. Following these initial studies, Different studies are conducted to improve the effectiveness and efficiency of DRL controller and also some researchers explored wider applications, including controlling and amplifying VIV, optimizing heat transfer, reducing wake flow for stealth purposes, and energy harvesting. These developments highlight DRL's potential to handle complex flow control problems in diverse and challenging conditions, paving the way for practical applications.

4. Research Gaps and Future Perspectives

This section begins with a review of various aspects of DRL-based AFC for bluff bodies, highlighting the research gaps associated with each aspect. Next, the available platforms and frameworks are introduced, followed by the presentation of a comprehensive framework. Finally, the challenges and future perspectives are discussed.

4.1. Research gaps

The application of DRL in AFC showed a rapid growth in the past years. Although these studies provided valuable insights, but there are still significant gaps hindering a full understanding of the field. Identifying and addressing them is essential for evaluating progress. This process is

454 also critical for advancing the boundaries of DRL applications in AFC. The review looked at 35
455 studies published between 2017 and 2024, with the majority of them concentrated on 2022
456 and 2023.

457 This section categorizes the research gaps related to DRL-based AFC of bluff bodies into
458 distinct themes, such as test categories, RL hyperparameters, reward functions, and additional
459 factors like actuation strategies, sensor design, and training methodologies. Table 2 represents
460 the common parameters and their value or percentage of their usage. Together, these
461 elements illustrate an emerging field that, while rich in potential, still faces foundational
462 challenges.

Table 2) Overview of key trends and parameters in RL for recent flow control studies

Category	Aspect	Details	Value / % of studies
Research Focus	Study Period	2017 – 2024	35 studies
	Shape	Single Circular Cylinder	80.00%
		Square Cylinder	8.60%
	Reynolds Number (Re)	100	77.14%
		200, 300	28.60%
		1000	20.00%
Test Category	Type of Simulation	2D, Num	88.57%
		3D, Num	5.71%
		Combined (Num+Exp)	5.71%
		3D, Exp	2.86%
CFD Tools	Most Common Software	OpenFOAM	28.60%
		FEniCS	22.90%
		Nek5000	8.60%
Simulation Domains & Num Time Steps	Most Common Domain	Rectangular Domain (22D, 4.1D)	45.70%
		Other Rectangular Domains	31.40%
		Circular and Cubic Domains	22.90%
	Δt_{sim}	Specified (0.0001 to 20.0)	51.40%
		Not Specified	28.60%
Hyper-parameters	RL Algorithm	PPO	74.30%
		SAC	20.00%
	Network Architecture	2 hidden layers, 512 neurons	42.86%
	Activation Functions	ReLU	17.14%
	Optimizer	Adam Optimizer	37.14%
	Learning Rate	Actor-Critic (10^{-3} , 10^{-4})	25.10%
	Discount Factor (γ)	Common values: 0.99, 0.97	34.30%
	likelihood-ratio clipping (ϵ)	0.2	25.70%
Reward Functions	Main Objectives	Drag Reduction	80.00%
		Lift Control	60.00%
Actuation and Control	Common Actuators	Synthetic/Control Jets	57.10%
	Control Action Type	Continuous Action Space	85.70%
	Common Distribution	Gaussian	42.90%
Sensors	Common Sensors	Velocity Sensors	65.00%
		Pressure Sensors	41.00%
	Sensor Placement	Wake Region	40.00%
		Around Cylinder	35.00%
Training time and procedure	Episode number	200	20%
		400	17%
	Parallel Environments	1-80	48.57%

4.1.1. Research Focus

The research focus of the reviewed studies shows an overwhelming preference for investigating single circular cylinders. This geometry dominates the field, appearing in roughly 80% of the research. Square cylinders, while providing a contrast in behavior, appear in far less studies. Other geometries and configurations are scarce (Figure 7). This focus on circular cylinders is likely due to their predictable aerodynamic properties, which provide a stable setup for testing DRL strategies. However, this emphasis leads to a significant gap with limited exploration of more complex shapes that may better represent real-world applications, such as rectangular, triangular, and trapezoidal prisms commonly found in structures like bridges, high-rise buildings, and offshore platforms.

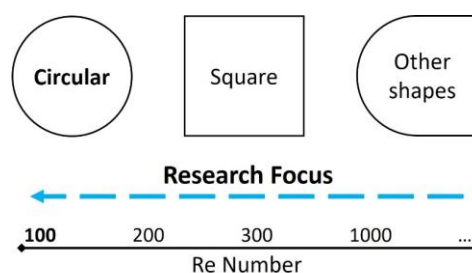


Figure 7) Research focus, the geometry and Re number

As a critical parameter in flow dynamics, the choice of Re number also reveals key trends. In most cases, researchers selected $Re=100$ for their study (cited in 77%). This is not surprising, as $Re = 100$ represents a transitional flow regime that is both computationally manageable and rich in features that DRL can effectively learn to control. Additionally, it is the value used by Rabault et al. in their pioneering study (Rabault et al., 2019). Other commonly used values include $Re=200$ and 300 , each appearing in 10 studies, while higher values such as $Re=1000$ are relatively scarce, appearing in only seven cases (Figure 7). The tendency to lower Re arises from practical constraints. Simulations at higher Re demand vastly more computational resources to capture turbulent flow complexity, making stable DRL training harder due to

larger state and action spaces. This tendency underscores the need for further research to scale DRL for realistic, turbulent flows.

4.1.2 Test Category

One of the most striking patterns in this analysis is the overwhelming emphasis on 2D numerical simulations. Of the 35 studies reviewed, 88.57% focused on this, whereas 3D numerical simulations accounted for only 5.71%. Similarly, 5.71% of the studies combined numerical simulations with experimental testing, and only one study focused exclusively on 3D experiments. This heavy reliance on 2D simulations, although they provide simplified and computationally efficient environments, highlights a research gap. 2D-trained models are invaluable for initial testing and hypothesis, but training models for 3D environments and experiments are necessary for real-world applications. The scarcity of 3D studies points to the ongoing challenge of extending DRL frameworks into more complex and real-world behaviors emerge more prominently (Figure 8).

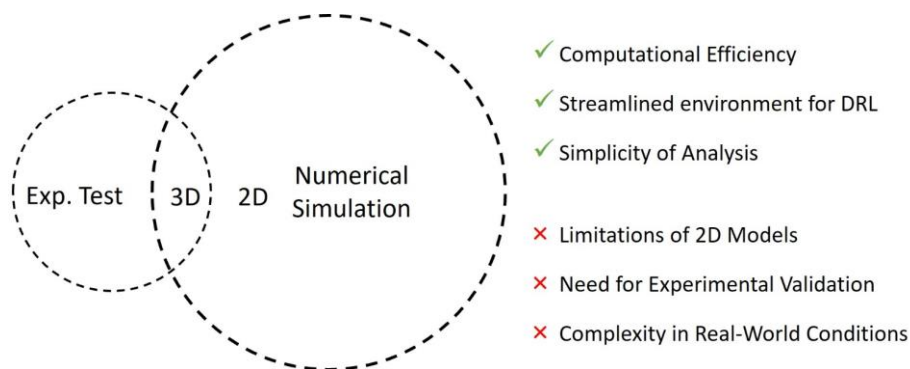


Figure 8) Test category of AFC using DRL

Among the studies reviewed, only two (Amico et al., 2022; Fan et al., 2020) provided detailed descriptions of their experimental setups, an alarming gap given the importance of experiments in validating DRL models. One study (Amico et al., 2022) measured static and fluctuating pressures, drag, and collected data through a highly detailed setup involving a NI-cDAQ chassis. Another study (Fan et al., 2020) involved towing a bluff cylindrical body in a fluid

500 tank, with custom-built rotating control cylinders and data collection interfaces. These studies
501 provide rare insight into experimental work but underscore the general lack of experimental
502 validation across the literature. This gap suggests that many studies remain confined to
503 idealized numerical environments, limiting the transferability of DRL techniques to real-world
504 systems where scale effects, turbulence, and boundary conditions are far more complex and
505 unpredictable.

506 4.1.3 CFD Solvers and Numerical Methods

507 The CFD solvers used in the studies reveal a diverse set of tools, yet OpenFOAM and FEniCS
508 are found to be the most widely adopted. OpenFOAM appears in about 30% of the studies,
509 while FEniCS is utilized in 20%. These tools offer open-source, highly adaptable platforms that
510 align well with the exploratory nature of DRL research, where custom modifications and
511 flexibility are often required. Other solvers, like Ansys Fluent, lattice-Boltzmann method,
512 Nek5000, and Alya appear in fewer studies. A comparison between flow solvers is presented
513 in Table 3.

Table 3) Comparison of CFD solvers for DRL

Feature	Alya	FEniCS	OpenFOAM	Fluent	Nek5000	Lattice Boltzmann Method
Numerical Method	Finite Volume	Finite Element	Finite Volume	Finite Volume	Spectral Element	Lattice Boltzmann
Scalability	Optimized for HPC	Suitable for HPC	Suitable for HPC	Limited HPC scalability	Suitable for HPC	Suitable for HPC
Turbulence Models	DNS, LES, RANS	Manual implementation	DNS, LES, DES, RANS	LES, DES, RANS	DNS, LES, limited RANS	DNS, LES, limited RANS
Multiphysics Flexibility	FSI, heat transfer, electromagnet	Flexible for PDEs	FSI, combustion, acoustics, plasma, etc.	FSI, combustion, acoustics, plasma, etc.	Limited to fluid dynamics and FSI	Mostly limited to fluid flows
Moving Mesh Support	Good support	Possible but complex	Good support	Strong support	Limited support	No traditional meshing
Interaction with ML	Needs external coupling	Easily integrated	Can be coupled via Python	Limited direct coupling	Limited direct coupling	Recent advances for ML coupling

FEniCS and OpenFOAM appear to be the best choices for solvers as they are easy to integrate with ML algorithms and show superior capabilities in flow simulation, respectively. Further, Alya is worth consideration as it is optimized for HPCs which accelerates the flow simulation in 3D turbulent cases.; however, further studies are required for DRL compatibility.

