

Robust and Adaptive Deep Reinforcement Learning for Enhancing Flow Control around a Square Cylinder with Varying Reynolds Numbers

Wang Jia (贾旺)¹ and Hang Xu (徐航)¹

State Key Lab of Ocean Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China

(*Electronic mail: hangxu@sjtu.edu.cn)

(Dated: 30 April 2024)

The present study applies a Deep Reinforcement Learning (DRL) algorithm to Active Flow Control (AFC) of a two-dimensional flow around a confined square cylinder. Specifically, the Soft Actor-Critic (SAC) algorithm is employed to modulate the flow of a pair of synthetic jets placed on the upper and lower surfaces of the confined squared cylinder in flow configurations characterized by Re of 100, 200, 300, and 400. The investigation starts with an analysis of the baseline flow in the absence of active control. It is observed that at $Re = 100$ and $Re = 200$, the vortex shedding exhibits mono-frequency characteristics. Conversely, at $Re = 300$ and $Re = 400$, the vortex shedding is dominated by multiple frequencies, which is indicative of more complex flow features. With the application of the SAC algorithm, we demonstrate the capability of DRL-based control in effectively suppressing vortex shedding, while significantly diminishing drag and fluctuations in lift. Quantitatively, the data-driven active control strategy results in a drag reduction of approximately 14.4%, 26.4%, 38.9%, and 47.0% for $Re = 100, 200, 300$, and 400 , respectively. To understand the underlying control mechanism, we also present detailed flow field comparisons, which showcase the adaptability of DRL in devising distinct control strategies tailored to the dynamic conditions at varying Re . These findings substantiate the proficiency of DRL in controlling chaotic, multi-frequency dominated vortex shedding phenomena, underscoring the robustness of DRL in complex AFC problems.

I. INTRODUCTION

The examination of fluid dynamics around square cylinders is of paramount importance within both scholarly inquiry and practical engineering realms, encompassing applications such as maritime structures, skyscrapers, towers, bridge piers, and heat exchangers.¹⁻⁴ The vortex shedding phenomenon of square cylinders, that is, the alternating shedding vortices formed behind the square cylinder, will generate periodic lateral forces, which may cause structural fatigue, resonance, noise, reduced performance, and safety risks.⁵ In view of this, mitigating the adverse effects of fluid forces around square cylinders has become a critical research effort. Flow control technology changes the flow pattern of fluid and is widely used in marine, construction, aviation, automobile, energy and other fields.⁶ It can reduce drag, improve aerodynamic performance, increase lift or load-bearing capacity, reduce vibration and noise, and control heat and mass transfer, which is of great significance to improving efficiency, performance and sustainable development. Active flow control technology, with its ability to sense and accurately regulate fluid flow in real time, demonstrates significant performance advantages over traditional passive control methods.^{6,7}

While active flow control technology has advanced significantly, its widespread adoption has been limited by high cost and technical complexity, reliance on high-performance computing resources, the need for a deep understanding of fluid dynamics during design and optimization, and the environment challenges with variability in operating conditions.^{8,9} The rapid evolution of machine learning in recent years has not only achieved significant advancements within the domain of computer science but has also exerted profound impacts on specialized fields such as fluid dynamics and aerodynamics. In particular, the swift progression of machine learning has introduced new opportunities for research in active flow control, catalyzing progress within this sector.¹⁰ The spatiotemporal evolution of flow is predominantly governed by the nonlinear Navier-Stokes equations, which inherently encapsulate complex characteristics such as high-dimensionality, multifrequency, multimodality, and multiscale phenomena. These complexities substantially elevate the challenges and intricacies associated with flow control, rendering traditional analytical and numerical approaches computationally intensive and sometimes intractable.^{8,9,11,12} Under such circumstances, Deep Reinforcement Learning (DRL) offers a promising new solution to the flow control problem. Previous studies have demonstrated the potential of DRL in Active Flow Control (AFC) applications. For example, DRL has been successfully used to control flow separation around airfoils, reduce drag on bluff bodies, and optimize flow control actuators.¹³⁻¹⁵ These studies demonstrate that DRL can effectively learn complex control strategies and adapt to different flow conditions, outperforming traditional control methods in terms of performance and efficiency.^{16,17} Overall, the clever application of DRL in AFC demonstrates its potential to solve complex decision-making problems.

Deep learning (DL) has demonstrated remarkable proficiency in extracting nonlinear features from complex systems, such as chaotic systems.^{10,18} DL can learn and represent highly abstract features, thereby capturing the nonlinear patterns and dynamic changes inherent in complex systems.¹⁹⁻²¹ This feature enables DL to effectively extract features of complex systems.^{20,21} Reinforcement learning (RL) is a key subfield of machine learning that has attracted significant attention in recent years, particularly after the integration of deep neural networks to create the Deep Q-Network (DQN) artificial intelligence agent. The DQN, by

learning from high-dimensional inputs, has demonstrated remarkable performance in various complex tasks, marking an important milestone.²² The triumph of AlphaGo over the world Go champion in 2016 further highlighted RL's adeptness at complex decision-making and strategic planning.²³ These developments have significantly propelled the interest and perceived impact of DRL within the AI domain. The synergistic ability of DL combined with RL can solve decision-making problems that were previously considered intractable, especially in non-linear and high-dimensional fusion scenarios.^{24,25}

DRL provides an innovative approach to optimizing fluid flow control strategies in various areas, including wall turbulence control²⁶, channel flow²⁷, convection control²⁸, heat transfer^{28,29}, combustion turbulence^{30,31}, and bluff bodies³². The first application of neural networks for active turbulence control on walls was introduced by Lee *et al.*³³. They used blowing/suction devices to lower shear stress in turbulent channel flows. Their approach, leveraging Artificial Neural Networks (ANN) for closed-loop control, resulted in a notable drag reduction of up to 20%. Lee, Kim, and Lee³⁴ used DRL for turbulence control, achieving a 20% reduction in drag. This study demonstrates the efficacy of DRL in turbulence control and provides a physical interpretation of its mechanisms. Guastoni *et al.*, Guastoni *et al.*^{26,27} employed DRL algorithms to reduce the drag of turbulent flows confined within channels, achieving drag reductions of 43% in minimal channels and 30% in larger channels. Gerben Beintema and Toschi³⁵ successfully utilized small temperature fluctuations to reduce heat transfer in a two-dimensional Rayleigh-Bénard system, applying advanced DRL algorithms. Wang *et al.*²⁸ harnessed a DQN in synergy with three intricate techniques to elucidate the core mechanisms of heat and mass transfer in the ambit of closed-loop active control interactions.