Integrating machine learning libraries, such as TensorForce, PyTorch, and TensorFlow, into CFD simulations, done by using platforms like DRLinFluids, reflects a rising trend toward hybrid approaches in DRL applications (Figure 9). These tools are essential for managing the complexity of DRL, especially since traditional reinforcement learning methods often struggle with low sample efficiency. Although these frameworks may help the researchers to accelerate the learning process, but the high computational cost of CFD simulations remains a challenge. This bottleneck shows the necessity for innovative solutions to improve learning

effectiveness. A detailed review on available tools and platforms are presented in chapter 4.2.1. DRL Platforms.

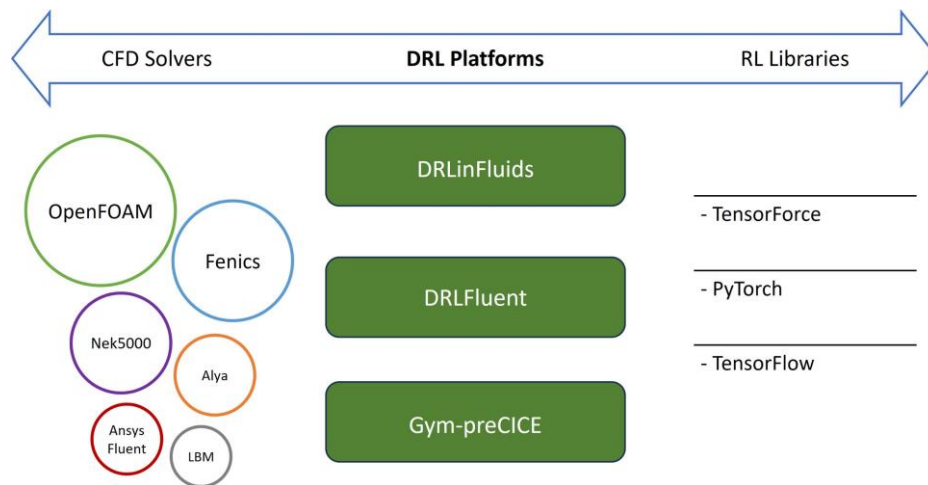


Figure 9) DRL frameworks, CFD solvers and ML libraries

4.1.4 Simulation Time Steps and Domains Shape

The time step in CFD simulations is a critical factor for accuracy and stability. It varies significantly in the reviewed studies, which show the complex nature of AFC problems. Time steps (Δt_{sim}) range from as low as 0.0001 to as high as 20.0, with around 51.4% of studies explicitly providing their chosen values. Some studies enforce solution stability through Courant–Friedrichs–Lewy (CFL) conditions, while others adopt adaptive schemes based on the evolving nature of the flow field. In case of simulation flow domain, the use of rectangular domains is prevalent with 77.14% of the studies. Most studies used rectangular domains for a confined circular cylinder, with a standardized configuration of (22D, 4.1D) being common in 45.7% of cases (“D” representing the cylinder diameter). Researchers study on confined cylinders rather than free-stream ones because the confined setup provides well-defined boundary conditions, making simulations more tractable and computationally efficient. The confinement also creates more complex flow interactions, such as stronger vortex-wall interactions, which present a more challenging problem for DRL. Additionally, confined

cylinder scenarios are more relevant to practical applications like pipelines and ducts, making the findings more directly applicable to real-world engineering problems.

4.1.5 Hyperparameters

When exploring RL for AFC, it is evident that a finely-tuned set of hyperparameters, external parameters that govern the learning process, can significantly influence the algorithm's performance. These hyperparameters span various components: algorithms, network architecture, activation functions, optimizer type, learning rates for the actor and critic, discount factor, exploration strategy, batch size, and likelihood-ratio clipping. While these elements are fundamental, their precise configurations across studies often remain under-explored, creating gaps in standardization and comparison.

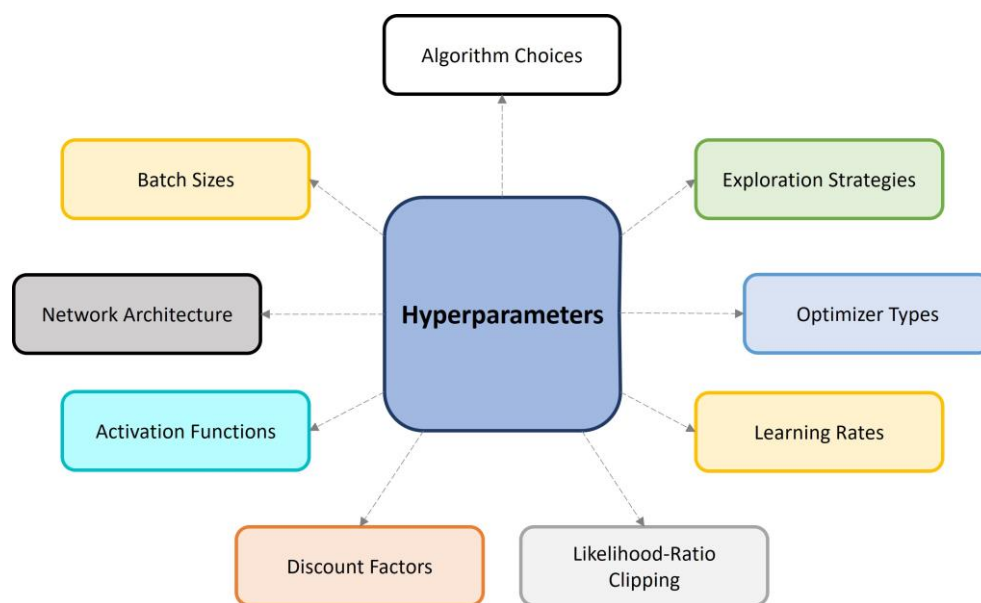


Figure 10) Main hyperparameters for DRL-based AFC

The control type is one of the most important factors in deciding which DRL algorithm may be adopted for AFC. DQN can only perform discrete control, which could make it inappropriate for flow control cases. PPO manages both discrete and continuous controls, allowing a greater range of AFC configurations. DDPG, SAC, and TD3 are purely continuous so that their actuator

555 adjustments can be smooth, which is generally required for AFC. The exploration strategy is
556 another critical aspect. DQN and PPO rely on stochastic policies, with DQN using ϵ -greedy
557 exploration, which might be inefficient for continuous action spaces. Noise-based exploration
558 is used in DDPG and TD3 which may be unstable in high-dimensional spaces. SAC uses entropy-
559 based exploration making the policy robust.

560 Training stability is another key factor because unstable training can result in unreliable
561 control. While PPO and TD3 are highly stable, DDPG is prone to instability. SAC enhances
562 stability with its entropy regularization which makes it another reliable option for AFC.

563 Sample efficiency is another parameter, where training data is expensive due to high
564 computational costs in CFD or experimental tests. DQN and DDPG are sample inefficient and
565 need huge amounts of training data. PPO is moderately sample efficient and is therefore an
566 option used very widely. SAC and TD3 are the most sample-efficient, allowing learning to
567 happen faster with the least number of interactions, which is quite attractive for AFC
568 applications.

569 When considering adaptability to flow variability, algorithms that generalize well to different
570 flow conditions are preferable. DQN fails due to its discrete nature. PPO and SAC, produce
571 very adaptable solutions that are best suited for real-world AFC applications, where flow
572 conditions are changing dynamically.

573 Another practical consideration is the computational cost. DQN and DDPG are lightweight,
574 but their poor performance limits their usefulness. PPO is moderately demanding and
575 balances efficiency with stability. SAC and TD3 are heavy, but they outperform the others,
576 converging faster, making the tradeoff worth it for many AFC scenarios.

577 Based on the reviewed PPO is the most widely used due to its stability and moderate
578 computational demands. SAC and TD3 are rising in fame owing to their efficiency and
579 robustness in handling complex flow behaviors. DQN and DDPG are rarely used due to their
580 limitations in control flexibility and stability. From these, it can be interpreted that researchers
581 may gravitate towards well-established methods such as PPO and SAC, driven by its strong
582 track record and widespread usage in RL libraries, rather than rigorously comparing
583 alternatives.

584 Regarding network architecture, 42.86% of the studies specify configurations using two
585 hidden layers, each with 512 neurons. This setup has become a de facto standard in the field
586 because it balances computational efficiency and the capacity to capture complex flow
587 dynamics. Other configurations with varying numbers of neurons, such as 64, 128, or 256,
588 appear less frequently, suggesting that many researchers stick to tried-and-true configurations
589 unless there is a compelling reason to deviate.

590 A similar pattern of limited exploration appears when examining activation functions. Out of
591 the 35 studies, only six mentioned using the Rectified Linear Unit (ReLU), a well-known and
592 widely used function due to its simplicity and effectiveness in mitigating the vanishing gradient
593 problem. The hyperbolic tangent (Tanh) appears in three studies, while the Swish function and
594 Sigmoid activation each receive a single mention. Strikingly, 22 studies (62.9%) have not
595 mentioned which activation functions were used at all, underscoring a documentation gap.
596 This lack of specificity leaves readers to assume that common activation functions were used,
597 or worse, that their impact on performance has been overlooked entirely. The optimizer type
598 is another crucial component in RL algorithms, influencing how efficiently the model learns
599 from data. The “Adam” optimizer is the most frequently mentioned (37.14%). Its popularity is

due to its adaptability and performance in training deep networks, particularly in environments with noisy gradients like those in fluid mechanic problems.

Learning rates for the actor and critic are specified in less than half of the reviewed studies. This omission highlights a critical gap, as the learning rate is one of the most sensitive hyperparameters in RL. Among the studies that do specify learning rates, the most common value is 10^{-3} , appearing in seven cases, with a smaller group opting for 10^{-4} in two instances. These values align with typical practice in RL, balancing convergence speed and stability. The discount factor (γ) is a parameter that controls how much the RL agent values future rewards compared to immediate ones. It is specified in only 13 cases, with the most common value of 0.99. This value is typical for tasks that require a balance between short-term and long-term performance, as is often the case in flow control over time. Other specified values include 0.97, 0.93, and 1.

The exploration strategy, specified in only 13 cases, allows the agent to gather information about its environment and discover new strategies. Among the specified strategies, maximum entropy, randomized actions, entropy regularization, and exploration noise are the most common. Techniques such as the Ornstein-Uhlenbeck process and Gaussian noise perturbation are also mentioned, showing the diverse range of approaches to exploration in RL. It must be noted that many of these strategies are inherent to specific algorithms, such as SAC's entropy-based exploration or PPO's policy-gradient approach. The batch size determines how many samples are processed before the model's weights are updated and can affect training stability and performance. A documentation gap appeared as this parameter is not mentioned in more than 77% of the reviewed articles. Finally, the likelihood-

ratio clipping parameter (ϵ) is only specified in 10 studies, with the majority of them (nine) using a default value of 0.2 adopted from the original PPO algorithm.

In summary, DRL for AFC heavily relies on selecting hyperparameters properly. Unfortunately, these settings are often under-reported that limits comparison across different studies. Improving documentation and experimenting with various configurations could improve progress and reproducibility of results in fluid dynamics.

For reproducibility, the description of the algorithm and hyperparameters is required. Creating a standard reporting table may be impractical as the models and related parameters may vary significantly. Under such circumstances, sharing the codes in online platforms is a more operative solution. Free access to the codes multiplies research gains since the work can be reused for new researches.

4.1.6 Reward Function

Effective strategies in DRL-based AFC depend on the precise design of reward functions, which are crucial for the learning process. The analyzed studies categorize reward functions into five groups with overlaps, which include Drag Reduction and Lift Optimization, Energy Efficiency, Flow Stabilization, VIV Control, and Heat Transfer Optimization (Figure 11).

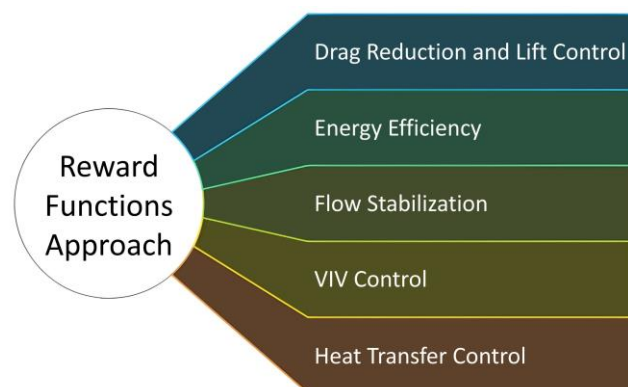


Figure 11) Reward Function approached in DRL-based AFC

Across the reviewed studies, the primary focus of most reward functions is clear: over 80% explicitly aim to reduce drag force and around 60% of the studies also target lift control, an equally important parameter, especially when maintaining stable flow is crucial. Rabault et al. in their pioneer study (Rabault et al., 2019), designed their reward function with a combination of lift and drag coefficient presented in Eq.(4, where $\langle . \rangle_T$ indicates the sliding average back in time over a duration corresponding to one vortex shedding cycle. The ANN tries to maximize the function to make it as little negative as possible, therefore minimizing drag and mean lift.

$$r_t = -\langle C_D \rangle_T - 0.2|\langle C_L \rangle_T| \quad (4)$$

This focus is understandable as drag and lift are key indicators of aerodynamic efficiency and even small deviation can translate into significant performance change. Researchers employed a variety of reward functions around this idea, including using time-averaged and instantaneous force measurements, both of which provide the agent with real-time feedback on its control actions.

In the study by Qin et al. the reward function is intricately designed using DMD to enhance the DRL model's to stabilize the vortex street and reduce drag (Qin et al., 2024). The function evaluates control effectiveness by encapsulating the global characteristics of the flow field through a weighted combination of dynamic mode amplitudes (Eq.(5). The reward function prioritizes the amplitude of the zeroth mode (α_0), which corresponds to the time-averaged flow and is crucial for maintaining stability. The remaining amplitudes (α_i), representing the unsteady components of the flow, are collectively given less weight.

$$r_t = 0.198 - 0.1|\alpha_0| - 0.0005 \sum_{i=1}^{28} |\alpha_i| \quad (5)$$

In the study by Jiang and Cao, designed their reward function based on the reduction of perturbation kinetic energy. The reward function for kinetic energy (Eq.(6), seeks to minimize these energy fluctuations within the fluid flow. The RL agent to optimize the oscillation control strategy, achieving lower energy fluctuations and reduced drag, thereby enhancing overall flow control efficiency (Jiang and Cao, 2023).

$$r_t = - \sum_{\{i\}}^{n_{probe}} [(u_i^2 + v_i^2) - (\bar{u}_i^2 + \bar{v}_i^2)] \quad (6)$$

In the study by Xu et al., the reward functions for the DRL-based AFC system is defined by the fluctuation of the drag and recirculation area, leading to enhanced flow control, reduced drag, and stabilized wake dynamics(Xu et al., 2020).

As the control process is conducted through some actuators like synthetic jets, or rotation, the energy efficiency of the system also becomes important. Xia et al. introduced a power-based reward function which calculates the net power savings by comparing the power saved from drag reduction against the power consumed by the control actuators, thereby incentivizing the agent to maximize drag reduction while minimizing actuator energy consumption (Xia et al., 2024).

In addition to studies related to lift and drag control, flow stabilization and energy efficiency, some researchers go further by incorporating secondary objectives such as heat transfer enhancement, and VIV control. In the study by Ren et al., the reward function is meticulously crafted to maximize the heat transfer from a circular cylinder experiencing VIV, employing RL techniques. The reward function (Eq.(7), is directly proportional to the Nusselt number. Where D_0 represents the cylinder's diameter, ΔT is the temperature difference between the cylinder's surface and the surrounding fluid, ∂T denotes the temperature gradient in the radial

direction, r and θ are the radial and angular coordinate, respectively. This formulation underscores the intricate relationship between the geometric and thermal parameters in enhancing the efficiency of heat transfer in systems subject to VIV (Ren et al., 2024).

$$r_t = \langle Nu \rangle = -\frac{D_0}{2\pi\Delta T} \int_0^{2\pi} \frac{\partial T}{\partial r} dr d\theta \quad (7)$$

The control of VIV itself studied by Zheng et al. with a reward function directly connected to the position and velocity of the cylinder (Zheng et al., 2022).

These added layers of complexity illustrate that it's not enough to simply reduce drag or stabilize lift; the solution also has to be efficient, stable, and adaptable in different conditions.

4.1.7 Actuation and Control Strategy

Actuators control flow in AFC applications, with their design being crucial to the control strategy. Synthetic jets are the most common, and they have been used in over half of the studies as they provide precise flow adjustments to reduce drag or stabilize lift. Their simple design, fast response, and ease of integration with DRL make them ideal for learning and adapting to flow changes. They are also compact and energy efficient, making them practical for both simulations and real-world applications.

Other approaches, like rotation of the main object or control rods and transverse oscillation, are the actuators used to manipulate the flow field. Plasma actuators, which are controlled by alternating current voltage, have also appeared. Though less common, these actuators have a unique advantage: they can induce rapid, localized changes in the flow, making them well-suited for scenarios where fast responses are needed (Figure 12).

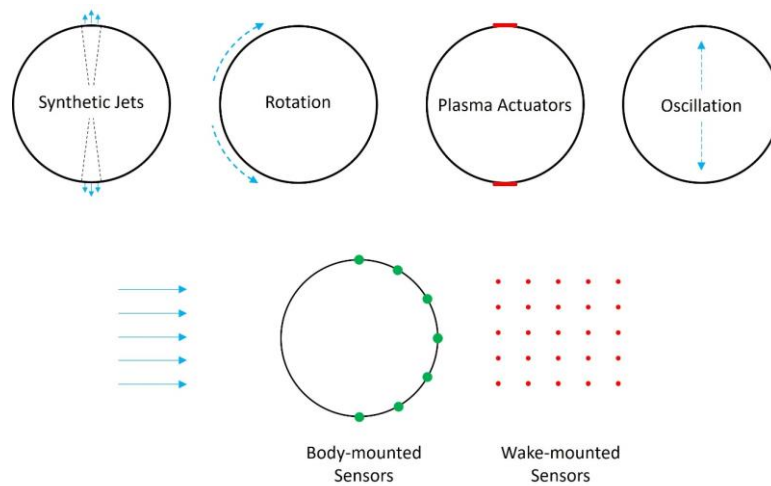


Figure 12) Actuators type and sensor placement

698 Actuation methods rely on advanced control strategies to enhance their effectiveness. About
699 20% of the studies employ ANNs to learn complex control actions to achieve this goal.

700 The structure of action spaces, which defines the range of possible actions an agent can take,
701 is also important in DRL. The review shows that most studies (about 85%) use continuous
702 action space. It offers high precision and flexibility which is required of AFC problems.

703 Regarding the underlying control distributions that dictate how actions are selected, the
704 Gaussian distribution stands out, appearing in 42.9% of cases. This is likely due to its natural
705 fit for modeling continuous, smooth action spaces. Other distributions, such as Beta (used in
706 8.6% of studies) and deterministic control (5.7%), show that researchers are experimenting
707 with different strategies to optimize control actions, though Gaussian remains the
708 predominant choice. Timing is also crucial in AFC strategies, with control steps varying widely
709 across studies to fit specific contexts. In some studies, researchers used fixed time steps, while
710 others tried to synchronize their actions with flow phenomena like vortex shedding. The
711 frequency of actions also differs in reviewed studies showing the varied nature of the
712 problems.

This variability highlights the challenge of aligning control timing with the natural dynamics of the problem. In scenarios with cylindrical objects controlled by RL agents, this challenge increases at higher Re numbers, where turbulence is introduced, and even small state changes can lead to significantly different outcomes.

4.1.8 Sensors

In a DRL control system, the sensors are the control agent's eyes and ears. These sensors capture the essential state variables required for the decision-making process. The type, number, and placement of these sensors affect the effectiveness of control policies. Common sensors measure variables like velocity, pressure, skin friction, and temperature in fluid dynamics, particularly flow around bluff bodies. The number and position of these sensors vary depending on the nature of the problem. Among the reviewed studies, 65% employed velocity sensors, making them the most widely used tool for measuring flow characteristics. Pressure sensors come next, used in 41% of the studies, which makes sense given their role in capturing flow forces that can disrupt stability. Some studies adopt a more integrated approach, with 24% combining both velocity and pressure sensors and 21% even incorporating past actions or dynamic parameters into the mix to provide richer, more comprehensive data for the agent.