DRL has also been successfully applied to the study of flow control strategies in confined cylindrical flow and blunt body flow scenarios. Rabault *et al.*³⁶ pioneered the use of DRL for AFC, achieving an 8% reduction in cylinder resistance in a 2D simulation at Re of 100. This work is the first to release open-source code that combines DRL with CFD coupling. In his review paper, he described how to use the class functions in the `Tensorforce` library to embed your own CFD environment into the DRL algorithm.⁸ Li and Zhang³⁷ successfully controlled the vortex shedding phenomenon behind a confined cylinder by embedding the physical information of hydrodynamics into the design of the reward function. Not only embedded physical information in the DRL control strategy but also validated the robustness of DRL-based control methods for various blockage ratios of confined cylinders. Paris, Beneddine, and Dandois³⁸ investigated the introduction of S-PPO-CMA optimization for sensor placement. At a Re of 120, they utilized this algorithm to train a control strategy that achieved an 18.4% reduction in drag. Wang *et al.*³⁹ introduced DRLinFluids, a Python-based platform for flow control, achieving drag reductions of approximately 8% and 13.6% for circular and square cylinders, respectively, at $Re = 100$.

Rabault *et al.*, Tang *et al.*, Heess *et al.*, Ren, Rabault, and Tang, Jia and Xu^{36,40–43} have applied the Proximal Policy Optimization (PPO) algorithm to AFC in the flow around a cylinder, and the research results consistently demonstrate a reduction in drag of approximately 8% at $Re = 100$. Tang *et al.*⁴⁰ utilized a flow control strategy based on DRL to actively reduce drag. They achieved approximately 21.6%, 32.7%, and 38.7% reduction in drag at $Re = 200, 300, 400$, respectively. Ren, Rabault, and Tang⁴² explored the efficacy of DRL in managing weak turbulence at $Re = 1000$, finding that DRL agents could identify strategies to reduce drag by up to 30%. The aforementioned studies comprehensively examined strategies for AFC using DRL, with a focus on flow around a confined cylinder as the test case. These investigations spanned multiple dimensions, including the robustness of the Re ranging from 100 to 1000, the exploration of probe positions and quantities, the incorporation of physical information into the reward function, and the enhancement of the PPO algorithm. The study on flow around a confined cylinder offers a comprehensive observation and analysis of DRL's application in AFC, providing valuable insights for this field.

Square cylinder flow and circular cylinder flow exhibit significant differences in physical phenomena and flow characteristics, primarily due to their geometric shapes and the fundamental principles of fluid dynamics.^{1–3} Square cylinders are prone to flow instabilities at lower Reynolds numbers, such as the von Kármán vortex street, with separation points fixed at sharp edges, while circular cylinders see separation points that move rearward with increasing Reynolds number.^{1–4} Additionally, vortex shedding around square cylinders occurs earlier and is more complex compared to the regular vortex patterns seen with circular cylinders.^{1–4,44–48} The influence of Reynolds number on drag and lift is more pronounced in square cylinder flow.^{1–4,49} Understanding these differences is crucial for the design and optimization of fluid control systems.

In recent years, there has been some research applying DRL algorithms to AFC control strategies within the context of flow around square cylinders. In the study conducted by Wang *et al.*³⁹, the investigation was exclusively focused on a Re of 100, where the drag reduction rate was observed to be 13.7%, and the wake vortices continued to exhibit an oscillating state. Chen *et al.*⁵⁰ used a DRL-based AFC method to reduce the vortex-induced vibration of a square cylinder when the $Re = 100$, and placed the injection actuators at the front, middle, and rear corners of the side walls of the square cylinder. The control results show that when the jet actuator is located at the rear corner, the control effect is faster and more obvious than at the front corner and middle position. The control results show that when the jet actuator is located at the rear corner, the drag reduction is the most effective with the least actuator energy consumption. Yan *et al.*⁵¹ studied the effect of synthetic jet placement on drag and lift around a square cylinder using Soft Actor-Critic (SAC) algorithm. They found jets at the front edge more effective than those placed at the rear, reducing drag coefficient by up to 44.4%, 60.6%, and 57.8% at Re of 500, 1000, and 2000, respectively. In another work by Yan *et al.*⁵², they investigated the effect of using a combination of eight jets located at four corners of rectangular cylinders at Re of 1000. With four or eight jets activated, they were able to reduce drag by 63.2% and 77.1% respectively on a square cylinder.

Despite the significant achievements of AFC technology based on DRL in reducing the drag and mitigating lift fluctuations

around square cylinders, a primary issue remains: the instability of the wake flow and the phenomenon of vortex shedding have not been effectively controlled. In the circular cylinder case, vortex shedding can be almost fully suppressed and only very weak jet flow is required to maintain the controlled state³⁹. Such a fully controlled vortex shedding pattern with minimal control energy has not been observed in existing literature for the square cylinder test case, even at a relatively low Re of 100. To the best of our knowledge, only one test case in the work by Chen *et al.*⁵⁰ showed signs of stabilized wake flow at $Re = 100$. The current state of the art is capable of sustaining the drag reduction throughout the control period, but with substantial vortex shedding in the downstream, which could induce adverse vibrations on the square cylinder itself or any downstream structures. Due to the instability of the wake flow and the persistence of vortex shedding, delving into the physical mechanisms behind DRL-based AFC strategies also remains a challenge. Therefore, a detailed investigation into the control effects of DRL-based AFC technology in managing the flow around square cylinders is particularly crucial to achieve a more stable wake flow and to deepen the understanding of its physical mechanisms.

In this study, we aim to explore the potential of DRL-based AFC technology in controlling the flow around square cylinders to suppress vortex shedding with minimal actuator excitation energy. To investigate the control mechanisms devised by DRL at various flow conditions, we examine flows with Re from 100 to 400. The development of an efficient, energy-saving, stable, and robust active flow control strategy is intended to deepen the understanding of the control mechanisms behind DRL-based AFC technology and promote its application in engineering practices, thereby advancing the technology both theoretically and practically. In Section II, the problem description is presented, followed by an in-depth overview of DRL and its formulation in addressing the AFC challenge. The main results and discussions are developed in Section III. Finally, Section IV concludes the paper by summarizing the key findings, discussing their implications, and suggesting potential directions for future research.