The sensor placement is just as critical as what they measure. Roughly 40% of the studies positioned sensors in the wake region behind the cylinder, which is the area with the most gradients. It's a logical choice as this region holds vital information about turbulence and vortex dynamics, which are key factors in flow control. Another 35% of studies placed sensors around the cylinder itself, where flow separation and surface interactions dominate (Figure 12). It is noteworthy that mounting the sensors on the body is more practical in real-world

applications as it simplifies installation and reduces vulnerability to disturbances. However, not all studies specify their sensor configurations, pointing to a reporting gap.

Noise, both in the flow and in the sensor data, adds another layer of complexity, but only 11.43% of the articles provide specific details about sensor noise levels. One article investigated signal-to-noise ratio under different conditions, while another employed a Kalman filter to improve data accuracy.

4.1.9 Training time and procedure

Training time of an RL agent depends on different parameters, such as time step, episode number and length and number of training environments (Figure 13). Training an RL agent for flow control is a time-intensive process, and the reviewed studies showcase a wide variation in the duration and complexity of these training procedures. Some agents converge quickly, in as few as 15 episodes, while others take considerably longer, requiring 600 to 1000 episodes. This disparity is unsurprising as it shows the diverse range of problem complexities and RL methods across studies.

Of the 35 studies, more than half specified the number of training iterations, providing insight into how long these systems take to learn effective control strategies. Around 20% required about 200 episodes for convergence, while another 17% needed around 400. Different methods of reward shaping, varying complexities in flow scenarios, and divergent convergence criteria are the main reasons behind these differences that make cross-study comparisons difficult.

Interestingly, episode length, the duration for which the agent interacts with the environment before resetting, also varies widely. Some researchers used non-dimensional time units, while

others relied on seconds, or even vortex shedding periods. This diversity shows that measurements must be done to each specific flow configuration.

The number of parallel environments used during training is another parameter that varies in different studies. Parallel environments allow multiple simulations to run concurrently, accelerating training in reinforcement learning by enabling agents to explore diverse scenarios simultaneously. Among 17 studies provided details on this aspect, a range from a single training environment to as many as 80 parallel agents is found.

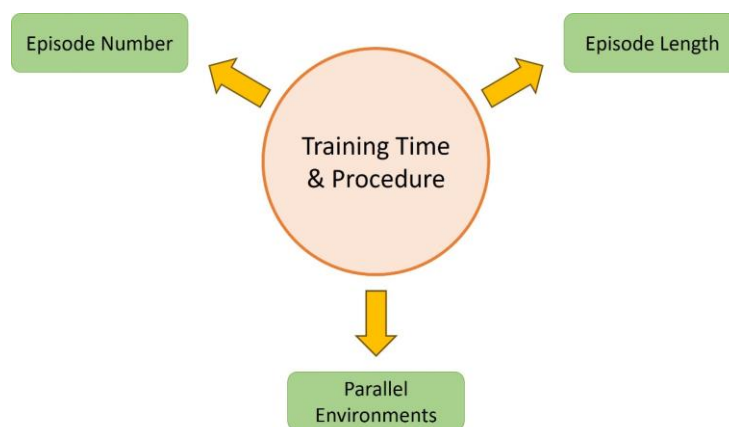


Figure 13) Effective parameters on training time and procedure

Currently, there are no standards for determining optimal training durations, episode lengths, or parallelization strategies, showing a research gap. Addressing this gap requires future studies to develop more standardized methodologies. This will facilitate a clearer understanding of the best ways to train RL agents to navigate the complexities of fluid dynamics.

4.2. DRL Frameworks for AFC

In this section, a review of open-source platforms for DRL-based AFC is conducted first, and then a comprehensive framework is presented for implementing DRL for flow control of bluff bodies.

4.2.1. DRL Platforms

In a DRL-based AFC setup, the agent learns optimal control strategies by interacting with a simulated or physical flow environment. The coupling process involves a feedback loop where the DRL agent receives state information, selects control actions, and evaluates rewards based on performance metrics. This process is generally done by two different frameworks: intrusive and non-intrusive. An intrusive framework embeds the DRL agent directly into the CFD environment, removing the necessity for external data exchange. As an example of such a framework, `drlFoam` (Weiner, 2024) couples a PyTorch-based DRL framework with OpenFOAM intrusively. The agent's model is directly loaded inside OpenFOAM, so no external communication is needed. Another example of an intrusive DRL framework is DRL-CFD, which, unlike the previous one, is based on the TensorFlow-based DRL package Tensorforce (Salehi, 2024).

In many cases, coupling algorithms run in separate programs, requiring an interface for communication between the agent and the environment. A simple way to handle this is through a "non-intrusive" approach, where the DRL code interacts with the CFD program as if it were a black box without being directly integrated into it. Wang et al. proposed DRLinFluids and employed OpenFOAM as a widely used and flexible Navier-Stokes solver in both industry and academia, along with Tensorforce or Tianshou as widely used versatile DRL packages. To enable flexible sharing of data and interaction with various modules, DRLinFluids employs regular expressions (Q. Wang et al., 2022). This approach makes using DRL for AFC easier, but it heavily depends on the data exchange between the agent and the environment. This interaction often requires stopping and restarting the simulation, which can slow things down and make the process less efficient. Another example of open-source frameworks is Relexi

(Kurz et al., 2022) which integrates TensorFlow Agents with a high-order spectral CFD code, FLEXI. the SmartSim (Partee et al., 2022) package is also utilized to facilitate information exchange between the DRL and CFD components of this framework.

Gym-preCICE(Shams and Elsheikh, 2023) integrates OpenFOAM with DRL using the external package preCICE, which is an open-source coupling library for partitioned multi-physics simulations, to handle information exchange. The non-intrusive strategy can even provide the possibility of coupling with commercial CFD software where the code is not accessible. DRLFluent (Mao et al., 2023) is a framework that couples Ansys Fluent with a Python DRL program through the omniORB package.

Non-intrusive platforms are generally more advantageous and recommended for DRL-based AFC implementations. Their capacity to interact with a variety of CFD solvers, including commercial software, provides unmatched flexibility and minimizes dependencies on specific solver architectures. Although there may be efficiency challenges due to coupling and interaction mechanism like Input/Output operations, their adaptability and broader applicability make them a preferred choice for real-world AFC problems.

4.2.2. Proposed Comprehensive Framework

Designing a clear framework for DRL-based AFC of bluff bodies is essential as the flow control problems are highly complex. Since this field is still emerging in the scientific community, a well-designed framework is key to guiding research, effectively combining simulations with experiments, and driving real-world applications forward. This framework includes a step-by-step approach for a DRL-based AFC problem with an example of flow control around a bluff body. Each step highlights key methodologies, from problem definition to real-world deployment, providing a comprehensive guide for researchers entering this field (Figure 14).

This framework consists of seven steps from the beginning of the problem definition to the end, with evaluation and validation of the results. The details for each step are presented in Table 4.

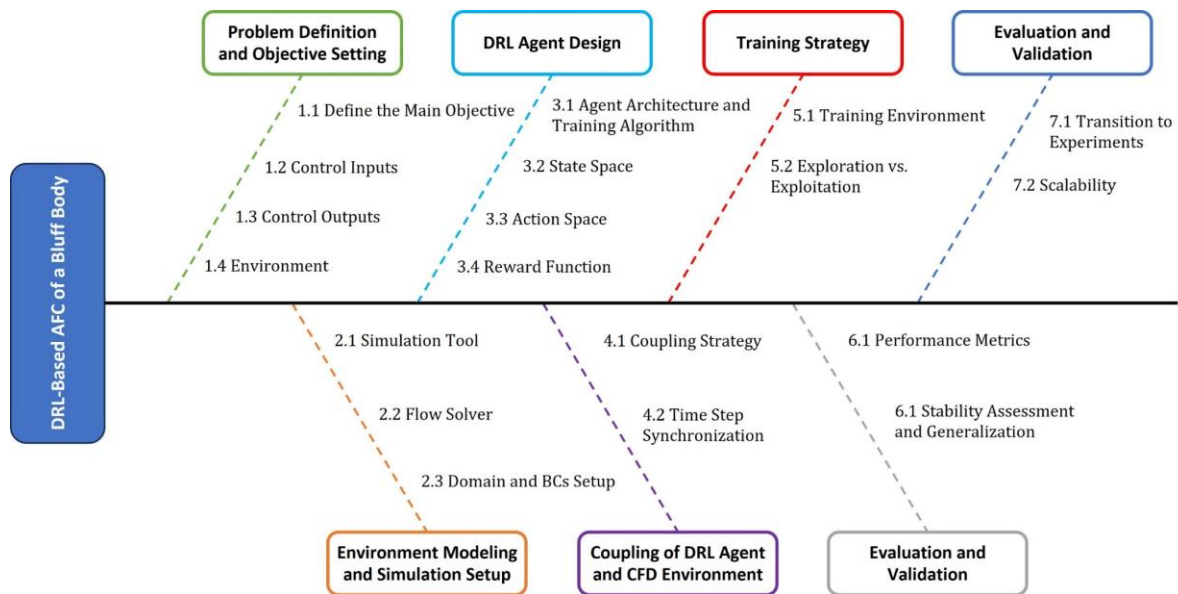


Figure 14) Proposed framework for DRL-based AFC of a bluff body

The first step involves defining the problem and setting main objectives. In the case of a bluff body, typical goals include flow stabilization, delaying flow separation, or energy efficiency. However, depending on the researcher's goal, it can also have broader applications like heat transfer and suppression or amplification of VIV. Next, the control inputs and outputs must be selected. Different actuators and corresponding inputs are available for control input with synthetic jets, which are the most common for bluff bodies defined by parameters such as the actuation amplitude, frequency, and phase. The control outputs are related to the main objectives, including key flow characteristics like wake structures and force variation. Environmental parameters, such as flow properties or the Reynolds number, must also be carefully specified.

833 Next comes environment modeling and simulation setup, where tools like OpenFOAM, ANSYS
834 Fluent, or any other PDE solver codes are used to create an accurate representation of the
835 flow system. The bluff body is modeled with actuators, such as synthetic jets or control rods,
836 and the simulation model is selected depending on the flow regime and computational
837 resources. It is also essential to properly define the boundary conditions.