II. PROBLEM SETUP AND METHODOLOGY

A. Numerical simulation setup

a. Model configuration The configuration studied in this paper is the two-dimensional (2D) flow around a square cylinder. The side length of the square cylinder is D . Fig. 1(a) illustrates the computational domain and coordinate system of the numerical model. The origin of the Cartesian coordinate system is established at the center of the square cylinder. The square cylinder is immersed in a rectangular domain of dimensions $40D$ (along the x -axis) and $20D$ (along the y -axis). Fig. 1(b) is a detailed view of the square cylinder, showing the position of a symmetrical pair of jets.

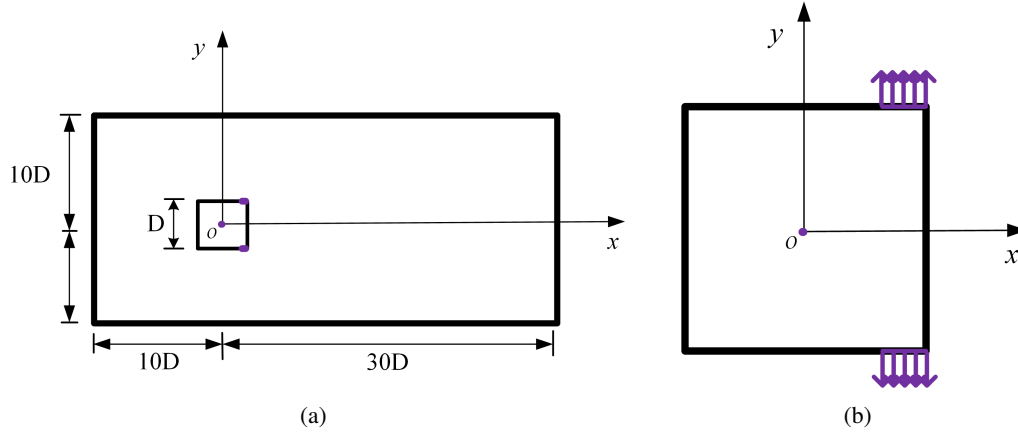


FIG. 1: Description of the 2D square cylinder numerical setup: (a) The computational domain has a length of $40D$ and a width of $20D$, with the square cylinder located at the center of the domain. Each side of the square cylinder is of length D . A uniform upstream velocity flows in the positive x -direction past the square cylinder towards the downstream. (b) This figure only shows the starting position of the jets on the square cylinder, with the scale ratios not reflecting actual proportions.

b. Governing equations To establish the physical model for numerical simulations, it is assumed that the flow is viscous and incompressible. Flow simulations are executed by solving the two-dimensional incompressible Navier-Stokes equations within the computational domain. The governing equations can be represented as follows:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot (\nabla \mathbf{u}) = -\nabla p + \frac{1}{Re} \Delta \mathbf{u}, \quad (1)$$

$$\nabla \cdot \mathbf{u} = 0. \quad (2)$$

Here, $\mathbf{u} = (U, V)$ represents the velocity vector, where U and V are the velocity components along the x - and y -axes, respectively. t denotes the dimensionless time unit. p is the thermodynamic pressure. $Re = \frac{\bar{U}D}{\nu}$ is the Re , where D is the side length of the square, ν denotes the kinematic viscosity of the fluid, and $\bar{U}$ is the mean velocity at the inlet.

c. Boundary conditions The left boundary of the computational domain is defined as the velocity inlet Γ_{in} , with the boundary conditions, $U = U_0$, $V = 0$. In this study, a uniform flow with a velocity of $U_0 = 2$ is imposed. The right boundary is defined as the pressure outlet Γ_{out} , with the boundary condition, the mean static pressure is set to zero. The spanwise boundaries are defined as symmetric boundary conditions Γ_s , with the boundary condition, the velocity component normal to each variable is zero. The positions of the two synthetic jet actuators near the trailing edge of the square cylinder are defined as the jet boundaries Γ_i ($i = 1, 2$). The boundary conditions impose a uniform velocity distribution along the normal direction. The width of the synthetic jets velocity distribution is $0.04D$. The jet velocities of the synthetic jets can be positive or negative, corresponding to blowing or suction, respectively. Maintaining the balance of the net mass flow rate for all jets ensures a zero net flow rate for the system, i.e., $V_{\Gamma_1} = -V_{\Gamma_2}$. Apart from the jets, the rest of the square cylinder adopts no-slip solid wall boundary conditions Γ_w , with the boundary conditions, $U = 0$, $V = 0$.

d. Solver details In this study, the incompressible flow solver is based on the open-source CFD package OpenFOAM, as described by Jasak *et al.*^{53,54} OpenFOAM is a widely used and validated CFD software package that provides robust numerical algorithms for solving the Navier-Stokes equations. The solver in OpenFOAM utilizes the finite volume method to discretize the computational domain into a mesh made up of control volumes. To ensure numerical stability, the time step is chosen as $\Delta t = 0.0005$.

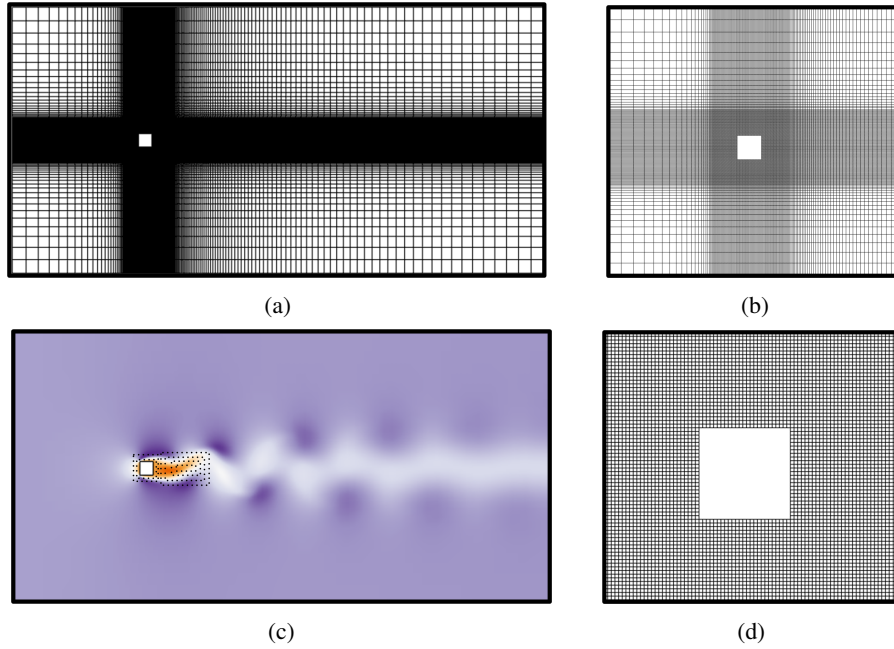


FIG. 2: Discretization of the computational domain. (a) Meshing of the global computational domain; (b) Mesh division near the square cylinder; (c) The velocity field around the square cylinder after the initiation of flow without flow control. The black dots represent the positions of 201 velocity probes; (d) More detailed mesh division around the square cylinder.

e. Grid system The computational domain is discretized using structured meshes and is refined around the surface of the square cylinder and in the wake flow region downstream of the cylinder, as shown in Fig. 2. The computational domain is discretized into 23264 grid elements using structured quadrilateral cells. Fig. 2(a) presents the global discretization of the computational domain, while Fig. 2(b) depicts a detailed treatment around the square cylinder to facilitate consideration of the impact of on flow simulation. Fig. 2(c) displays the velocity field around the square cylinder in the absence of flow control, primarily to show the placement of 201 probes. Fig. 2(d) details the mesh around the square cylinder with a closer view.

f. Quantities of interest The lift coefficient (C_L) and drag coefficient (C_D) are defined as

$$C_L = \frac{F_L}{0.5\rho\bar{U}^2D}, \quad C_D = \frac{F_D}{0.5\rho\bar{U}^2D}. \quad (3)$$

Here, F_L and F_D represent the lift and drag forces integrated on the surface of the cylinder, respectively, and ρ is the fluid density. The Strouhal number (St) is used to describe the characteristic frequency of oscillatory flow phenomena and is defined

as follows:

$$St = \frac{f \cdot D}{U}. \quad (4)$$

Where f_s is the shedding frequency calculated based on the periodic evolution of the C_L . D is the characteristic length, which is the side length of the square cylinder, and $\bar{U}$ is the mean velocity of the upstream flow.

g. *Grid independence* To investigate grid convergence and validate the numerical methods, the flow under $Re = 100$ was simulated to compute the desired quantities of interest. These results were then compared with the computations carried out by other published paper.

TABLE I: Mesh convergence and flow parameters for the 2D flow around a square cylinder at $Re = 100$.

Case	Mesh resolution	Δt	$\bar{C}_D$	$C_{D,\max}$	$C_{L,\max}$	St
Wang <i>et al.</i> ³⁹	23 125	0.0005	1.549	-	-	-
Coarse	8465	0.0005	1.532	1.562	0.326	0.145
Main	23 264	0.0005	1.548	1.559	0.321	0.142
Fine	35 179	0.0005	1.549	1.561	0.320	0.141

Note: Δt represents the time step size for CFD.

Table. I lists the simulation results using meshes of three different resolutions and compares them with the results calculated by Wang *et al.* $C_{D,\max}$ and $C_{L,\max}$ correspond to the maximum values of C_D and C_L , respectively. The C_D and C_L calculated with the coarse mesh deviate significantly from the results obtained by Wang *et al.*³⁹, rendering the precision unreliable. The fine mesh entails a substantial number of grid points, which is disadvantageous for subsequent training processes. This study aims to select discretization results with as few grid points as possible while ensuring accuracy, as this greatly influences the speed of training. The resolution of the main mesh used in this work is sufficiently high to achieve good agreement with the benchmark data.

B. Deep Reinforcement learning

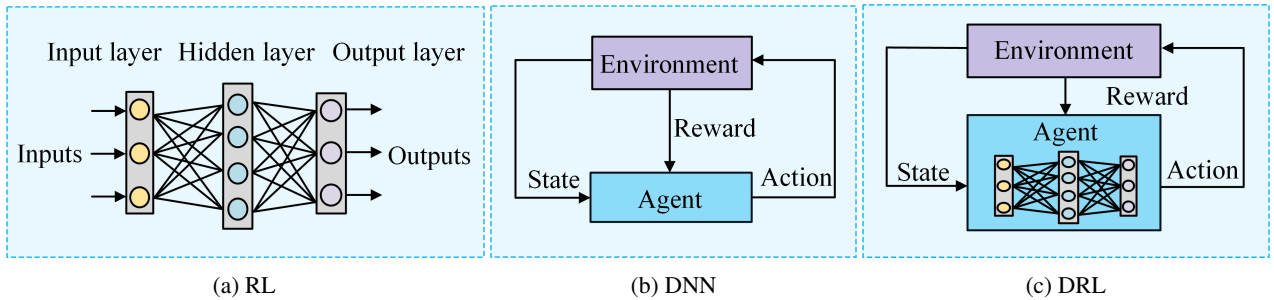


FIG. 3: Illustration of deep learning, RL, and DRL. (a) In RL, the agent interacts with the environment to learn and make decisions; (b) The relationship between inputs and outputs in DNN is established through the connections between layers of the neural network; (c) In DRL, leveraging deep neural networks to learn the mapping between states and actions is accomplished through the process of training the neural network.

Fig. 3 explains the basic elements of RL, DL, and DRL. As described in Fig. 3(a), an agent interacts with the environment by observing its state s_t and taking corresponding action a_t . In Fig. 3(b), it illustrates the relationship between input and output, showcasing how deep neural networks adeptly map high-dimensional input data to their corresponding outputs by learning complex mappings and automatically extracting features through advanced representations and patterns. Fig. 3(c) illustrates the method of applying DL within the realm of RL, merging the decision-making optimization capabilities of RL with the representational learning abilities of DL.