838 In step 3, the DRL agent is designed to act as the decision-making core of the system. Its core
839 is implemented with an algorithm implemented with DNN, like PPO or SAC, which are often
840 chosen for AFC of bluff bodies. Its state space includes features like velocity and pressure
841 fields, along with lift and drag coefficients. The agent's action space involves controlling the
842 actuators. It varies depending on the actuator, type such as jet parameters like amplitude,
843 frequency, and phase or rotation velocity for control rods. Finally, a reward function must be
844 defined carefully to ensure that the agent's actions are aligned with the main control goals.

845 Effective communication between the DRL agent and the CFD environment is a key for
846 successful training procedure. The CFD solver supplies data from flow field, while the DRL
847 agent sends updated actuator commands. Middleware libraries or APIs are often used to
848 manage data exchange, keeping everything in sync and reducing delays. For the case of a bluff
849 body, specialized platforms like DRLinFluids (Q. Wang et al., 2022) and DRLFluent (Mao et al.,
850 2023) are suggested as they couple DRL agent with capable flow solvers and make parallelism
851 available for high performance computing.

852 During the training phase (step 5), the DRL agent must balance exploration and exploitation
853 to refine its control policy. It starts with random actuation patterns to explore a broad action
854 space before gradually optimizing its decisions based on cumulative rewards. Parallel

855 simulations across multiple environments can significantly speed up the training process while
856 convergence is monitored using performance metrics.

857 Once training is complete (step 6), the agent's performance can be analyzed based on metrics
858 like drag reduction, flow stability, or other objectives. Evaluation and validation ensure the
859 DRL agent's robustness across varying flow conditions and Reynolds numbers, showing the
860 generalization, stability, and robustness of the trained agent in different conditions.

861 Finally, Testing the system in experimental setups, like in water or wind tunnels, connects
862 simulations to real-world performance. The Comparison between the simulation and
863 experiment outcomes the system's scalability and effectiveness in practical applications. This
864 structured approach provides a comprehensive pathway for leveraging DRL in Active Flow
865 Control applications.

Table 4) Comprehensive Framework for DRL-Based Active Flow Control (AFC) of a bluff body

1. Problem Definition and Objective Setting	
1.1 Define the Main Objective	Reduce drag, enhance lift, delay flow separation, suppress or amplify vibration, Heat Transfer, Energy Harvesting, ...
1.2 Control Inputs	Select control actuators (e.g., synthetic jets, plasma actuators, rotation, control rods, ...)
1.3 Control Outputs	Desired flow state (e.g., lower drag coefficient).
1.4 Environment	Flow domain, Re number, boundary conditions, and fluid properties.
2. Environment Modeling and Simulation Setup	
2.1 Simulation Tool	OpenFOAM, ANSYS Fluent, lattice-Boltzmann, or other CFD solvers.
2.2 Flow Solver	Choose simulation type depending on flow regime and computational resources (Laminar, Rans, LES, DNS, etc.)
2.3 Domain and Boundary Condition Setup	Define geometry (Bluff body) and actuators, grid resolution Set boundary conditions for inlet, outlet, walls etc.
3. DRL Agent Design	
3.1 Agent Architecture and Training Algorithm	Reinforcement learning algorithm (e.g., PPO, TD3) Implemented with DNN.
3.2 State Space	Observations from the flow field (e.g., velocity field, pressure distribution).
3.3 Action Space	Jet actuation parameters (e.g., frequency, amplitude, phase).
3.4 Reward Function	Minimize drag coefficient (Cd), Maximize lift coefficient (Cl), Smooth flow separation.
4. Interaction Between DRL Agent and CFD Environment	
4.1 Coupling Strategy	Two-way interaction between DRL agent and CFD solver. Flow field data → Agent → Control commands → Actuator input in CFD
4.2 Time Step Synchronization	Ensure CFD solver time steps align with DRL decision intervals.
5. Training Strategy	
5.1 Training Environment	Parallel simulations for faster convergence.
5.2 Exploration vs. Exploitation	Balance initial exploration with later exploitation of learned policies.
6. Evaluation and Validation	
6.1 Performance Metrics	Reduction in drag, improvement in lift, or delay in separation.
6.2 Stability Assessment and Generalization	Check if control is robust across varying flow conditions.
7. Deployment and Real-World Testing	
7.1 Transition to Experiments	Validate in physical wind/water tunnel setups.
7.2 Scalability	Evaluate feasibility for larger-scale applications.

4.3. Challenges and Future Perspectives

The review on the application of DRL-based AFC of bluff bodies showed notable progress in various aspects. However, the field is still in its early stages and requires significant advancements before becoming viable for application in real world. The required advancements cover several key areas, including refining reward functions, optimizing sensor placement, exploring transfer learning, and developing hybrid approaches, leading to wider applications, represented by the increasing number of arrows in Figure 15.

As mentioned in the last section, most of the studies are conducted in low Re numbers with 2D flow fields for the simulation. Only a few studies are conducted at higher Re numbers corresponding to weak turbulent flow. This focus arises from several reasons. At low Re corresponding to the laminar flow regime, the simpler and more predictable flow structure allows researchers to focus on specific phenomena without chaotic interactions. These conditions also provide an ideal training ground for DRL agents, enabling algorithm testing and refinement in controlled scenarios before scaling to high-Re flows. Additionally, designing and testing algorithms is more efficient as the computational cost is relatively lower. It is also noteworthy that research in low Re numbers has some direct applications in fields like microfluidics, underwater drones, and biomedical devices, making it practically valuable.

On the other hand, applying DRL in high Re numbers and turbulent flow regimes faces different challenges, which are mainly related to data requirements, computational costs, and algorithmic efficiency. Training agents demands vast datasets, but generating each episode in turbulent conditions is computationally intensive and time-consuming. The chaotic nature of turbulence requires extensive exploration across numerous states, increasing training demands. Accurate training also relies on multiple sensors to capture phenomena that are

considerably smaller than laminar flow, resulting in larger, slower neural networks prone to instability with noisy data. Turbulent flows also require dense grids and small time steps in CFD simulations, which, combined with DRL's iterative training, significantly amplify computational costs despite parallel computing advancements.

It must be noted that increasing the Re number and becoming turbulent also changes the flow field, which does not remain 2D anymore. Therefore, 3D setups are required, which increases the computational cost for both the CFD solver and DRL agent. The large number of mesh cells in 3D domains makes CFD simulation extremely time-consuming. They also require more sensors and actuators, which increases the complexity of the DNN and makes the training process more challenging.

Thus, the core DRL obstacles in 3D turbulent flows can be classified into expensive costs of flow solver computations and large dimensionality of state space specific for the DRL component, due to a turbulence's complexity and multitude of control factors. Several studies tried to improve the computational efficiency of CFD solvers. The use of parallel simulations alongside larger HPCs has sped up the flow solvers in the last few decades. In the past few years, GPU-optimized flow solvers appeared and performed DRL-based AFC for a turbulent separation (Font et al., 2025).

Challenges and novel efforts directed towards enhancing the overall efficiencies of different aspects of DRL, such as Multi-agent reinforcement learning (MARL), group invariant methods, and symmetries-based dimension reduction. In case of MARL, its implementation allows multiple agents to learn in a cooperative or competitive manner. This greatly reduces the burden of a single agent having to make decisions all by itself. As a result, a more effective and centralized decision-making policy is established, which works remarkably well in large scale

turbulent flows. The effectiveness of a MARL framework studied in drag reduction in channel flows turbulence is mitigated (Guastoni et al., 2023). Its improvement has also been studied in other cases like 2D and 3D Rayleigh-Benard Convection problems (Vasanth et al., 2024; Vignon et al., 2023a). Group invariant methods enable the DRL model to generalize the structure and ensure that the learned control policies are robust to transformations such as translations, rotations, or reflections. It prevents model from learning redundant patterns and improves sample efficiency, which is crucial in computationally expensive simulations. The study of Bewley et al. (Belus et al., 2019) established fundamental ideas that subsequently impacted MARL-based techniques for scalable MIMO flow control by introducing the utilization of locality and translational invariance in DRL for managing unstable falling liquid films.

Symmetries-based dimension reduction goes a step further in reducing the complication of the problem by determining the redundant variables and eliminating them, thus resulting in a lower-dimensional yet representative state space. In 3D turbulence control, where it is not possible to resolve all flow structures, this approach allows the DRL to work only on the critical characteristics that really affect the flow behavior. Jeon et al. (Jeon et al., 2024) explored how symmetry-based dimension reduction can enhance DRL efficiency by leveraging underlying invariances to reduce the complexity of the state space. It is shown that the use of symmetries improves the learning speed and reduces the computational expenses, thereby making the control policies more generalizable. All three approaches combined assist DRL to better handle the high dimensionality and chaotic nature of turbulent flows. The MARL distributes the control complexity, the Group Invariant Methods help generalization, whilst the Symmetries-based Dimension Reduction helps in problem simplification, which in turn leads

935 to faster learning, due to lower computational expenditure and more robust flow control
936 strategies.

937 Direct implementation of DRL in experiments is a promising way for future research that helps
938 address issues of CFD computational cost. Although experimental studies on DRL-based AFC
939 are rare, but recent studies like DRLinWT (Dong et al., 2024), trying to couple DRL with wind
940 tunnel testing have received some credibility in overcoming certain key challenges. It allows
941 researchers to rely less on high-cost simulations and validate DRL strategies under real-world
942 conditions in an interactive platform.

943 Another important consideration for future research is broadening the scope of objectives.
944 Currently, the majority of studies focus on controlling flow to reduce mean drag and lift
945 fluctuations around a fixed cylinder. While this goal is highly relevant to applications, it
946 represents only a fraction of the potential applications of AFC of bluff bodies. For instance,
947 energy harvesting from VIV offers a promising direction. Optimizing flow conditions could
948 enhance the performance and efficiency of these mechanisms. DRL-based AFC could also be
949 employed for noise reduction in aerodynamic and hydrodynamic systems, such as minimizing
950 flow-induced noise in underwater vehicles or aircraft components. In addition, AFC goals
951 could extend to improving heat transfer efficiency in thermal management systems, such as
952 electronic cooling devices or heat exchangers, where precise control of flows can significantly
953 boost performance. Expanding DRL-based AFC research beyond drag reduction can tackle
954 more real-world challenges and benefit engineering, energy, and environmental systems. This
955 broader focus will make DRL algorithms more versatile and speed up their practical use.

956 Redesigning reward functions is another important direction for future research. In most
957 current studies, reward functions are primarily based on directly minimizing drag and lift.

958 However, other parameters closely tied to the underlying physics of the flow are often
959 overlooked. Integrating relevant physical parameters, such as vortex shedding frequency,
960 separation point, flow energy distribution, and wake vorticity, into reward functions can
961 improve the performance of the control system and potentially reduce training time. This is
962 particularly crucial in high Reynolds number regimes and turbulent flows, where
963 computational costs and training durations are significant challenges.

964 Transfer learning is recognized as a crucial strategy for reducing the significant computational
965 costs associated with training DRL agents, especially in High Re numbers and turbulent flows.
966 The main idea is to train agents in simpler and less computationally demanding environments,
967 such as low Re number flows or simplified geometries, and then transfer the trained agent to
968 test on more complex scenarios. This approach allows agents to start with a baseline
969 understanding of flow dynamics, and reduces the need for extensive exploration in complex
970 environments. Beyond reducing training costs, transfer learning can also improve the
971 robustness and generalizability of DRL algorithms.

972 Another important research area is improving sensor placement and reducing reliance on
973 excessive sensors. Many current studies use sensors placed in the wake, but this setup is often
974 impractical in the real world and even in experimental tests due to interferences with the wake
975 flow. Removing wake sensors and reducing their number mounted on the object leads to
976 partial measurement, which can reduce the performance of DRL systems, showing the need
977 for smarter algorithms to make up for missing flow data. This challenge grows in turbulent
978 flows, where capturing complex dynamics requires a large number of sensors and 3D domains.

979 The promising study by Wang et al. (Wang et al., 2024) demonstrated that wake-mounted
980 sensors can be removed while maintaining high control performance by defining dynamic

981 features based on previous states and actions. Further studies on more complex real-world
982 cases may lead to more practical sensor placements.

983 Combining DRL with other methods, such as Genetic Programming, model-based approaches,
984 or traditional control strategies, is promising for improving flow control capabilities. These
985 hybrid strategies combine different methods to overcome DRL's weaknesses, especially in
986 more challenging conditions like high Re numbers.

987 In summary, while DRL-based AFC of bluff bodies has made significant progress, its application
988 in real world is still blocked. Researchers must tackle remaining challenges like computational
989 limits, refine reward functions, apply transfer learning, optimize sensor placement, and create
990 hybrid approaches to boost efficiency and scalability. Additionally, widening the scope of
991 objectives to more practical and trending topics like energy harvesting, noise reduction, and
992 thermal management will further enhance DRL applications.

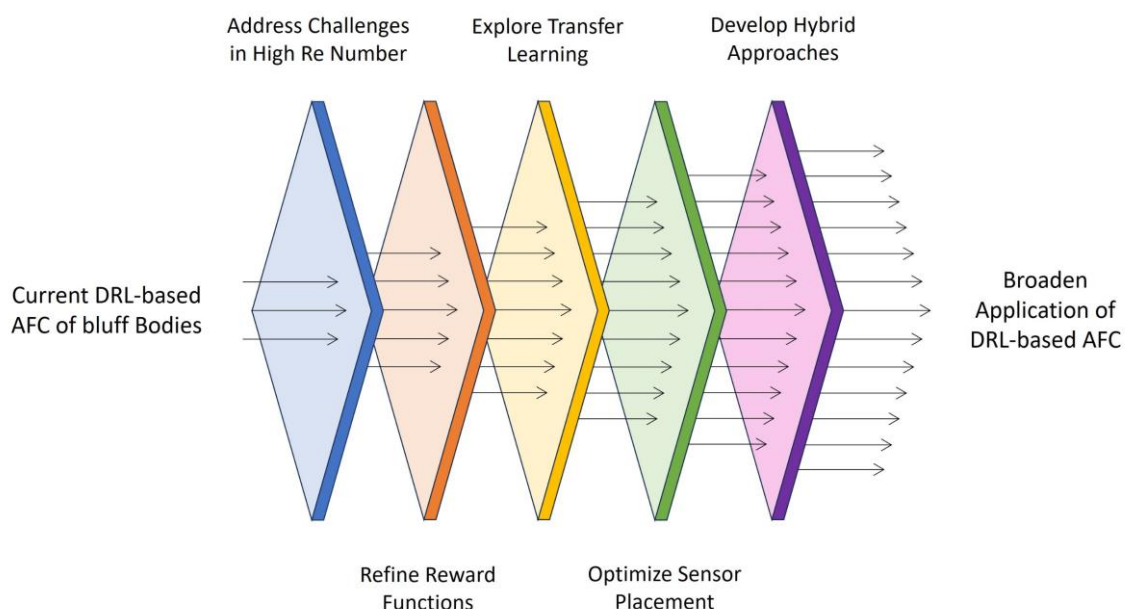


Figure 15) Challenges and future perspective to broaden application of DRL-based AFC

5. Conclusion

This review examined the integration of DRL for AFC of bluff bodies and critical advancements, persistent limitations, and emerging opportunities in the field are highlighted.

The predominance of 2D numerical simulations at low Re numbers, has provided foundational insights. However, transitioning from these simplified numerical environments to 3D turbulent flow conditions is still challenging due to mostly computational challenges. Future efforts must focus on scaling these approaches to higher Re numbers and 3D environments to validate their robustness and generalizability.

The predominance of 2D numerical simulations at low Re numbers has provided foundational insights. However, transitioning from these simplified numerical environments to 3D turbulent flow conditions is still challenging. Efforts have been made to extend the field, addressing challenges related to the flow solver and DRL components by proposing approaches such as accelerated GPU-based flow solvers and improving DRL algorithms using MARL, group-invariant methods, and symmetry-based dimension reduction, but further studies are required.

PPO emerged as the dominant DRL algorithm due to its balance between training stability and policy adaptability. Most studies employed deep neural network architectures with two hidden layers of 512 neurons, highlighting a standardized approach to network design. The Gaussian distribution is commonly used control action representation, facilitating smooth and continuous actuation. While these trends demonstrate a degree of convergence in methodological choices, inconsistencies in reporting hyperparameters and training procedures result in obstacles to reproducibility and cross-study comparisons.

1015 Reward function design was identified as another influential factor in the success of DRL-
1016 based AFC. While early studies focused mainly on flow stabilization by minimizing drag and
1017 lift fluctuations, recent advancements have explored objectives such as vibration control and
1018 heat transfer enhancement. Integrating more physical parameters into reward functions can
1019 be a promising direction for future research. The actuators and sensor placement also play
1020 crucial roles in the effectiveness of the control system. Synthetic jets are the preferred
1021 actuation mechanism due to their precision and adaptability, while sensor placement in the
1022 wake region continues to dominate most setups. However, real-world implementation favors
1023 sensors mounted directly on the bluff body surface. Eliminating the wake-mounted sensors
1024 leads to partial measurement and lower effectiveness. Therefore, algorithm improvement is
1025 required to make up for missing flow data. This challenge grows in turbulent flows, where
1026 capturing complex dynamics requires a large number of sensors and 3D domains.

1027 In conclusion, while significant progress has been made in the DRL-based AFC of bluff bodies,
1028 critical challenges remain. By addressing the research gaps and challenges through a
1029 comprehensive framework, DRL-based AFC systems have the potential to broaden their
1030 capabilities to be used in real-world applications.

References

- Amico, E., Cafiero, G., Iuso, G., 2022. Deep reinforcement learning for active control of a three-dimensional bluff body wake. *Phys. Fluids* 34. <https://doi.org/10.1063/5.0108387>
- Andrychowicz, M., Wolski, F., Ray, A., Schneider, J., Fong, R., Welinder, P., McGrew, B., Tobin, J., Pieter Abbeel, O., Zaremba, W., 2017. Hindsight experience replay. *Adv. Neural Inf. Process. Syst.* 30.
- Barbagallo, A., Dergham, G., Sipp, D., Schmid, P.J., Robinet, J.-C., 2012. Closed-loop control of unsteadiness over a rounded backward-facing step. *J. Fluid Mech.* 703, 326–362. <https://doi.org/DOI: 10.1017/jfm.2012.223>
- Bellemare, M.G., Dabney, W., Munos, R., 2017. A distributional perspective on reinforcement learning, in: *International Conference on Machine Learning*. PMLR, pp. 449–458.
- Bellman, R., 1957. A Markovian Decision Process. *Indiana Univ. Math. J.* 6, 679–684. <https://doi.org/10.1512/iumj.1957.6.56038>
- Belus, V., Rabault, J., Viquerat, J., Che, Z., Hachem, E., Reglade, U., 2019. Exploiting locality and translational invariance to design effective deep reinforcement learning control of the 1-dimensional unstable falling liquid film. *AIP Adv.* 9, 125014. <https://doi.org/10.1063/1.5132378>
- Brackston, R.D., García de la Cruz, J.M., Wynn, A., Rigas, G., Morrison, J.F., 2016. Stochastic modelling and feedback control of bistability in a turbulent bluff body wake. *J. Fluid Mech.* 802, 726–749. <https://doi.org/DOI: 10.1017/jfm.2016.495>
- Castellanos, R., Cornejo Maceda, G.Y., de la Fuente, I., Noack, B.R., Ianiro, A., Discetti, S., 2022. Machine-learning flow control with few sensor feedback and measurement noise. *Phys. Fluids* 34. <https://doi.org/10.1063/5.0087208>
- Chen, J., Hachem, E., Viquerat, J., 2021. Graph neural networks for laminar flow prediction around random two-dimensional shapes. *Phys. Fluids* 33.
- Chen, W.-L., Huang, Y., Chen, C., Yu, H., Gao, D., 2022. Review of active control of circular cylinder flow. *Ocean Eng.* 258, 111840. <https://doi.org/https://doi.org/10.1016/j.oceaneng.2022.111840>
- Dabney, W., Rowland, M., Bellemare, M., Munos, R., 2018. Distributional reinforcement learning with quantile regression, in: *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Dahan, J.A., Morgans, A.S., Lardeau, S., 2012. Feedback control for form-drag reduction on a bluff body with a blunt trailing edge. *J. Fluid Mech.* 704, 360–387. <https://doi.org/DOI: 10.1017/jfm.2012.246>
- Degrave, J., Felici, F., Buchli, J., Neunert, M., Tracey, B., Carpanese, F., Ewalds, T., Hafner, R., Abdolmaleki, A., de las Casas, D., Donner, C., Fritz, L., Galperti, C., Huber, A., Keeling, J., Tsimpoukelli, M., Kay, J., Merle, A., Moret, J.-M., Noury, S., Pesamosca, F., Pfau, D., Sauter, O., Sommariva, C., Coda, S., Duval, B., Fasoli, A., Kohli, P., Kavukcuoglu, K.,