From Fig. 3, we can understand that DRL is a combination of the decision-making capability of DR and the mapping capability of DL. How to perform training using DL and RL will be described in the specific algorithm. This paper adopts the SAC algorithm, a RL algorithm proposed by Haarnoja *et al.*⁵⁵ in 2018. The SAC algorithm is an advanced framework that merges the principles of DRL with maximum entropy theory. It leverages deep neural networks to serve as approximators for value and policy functions, aiming to maximize the entropy of a_t to promote exploration and the development of robust strategies. The SAC

algorithm exhibits exceptional performance and applicability in handling problems within continuous action spaces, fostering exploration and the learning of diverse strategies, adaptively adjusting entropy coefficients, and facilitating offline learning^{55,56}. Consequently, this paper selects the SAC algorithm to control fluid flow.

The mathematical foundation of RL is encapsulated by the Markov Decision Process (MDP), a model that rigorously defines the dynamics of decision-making environments where outcomes are partly random and partly under the control of a decision maker^{57–59}. An MDP is characterized by a tuple (S, A, R, P, ρ_0) , where S represents the set of all possible s_t within the environment. A encapsulates the pertinent information describing the current circumstances of both the environment and the agent. A represents the set of possible a_t that an agent can take. $R : S \times A \rightarrow \mathbb{R}$ is the reward function, providing the immediate reward r_t received after an agent transitions from s_t to s_{t+1} due to a_t at time t . r_t is the feedback signal the agent receives based on the a_t it takes and its current s_t . $P : S \times A \times S \rightarrow \mathbb{R}$ defines the s_t transition probability function, indicating the likelihood of moving from s_t to s_{t+1} upon taking a_t . ρ_0 is the initial s_t distribution function, ensuring that the sum of probabilities across all s_t in S equals one ($\sum_{s \in S} \rho_0(s) = 1$).

The MDP framework provides a precise structure for RL problems, defining key elements such as S spaces, A spaces, reward functions, and state transition probabilities. Within this framework, the SAC algorithm introduces entropy regularization, which encourages exploration by maximizing policy entropy. This regularization enhances action diversification, preventing premature convergence to local optima and greatly improving the algorithm’s exploration capability. The core optimization goal of the SAC algorithm is as follows^{55,60,61}:

$$\pi^* = \arg \max_{\pi} \sum_t \mathbb{E}_{(s_t, a_t) \sim \rho_{\pi}} [r(s_t, a_t) + \alpha \mathcal{H}(\pi(\cdot | s_t))], \quad (5)$$

where π^* denotes the optimal policy, $r(s_t, a_t)$ represents the reward for taking a_t in s_t , α signifies the weight of the entropy term in the objective, and $\mathcal{H}(\pi(\cdot | s_t))$ is the entropy of policy π given s_t . This inclusion of entropy in its objective ensures SAC maintains a balance between exploiting known rewards and exploring novel a_t , rendering it highly versatile for a broad spectrum of tasks.

In the SAC algorithm, the state value function $V(s_t)$ and the action value function $Q(s_t, a_t)$ are pivotal in determining the optimal policy. These functions facilitate the algorithm’s decision-making process by estimating the expected returns from s_t and a_t , leveraging the Bellman equation for recursive computation. The SAC algorithm approximates these value functions using DL techniques, enabling efficient handling of complex, high-dimensional environments. Below are the formulations of these value functions within the SAC framework:

- $Q(s_t, a_t)$ estimates the expected return of taking an a_t in s_t and thereafter following the current policy. In the SAC algorithm, $Q(s_t, a_t)$ is updated based on the Bellman optimality principle, incorporating both the immediate r_t and the discounted future s_t value, adjusted for the entropy of subsequent policy actions. The typical formulation for updating $Q(s_t, a_t)$ in SAC is:

$$Q(s_t, a_t) = r(s_t, a_t) + \gamma \mathbb{E}_{s_{t+1}} [V(s_{t+1})] \quad (6)$$

where $r(s_t, a_t)$ is the reward received after executing a_t in s_t , γ is the discount factor, and $\mathcal{E}$ denotes the environment dynamics that govern the transition to the next state s_{t+1} . These formulations underscore the SAC algorithm’s dual emphasis on maximizing expected returns and promoting policy entropy, thus ensuring a balanced approach to exploration and exploitation in continuous action spaces. The use of DL models to approximate $V(s_t)$ and $Q(s_t, a_t)$ allows SAC to effectively navigate and learn from complex, high-dimensional S and A spaces.

- $V(s_t)$ in SAC is aimed at evaluating the expected return of being in s_t and acting according to the current policy π . It is defined as the expected value of the $Q(s_t, a_t)$, minus the product of the temperature parameter α and the policy’s entropy, which encourages exploration by valuing states based not only on immediate r_t but also on the diversity of a_t taken. The formal definition is:

$$V(s_t) = \mathbb{E}_{a_t \sim \pi} [Q(s_t, a_t) - \alpha \log \pi(a_t | s_t)] \quad (7)$$

In the SAC algorithm, the learning of the $V(s_t)$, and $Q(s_t, a_t)$ is facilitated through deep neural networks.⁵⁵ These networks utilize experience samples collected from the environment, updating network parameters via the back-propagation algorithm to minimize the loss functions mentioned above. The capability of DL allows SAC to handle high-dimensional S and A spaces, and by employing an end-to-end learning approach, it directly extracts useful features from raw observation data, greatly enhancing the algorithm’s generalization ability and applicability.^{55,56,61}