- Hassabis, D., Riedmiller, M., 2022. Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature* 602, 414–419. <https://doi.org/10.1038/s41586-021-04301-9>
- Derakhshandeh, J.F., Alam, M.M., 2019. A review of bluff body wakes. *Ocean Eng.* 182, 475–488.
- Dong, X.(董欣辉), Hong, H.(洪昊), Deng, X.(邓晓龙), Zhong, W.(钟薇), Hu, G.(胡钢), 2023. Surrogate model-based deep reinforcement learning for experimental study of active flow control of circular cylinder. *Phys. Fluids* 35, 105147. <https://doi.org/10.1063/5.0170316>
- Dong, X.(董欣辉), Wang, Z.(王卓然), Lin, P.(林鹏飞), Wang, Q.(王秋垒), Hu, G.(胡钢), 2024. An interactive platform of deep reinforcement learning and wind tunnel testing. *Phys. Fluids* 36, 115197. <https://doi.org/10.1063/5.0238959>
- Fan, D., Yang, L., Wang, Z., Triantafyllou, M.S., Karniadakis, G.E., 2020. Reinforcement learning for bluff body active flow control in experiments and simulations. *Proc. Natl. Acad. Sci.* 117, 26091–26098. <https://doi.org/10.1073/pnas.2004939117>
- Font, B., Alcántara-Ávila, F., Rabault, J., Vinuesa, R., Lehmkuhl, O., 2025. Deep reinforcement learning for active flow control in a turbulent separation bubble. *Nat. Commun.* 16, 1422. <https://doi.org/10.1038/s41467-025-56408-6>
- Francis, S., Swain, A., 2024. Modelling and harnessing energy from flow-induced vibration, particularly VIV and galloping: An explicit review. *Ocean Eng.* 312, 119290. <https://doi.org/https://doi.org/10.1016/j.oceaneng.2024.119290>
- Frank, M., Drikakis, D., Charissis, V., 2020. Machine-Learning Methods for Computational Science and Engineering. *Computation* 8, 15. <https://doi.org/10.3390/computation8010015>
- Fujimoto, S., Hoof, H., Meger, D., 2018. Addressing function approximation error in actor-critic methods, in: *International Conference on Machine Learning*. PMLR, pp. 1587–1596.
- Gad-el-Hak, M., 2000. *Flow Control*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511529535>
- Garnier, P., Viquerat, J., Rabault, J., Larcher, A., Kuhnle, A., Hachem, E., 2021. A review on deep reinforcement learning for fluid mechanics. *Comput. Fluids* 225, 104973. <https://doi.org/https://doi.org/10.1016/j.compfluid.2021.104973>
- Ghraieb, H., Viquerat, J., Larcher, A., Meliga, P., Hachem, E., 2021. Single-step deep reinforcement learning for open-loop control of laminar and turbulent flows. *Phys. Rev. Fluids* 6, 053902. <https://doi.org/10.1103/PhysRevFluids.6.053902>
- Guastoni, L., Rabault, J., Schlatter, P., Azizpour, H., Vinuesa, R., 2023. Deep reinforcement learning for turbulent drag reduction in channel flows. *Eur. Phys. J. E* 46, 27. <https://doi.org/10.1140/epje/s10189-023-00285-8>
- He, X.-J., Wang, Y.-Z., Hua, Y., Chen, Z.-H., Li, Y.-B., Wu, W.-T., 2023. Policy transfer of

- reinforcement learning-based flow control: From two- to three-dimensional environment. *Phys. Fluids* 35. <https://doi.org/10.1063/5.0147190>
- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. *Neural Networks* 2, 359–366. [https://doi.org/https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/https://doi.org/10.1016/0893-6080(89)90020-8)
- Jebelli, M., Masdari, M., 2022. Interaction of two parallel free oscillating flat plates and VIV of an upstream circular cylinder in laminar flow. *Ocean Eng.* 259, 111876.
- Jebelli, M., Shariloo, K., Masdari, M., 2024. Vortex-plate interaction in wake-induced vibration: A study of single and parallel flat plates behind a cylinder. *Ocean Eng.* 300, 117340. <https://doi.org/https://doi.org/10.1016/j.oceaneng.2024.117340>
- Jeon, J., Rabault, J., Vasanth, J., Alcántara-Ávila, F., Baral, S., Vinuesa, R., 2024. Advanced deep-reinforcement-learning methods for flow control: group-invariant and positional-encoding networks improve learning speed and quality. *arXiv Prepr. arXiv2407.17822*.
- Jiang, H., Cao, S., 2023. Reinforcement learning-based active flow control of oscillating cylinder for drag reduction. *Phys. Fluids* 35. <https://doi.org/10.1063/5.0172081>
- Koizumi, H., Tsutsumi, S., Shima, E., 2018. Feedback Control of Karman Vortex Shedding from a Cylinder using Deep Reinforcement Learning, in: 2018 Flow Control Conference. American Institute of Aeronautics and Astronautics, Reston, Virginia, p. 3691. <https://doi.org/10.2514/6.2018-3691>
- Koza, J., 1994. Genetic programming as a means for programming computers by natural selection. *Stat. Comput.* 4, 87–112. <https://doi.org/10.1007/BF00175355>
- Kurz, M., Offenhäuser, P., Viola, D., Resch, M., Beck, A., 2022. Relexi—A scalable open source reinforcement learning framework for high-performance computing. *Softw. Impacts* 14, 100422.
- Li, J., Zhang, M., 2022. Reinforcement-learning-based control of confined cylinder wakes with stability analyses. *J. Fluid Mech.* 932, A44. <https://doi.org/10.1017/jfm.2021.1045>
- Lillicrap, T.P., 2015. Continuous control with deep reinforcement learning. *arXiv Prepr. arXiv1509.02971*.
- Liu, X., MacArt, J.F., 2024. Adjoint-based machine learning for active flow control. *Phys. Rev. Fluids* 9, 013901. <https://doi.org/10.1103/PhysRevFluids.9.013901>
- Mao, Y., Zhong, S., Yin, H., 2023. DRLFluent: A distributed co-simulation framework coupling deep reinforcement learning with Ansys-Fluent on high-performance computing systems. *J. Comput. Sci.* 74, 102171. <https://doi.org/10.1016/j.jocs.2023.102171>
- Mao, Y., Zhong, S., Yin, H., 2022. Active flow control using deep reinforcement learning with time delays in Markov decision process and autoregressive policy. *Phys. Fluids* 34. <https://doi.org/10.1063/5.0086871>
- McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* 5, 115–133. <https://doi.org/10.1007/BF02478259>
- Mei, Y.-F., Zheng, C., Aubry, N., Li, M.-G., Wu, W.-T., Liu, X., 2021. Active control for

- enhancing vortex induced vibration of a circular cylinder based on deep reinforcement learning. *Phys. Fluids* 33. <https://doi.org/10.1063/5.0063988>
- Mnih, V., 2016. Asynchronous Methods for Deep Reinforcement Learning. *arXiv Prepr. arXiv1602.01783*.
- Mnih, V., 2013. Playing atari with deep reinforcement learning. *arXiv Prepr. arXiv1312.5602*.
- Morton, J., Jameson, A., Kochenderfer, M.J., Witherden, F., 2018. Deep dynamical modeling and control of unsteady fluid flows. *Adv. Neural Inf. Process. Syst.* 31.
- Noack, B.R., 2023. Deep reinforcement learning-based active flow control of vortex-induced vibration of a square cylinder. *Phys. Fluids* 35. <https://doi.org/10.1063/5.0152777>
- Paris, R., Beneddine, S., Dandois, J., 2021. Robust flow control and optimal sensor placement using deep reinforcement learning. *J. Fluid Mech.* 913, A25. <https://doi.org/10.1017/jfm.2020.1170>
- Part 2: Kinds of RL Algorithms — Spinning Up documentation [WWW Document], n.d. URL https://spinningup.openai.com/en/latest/spinningup/rl_intro2.html (accessed 2.11.24).
- Partee, S., Ellis, M., Rigazzi, A., Shao, A.E., Bachman, S., Marques, G., Robbins, B., 2022. Using Machine Learning at scale in numerical simulations with SmartSim: An application to ocean climate modeling. *J. Comput. Sci.* 62, 101707. <https://doi.org/https://doi.org/10.1016/j.jocs.2022.101707>
- Pivot, C., Cordier, L., Mathelin, L., 2017. A continuous reinforcement learning strategy for closed-loop control in fluid dynamics, in: 35th AIAA Applied Aerodynamics Conference. American Institute of Aeronautics and Astronautics, Reston, Virginia, p. 3566. <https://doi.org/10.2514/6.2017-3566>
- Qin, S., Wang, S., Sun, G., 2024. An Application of Data Driven Reward of Deep Reinforcement Learning by Dynamic Mode Decomposition in Active Flow Control, in: International Conference on Computational & Experimental Engineering and Sciences. Springer, pp. 1195–1217. https://doi.org/10.1007/978-3-031-44947-5_92
- Rabault, J., Kuchta, M., Jensen, A., Réglade, U., Cerardi, N., 2019. Artificial neural networks trained through deep reinforcement learning discover control strategies for active flow control. *J. Fluid Mech.* 865, 281–302. <https://doi.org/10.1017/jfm.2019.62>
- Rabault, J., Kuhnle, A., 2023. Deep Reinforcement Learning Applied to Active Flow Control, in: Mendez, M.A., Ianiro, A., Noack, B.R., Brunton, S.L. (Eds.), *Data-Driven Fluid Mechanics: Combining First Principles and Machine Learning*. Cambridge University Press, Cambridge, pp. 368–390. [https://doi.org/DOI: 10.1017/9781108896214.025](https://doi.org/DOI:10.1017/9781108896214.025)
- Rabault, J., Kuhnle, A., 2019. Accelerating deep reinforcement learning strategies of flow control through a multi-environment approach. *Phys. Fluids* 31. <https://doi.org/10.1063/1.5116415>
- Rashidi, S., Hayatdavoodi, M., Esfahani, J.A., 2016. Vortex shedding suppression and wake control: A review. *Ocean Eng.* 126, 57–80. <https://doi.org/https://doi.org/10.1016/j.oceaneng.2016.08.031>