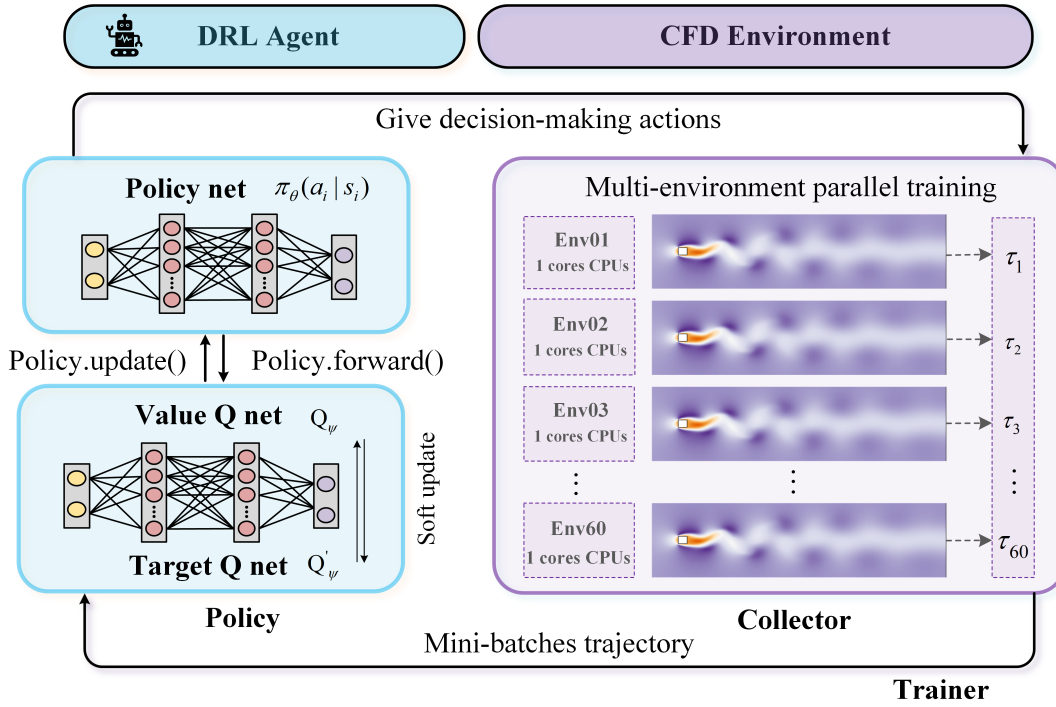


FIG. 4: Describes an agent observing states in the environment, taking actions, and obtaining rewards from the environment to learn and optimize its behavior. Posits a scenario with 60 parallel CFD environments, each harnessing the computational power of 1 CPU, cumulatively utilizing 60 cores. These environments are independent CFD environments, capable of concurrently generating 60 separate trajectories to facilitate gradient updates. Neural networks are used to approximate value and policy functions, which are an important part of this computational process.

C. Formulation of the AFC problem using DRL

The study utilizes DRL algorithms to control two synthetic jets placed on two sides of a square cylinder. By adjusting the airflow around the cylinder through suction or blowing from the jets, the research aims to alter the flow field, affecting the velocity distribution and pressure gradient, in order to minimize drag and suppress vortex shedding. This section outlines how CFD as an environment for DRL algorithms can be coupled with DRL algorithms to solve AFC problems. The integration of CFD into DRL algorithms necessitates an initial comprehension of the fundamental components of DRL algorithms, namely the environment and the agent. Embedding CFD within the DRL algorithm's environment, and utilizing the decision-making capabilities of the agent, aims to achieve the objectives of reducing vortex shedding and drag around a bluff body. This is facilitated through defining key components such as the agent's s_t and a_t , and designing a reward function r_{T_i} to complete the agent's configuration. Ultimately, the agent, by observing the current s_t and selecting corresponding a_t , obtains r_t based on the feedback from the environment. The environment, in turn, transitions to a subsequent s_t based on the agent's a_t and the current s_t , providing corresponding r_t . This iterative process underpins the synergy between CFD and DRL, offering a novel approach to optimizing fluid dynamics problems through learned behaviors. The design of the fundamental framework, as delineated in Fig. 4, draws upon the collective works of Rabault *et al.*³⁶, Rabault *et al.*⁸, Wang *et al.*³⁹, Li and Zhang³⁷, Paris, Beneddine, and Dandois³⁸ and Jia and Xu⁴³.

- s_t : At each temporal juncture, the s_t is constituted by instantaneous flow field or pressure data collected from specific locations within the computational domain of the numerical simulation. Around the bluff body and in the wake region, 201 probes are strategically placed to capture the transient states of the CFD environment. The placement of these probes is informed by the research findings of Wang *et al.*³⁹, Li and Zhang³⁷, Paris, Beneddine, and Dandois³⁸, ensuring comprehensive coverage around the bluff body and within the wake region. This arrangement scrutinizes the placement of probes to ensure they encompass the maximum values and boundary locations of velocity fields, pressure fields, and their fluctuating counterparts, thereby capturing key flow characteristics essential for accurate analysis and decision-making within the DRL framework.
- a_t : the a_t of the agent are defined as the velocity of the synthetic jets, with the stipulation that the magnitude of these action cannot exceed 2% of the inlet velocity. This constraint is designed with the intention of achieving drag reduction using

the smallest possible jet velocity, thus ensuring the energy efficiency of the AFC technology. To guarantee the smoothness of continuous a_t provided by the agent to the environment, a smoothing function is employed between adjacent a_t . Let a_t and a_{t+1} represent the a_t magnitudes at consecutive time steps. The smoothing function S applied between these a_t can be defined by a mathematical relationship that ensures a gradual transition from a_t to a_{t+1} , thereby mitigating abrupt changes in the jet velocity. Smoothing function S can be defined as:

$$S(V_{\Gamma_1, T_i}, a, V_{\Gamma_1, T_{i-1}}) = V_{\Gamma_1, T_i} + \beta \cdot (a - V_{\Gamma_1, T_{i-1}}), \quad (8)$$

where V'_{Γ_1, T_i} is the updated value at time step i , V_{Γ_1, T_i} is the current value at time step i , a represents the target a_t magnitude, $V_{\Gamma_1, T_{i-1}}$ is the value at the previous time step $i - 1$, and β is a coefficient determining the extent of adjustment towards the target a . This function effectively interpolates between the previous value and the target a_t , with β controlling the smoothness of the transition.

- r_t : The design of the reward function is crucial for guiding effective learning, particularly in fluid dynamics applications like reducing drag around a bluff body, where selecting appropriate performance metrics is key. In this study, the reward function incorporates both the C_D and C_L , which are critical indicators of performance in fluid flow past objects.

The specific reward function takes the following form:

$$r_{T_i} = C_{D,0} - (C_D)_{T_i} - \omega |(C_L)_{T_i}|. \quad (9)$$

$C_{D,0}$ represents the baseline C_D , serving as a reference point for drag coefficient. The addition of the reference point is to drive the overall reward to positive values, which helps with the convergence during training. $(C_D)_{T_i}$ denotes the C_D at time step T_i , with the objective of minimizing this value relative to the baseline. $(C_L)_{T_i}$ signifies the C_L at time step T_i , whose absolute value is penalized to mitigate lift forces that may destabilize the flow around the bluff body. ω is a weighting factor that quantifies the trade-off between minimizing drag and controlling lift fluctuations. Since the flow we are interested in is highly directional, the weighting factor is typically between 0.1 and 0.2, indicating that drag reduction and lift reduction are valued differently. In this study, the exact value of hyperparameter ω is tuned for performance by following the advice given by Rabault *et al.*³⁶.

In this study, numerical simulations were conducted with a time step size of 5×10^{-4} seconds, and the duration of the control time step was set to 0.025 seconds, equivalent to 25 numerical simulation time steps. Consequently, the total duration of a training dataset comprising 100 steps amounts to 1.25 seconds, which corresponds to 2500 numerical simulation time steps. In scenarios with Re ranging from 100 to 400, the vortex shedding period is approximately between 0.174 and 0.191 seconds. Thus, one epoch is designed to last for 1.25 seconds, equating to about 6.6 to 7.2 vortex shedding periods. Each training epoch encompasses several vortex shedding periods, allowing the agent to thoroughly observe and adapt to the flow dynamics within multiple shedding periods.

D. Parallelization of DRL-based AFC problem

This parallel strategy draws inspiration from the research conducted by Rabault and Kuhnle⁶² and Wang *et al.*³⁹, while also incorporating insights from the work of Jia and Xu⁴³. For a more comprehensive understanding, we encourage readers to consult their work and the references therein. The DRL training and CFD simulations presented in this study were performed on a high-performance computing system. The computational framework employed for the analyses is powered by an Intel® Xeon® Platinum 8358 CPU, operating at 2.60GHz. This system is fortified with a total of 64 cores, distributed evenly across two sockets, with each socket housing 32 cores. The CFD were conducted utilizing OpenFOAM® version 8⁵⁴, an acclaimed open-source software platform. For the DRL aspect, the open-source Python framework Tianshou⁶³ was employed. Fig. 4 illustrates the allocation and utilization of CPU cores during the coupled training of DRL algorithms and CFD simulations, showcasing the fundamental parallelization strategy.

III. RESULTS AND DISCUSSION

Section III A details the baseline fluid dynamics around a square cylinder for various Re (100, 200, 300 and 400), with a focus on the examination of fluid dynamic parameters such as C_D , C_L , and St . The analysis furnishes a reference for selecting pertinent coefficients for AFC strategies and lays the groundwork for a comparative assessment of control effectiveness. In Section III B, the SAC algorithm is employed to evaluate the control performance and reliability of applying AFC techniques on the flow around a square cylinder at various Re . The primary focus is to ascertain whether the DRL agent can effectively manipulate

vortex shedding in flow configurations at increasing Re to achieve a reduction in drag and control over lift. Section III C discusses detailed comparison of flow fields before and after control application, revealing the mechanisms behind the algorithm's effectiveness in modifying fluid behavior and reducing flow separation.

A. Flow Analysis of a Square Cylinder at Different Reynolds Numbers

The analysis of the flow around a square cylinder at $Re = 100$, $Re = 200$, $Re = 300$ and $Re = 400$ is depicted in Fig. 5, which present the variation of C_D , C_L , and PSD (Power Spectral Density) of C_L over time. Starting with $Re = 100$, Fig.5 (a) elucidates the temporal evolution of the C_D . Initially, there is a marked transient phase where C_D exhibits a gradual decrease, then stabilizing to a quasi-steady state as the flow regime reaches equilibrium. This behavior is indicative of the flow adjustment from an initial transient response to a stable vortex shedding pattern. In Fig.5 (b), following a transition phase, the C_L exhibits periodic oscillations within a magnitude of ± 0.2 . These oscillations are symmetric about the zero baseline, indicating that the lift acting on the cylinder has an alternating character due to the periodic shedding of vortices. Fig.5 (c) is the PSD of the C_L . The non-dimensionalized frequency $fD/U = 0.142$ corresponding to the peak value reflects the main vortex shedding frequency, indicating the periodicity of the lift exerted on the object due to the alternating separation of the vortex. At a Re of 200, the development patterns of drag and lift around a square cylinder are fundamentally similar to those observed at $Re = 100$. However, as depicted in Fig.5 (d), the transition to a stable phase of periodic vortex shedding around the square cylinder occurs more rapidly at $Re = 200$. Correspondingly, as shown in Fig.5 (e), the C_L of the square cylinder also progresses through a transition phase before entering a periodic stage, where the magnitude of C_L oscillations is greater compared to the scenario at $Re = 100$. Similarly, Fig.5 (f) illustrates a PSD with a distinct peak, indicating that the phenomenon of vortex shedding is dominated by a single frequency, and it is consistent to the shedding frequency for $Re = 100$.

As we increase the Re to 300, the C_D around the square cylinder, as illustrated in Fig.5 (g), transitions at a faster rate from an initial transient stage to a phase characterized by periodic oscillations. The periodicity of the C_D oscillations at $Re = 300$ is significantly different from that observed at $Re = 100$ and $Re = 200$. In the stable phase of periodic oscillations at $Re = 300$, the drag curve exhibits multiple distinct extremal points within each period. Correspondingly, in Fig.5 (h), the C_L also displays multiple oscillation amplitudes, oscillating not at a singular frequency but rather at multiple frequencies that are harmonics of the lowest peak frequency. The PSD calculations for the lift, as shown in Fig.5 (i), further corroborate that the periodic oscillations of lift are governed by multiple frequencies. The largest spectral density still peaks close to the frequency at $Re = 100$ or 200. However, the emergence of harmonics indicate the onset of multi-scale in the underlying flow. When the Re reaches 400, both the amplitude of the C_D (Fig.5 (j)) and C_L (Fig.5 (k)) surpass those observed at other Re , accompanied by a more complex fluctuation frequency spectrum. The power spectral density graph reveals even more harmonics at higher frequencies, shown in Fig.5 (l), indicating the predominance of small-scale vortices at higher frequencies in the vortex shedding phenomenon.