- Ren, F., Rabault, J., Tang, H., 2021a. Applying deep reinforcement learning to active flow control in weakly turbulent conditions. *Phys. Fluids* 33. <https://doi.org/10.1063/5.0037371>
- Ren, F., Wang, C., Tang, H., 2021b. Bluff body uses deep-reinforcement-learning trained active flow control to achieve hydrodynamic stealth. *Phys. Fluids* 33. <https://doi.org/10.1063/5.0060690>
- Ren, F., Zhang, F., Zhu, Y., Wang, Z., Zhao, F., 2024. Enhancing heat transfer from a circular cylinder undergoing vortex induced vibration based on reinforcement learning. *Appl. Therm. Eng.* 236, 121919. <https://doi.org/10.1016/j.applthermaleng.2023.121919>
- Richard S. Sutton, A.G.B., 2018. Reinforcement Learning: An Introduction, second edi. ed, A Bradford Book. Bradford Books.
- Sakamoto, T., Okabayashi, K., 2024. Optimization of fluid control laws through deep reinforcement learning using dynamic mode decomposition as the environment. *AIP Adv.* 14, 115204. <https://doi.org/10.1063/5.0237682>
- Salehi, S., 2024. An efficient intrusive deep reinforcement learning framework for OpenFOAM. *Meccanica*. <https://doi.org/10.1007/s11012-024-01830-1>
- Schulman, J., 2015. Trust Region Policy Optimization. *arXiv Prepr. arXiv1502.05477*.
- Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O., 2017. Proximal policy optimization algorithms. *arXiv Prepr. arXiv1707.06347*.
- Shams, M., Elsheikh, A.H., 2023. Gym-preCICE: Reinforcement learning environments for active flow control. *SoftwareX* 23, 101446. <https://doi.org/10.1016/j.softx.2023.101446>
- Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., Hassabis, D., 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 484–489. <https://doi.org/10.1038/nature16961>
- Suárez, P., Alcantara-Avila, F., Miró, A., Rabault, J., Font, B., Lehmkuhl, O., Vinuesa, R., 2024. Active flow control for drag reduction through multi-agent reinforcement learning on a turbulent cylinder at $Re_D = 3900$. *arXiv Prepr. arXiv2405.17655*.
- Suárez, P., Alcántara-Ávila, F., Miró, A., Rabault, J., Font, B., Lehmkuhl, O., Vinuesa, R., 2023. Active flow control for three-dimensional cylinders through deep reinforcement learning. *arXiv Prepr. arXiv2309.02462*. <https://doi.org/https://doi.org/10.48550/arXiv.2309.02462>
- Sutton, R.S., 1988. Learning to predict by the methods of temporal differences. *Mach. Learn.* 3, 9–44. <https://doi.org/10.1007/BF00115009>
- Tang, H., Rabault, J., Kuhnle, A., Wang, Y., Wang, T., 2020. Robust active flow control over a range of Reynolds numbers using an artificial neural network trained through deep reinforcement learning. *Phys. Fluids* 32. <https://doi.org/10.1063/5.0006492>
- Tokarev, M., Palkin, E., Mullyadzhyanov, R., 2020. Deep Reinforcement Learning Control of

- Cylinder Flow Using Rotary Oscillations at Low Reynolds Number. *Energies* 13, 5920. <https://doi.org/10.3390/en13225920>
- Varela, P., Suárez, P., Alcántara-Ávila, F., Miró, A., Rabault, J., Font, B., García-Cuevas, L.M., Lehmkuhl, O., Vinuesa, R., 2022. Deep Reinforcement Learning for Flow Control Exploits Different Physics for Increasing Reynolds Number Regimes. *Actuators* 11, 359. <https://doi.org/10.3390/act11120359>
- Vasanth, J., Rabault, J., Alcántara-Ávila, F., Mortensen, M., Vinuesa, R., 2024. Multi-agent Reinforcement Learning for the Control of Three-Dimensional Rayleigh–Bénard Convection. *Flow, Turbul. Combust.* <https://doi.org/10.1007/s10494-024-00619-2>
- Vignon, C., Rabault, J., Vasanth, J., Alcántara-Ávila, F., Mortensen, M., Vinuesa, R., 2023a. Effective control of two-dimensional Rayleigh–Bénard convection: Invariant multi-agent reinforcement learning is all you need. *Phys. Fluids* 35, 65146. <https://doi.org/10.1063/5.0153181>
- Vignon, C., Rabault, J., Vinuesa, R., 2023b. Recent advances in applying deep reinforcement learning for flow control: Perspectives and future directions. *Phys. Fluids* 35. <https://doi.org/10.1063/5.0143913>
- Wang, J., Feng, L., 2018. *Flow Control Techniques and Applications, Flow Control Techniques and Applications, Cambridge Aerospace Series.* Cambridge University Press, Cambridge. <https://doi.org/10.1017/9781316676448>
- Wang, J., Geng, L., Ding, L., Zhu, H., Yurchenko, D., 2020. The state-of-the-art review on energy harvesting from flow-induced vibrations. *Appl. Energy* 267, 114902. <https://doi.org/10.1016/j.apenergy.2020.114902>
- Wang, Q., Yan, L., Hu, G., Chen, W., Rabault, J., Noack, B.R., 2024. Dynamic feature-based deep reinforcement learning for flow control of circular cylinder with sparse surface pressure sensing. *J. Fluid Mech.* 988, A4. <https://doi.org/10.1017/jfm.2024.333>
- Wang, Q., Yan, L., Hu, G., Li, C., Xiao, Y., Xiong, H., Rabault, J., Noack, B.R., 2022. DRLinFluids: An open-source Python platform of coupling deep reinforcement learning and OpenFOAM. *Phys. Fluids* 34. <https://doi.org/10.1063/5.0103113>
- Wang, Y.-Z., Hua, Y., Aubry, N., Chen, Z.-H., Wu, W.-T., Cui, J., 2022. Accelerating and improving deep reinforcement learning-based active flow control: Transfer training of policy network. *Phys. Fluids* 34. <https://doi.org/10.1063/5.0099699>
- Weiner, A., 2024. drlFoam. GitHub repository. [WWW Document]. URL <https://github.com/OFDataCommittee/drlfoam>
- Xia, C., Zhang, J., Kerrigan, E.C., Rigas, G., 2024. Active flow control for bluff body drag reduction using reinforcement learning with partial measurements. *J. Fluid Mech.* 981, A17. <https://doi.org/10.1017/jfm.2024.69>
- Xie, F., Zheng, C., Ji, T., Zhang, X., Bi, R., Zhou, H., Zheng, Y., 2023. Deep Reinforcement Learning: A New Beacon for Intelligent Active Flow Control. *Aerosp. Res. Commun.* 1, 11130. <https://doi.org/10.3389/arc.2023.11130>
- Xie, Z., Hu, H., Chen, J., Song, J., Lu, T., Ren, F., 2023. Applying reinforcement learning to

- mitigate wake-induced lift fluctuation of a wall-confined circular cylinder in tandem configuration. *Phys. Fluids* 35. <https://doi.org/10.1063/5.0150244>
- Xu, H., Zhang, W., Deng, J., Rabault, J., 2020. Active flow control with rotating cylinders by an artificial neural network trained by deep reinforcement learning. *J. Hydrodyn.* 32, 254–258. <https://doi.org/10.1007/s42241-020-0027-z>
- Yan, L., Hu, G., Chen, W., Noack, B.R., 2024a. Deep reinforcement transfer learning for active flow control of a 3D square cylinder under state dimension mismatch. *arXiv Prepr. arXiv2401.12543*.
- Yan, L., Li, Y., Liu, B., Hu, G., 2024b. Aerodynamic force reduction of rectangular cylinder using deep reinforcement learning-controlled multiple jets. *Phys. Fluids* 36.
- Yan, L., Zhang, X., Song, J., Hu, G., 2024c. Active flow control of square cylinder adaptive to wind direction using deep reinforcement learning. *Phys. Rev. Fluids* 9, 94607. <https://doi.org/10.1103/PhysRevFluids.9.094607>
- Ye, S., Zhang, Z., Wang, Y., Huang, C., 2024. Reducing flow fluctuation using deep reinforcement learning with a CNN-based flow feature model. *Ocean Eng.* 306, 118089. <https://doi.org/https://doi.org/10.1016/j.oceaneng.2024.118089>
- Yousif, M.Z., Kolesova, P., Yang, Y., Zhang, M., Yu, L., Rabault, J., Vinuesa, R., Lim, H.-C., 2023. Optimizing flow control with deep reinforcement learning: Plasma actuator placement around a square cylinder. *Phys. Fluids* 35. <https://doi.org/10.1063/5.0174724>
- Zhang, F., Hu, H., Zhang, H., Zhang, M., Song, J., Meng, Y., 2024. Predicting pressure fields from incomplete velocity fields based on deep convolutional neural network. *Ocean Eng.* 309, 118578. <https://doi.org/https://doi.org/10.1016/j.oceaneng.2024.118578>
- Zhao, F., Zhou, Y., Ren, F., Tang, H., Wang, Z., 2024. Mitigating the lift of a circular cylinder in wake flow using deep reinforcement learning guided self-rotation. *Ocean Eng.* 306, 118138. <https://doi.org/https://doi.org/10.1016/j.oceaneng.2024.118138>
- Zhao, M., 2023. A review of recent studies on the control of vortex-induced vibration of circular cylinders. *Ocean Eng.* 285, 115389. <https://doi.org/https://doi.org/10.1016/j.oceaneng.2023.115389>
- Zheng, C., Ji, T., Xie, F., Zhang, X., Zheng, H., Zheng, Y., 2021. From active learning to deep reinforcement learning: Intelligent active flow control in suppressing vortex-induced vibration. *Phys. Fluids* 33. <https://doi.org/10.1063/5.0052524>
- Zheng, C., Xie, F., Ji, T., Zhang, X., Lu, Y., Zhou, H., Zheng, Y., 2022. Data-efficient deep reinforcement learning with expert demonstration for active flow control. *Phys. Fluids* 34. <https://doi.org/10.1063/5.0120285>