Overall, through the analysis of the characteristics of C_D , C_L and f across four different Re , it is observed that at $Re = 100$ and $Re = 200$, the flow around the square cylinder is characterized by a vortex shedding pattern dominated by a single frequency. Conversely, at $Re = 300$ and $Re = 400$, the vortex shedding pattern around the square cylinder becomes more chaotic and multiple harmonic frequencies emerge. This transition indicates an increase in flow complexity as the Re rises, implying that flow control encounters greater challenges at higher Re .

B. Control effect based on SAC algorithm

In the previous section, we observed that at $Re = 300$ and $Re = 400$, wake vortex shedding behind the square cylinder exhibits a complex pattern dominated by multiple frequencies, in stark contrast to the periodic, single-frequency dominated shedding observed at $Re = 100$ and $Re = 200$. Although Wang *et al.*³⁹ has demonstrated that an DRL algorithm can achieved a drag reduction effect of 13.7% at $Re = 100$, the vortex shedding phenomenon still prevails. Whether an effective control of the wake vortex can be achieved at high Re still remain elusive. Therefore, this section aims to verify whether active flow control based on an DRL algorithm can intelligently identify control strategies at a range of Re to suppress the shedding of wake vortices, thus reducing drag and mitigating lift oscillations.

We use the SAC algorithm to implement active flow control, and use a pair of synthetic jets to control the flow field of a square cylinder. Fig. 6 presents the reward function profiles from the DRL training process using the SAC algorithm across $Re = 100$, 200, 300 and 400 respectively. As depicted in Fig. 6(a), at $Re = 100$, the SAC agent's learning trajectory is initially characterized by lower rewards, yet it exhibits rapid improvement within the initial episodes. The learning curve swiftly stabilizes, culminating in a plateau around the 2,500th episode, indicative of a consolidated learning outcome. The compact 2σ interval observed in later stages signifies consistent agent performance, implying a stable and convergent training process at this Re . Based on the definition of the reward function in Eq. (9), a positive reward value indicates a successful reduction in the mean drag, suggesting that the agent have learned from simulation data to achieve the objective at $Re = 100$.

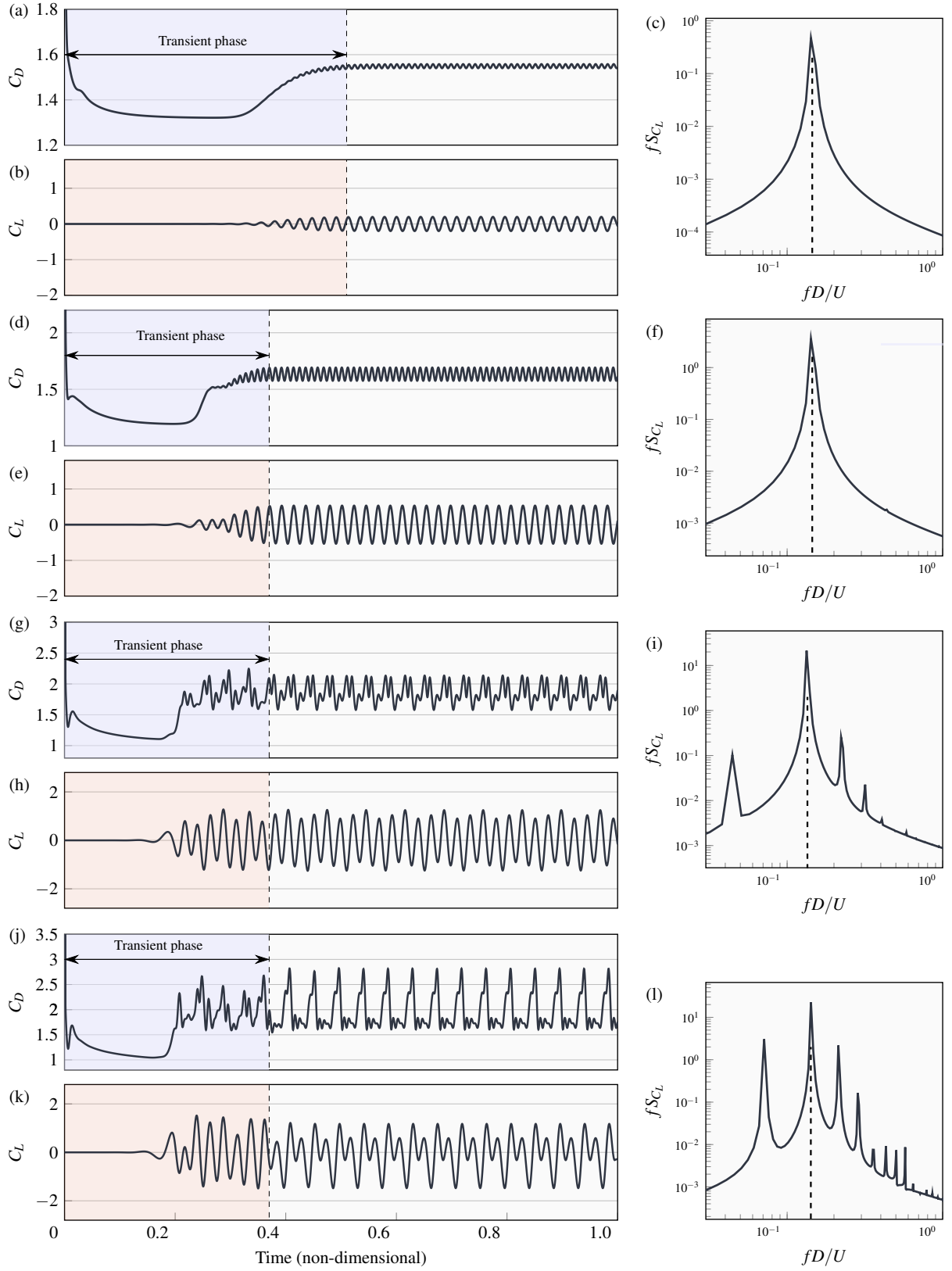


FIG. 5: Unsteady flow past a square cylinder. For $Re = 100$ (a) time series of C_D ; (b) time series of C_L ; (c) PSD of C_L ; For $Re = 200$ (d) time series of C_D ; (e) time series of C_L ; (f) PSD of C_L . For $Re = 300$ (g) time series of C_D ; (h) time series of C_L ; (i) PSD of C_L ; For $Re = 400$ (j) time series of C_D ; (k) time series of C_L ; (l) PSD of C_L .

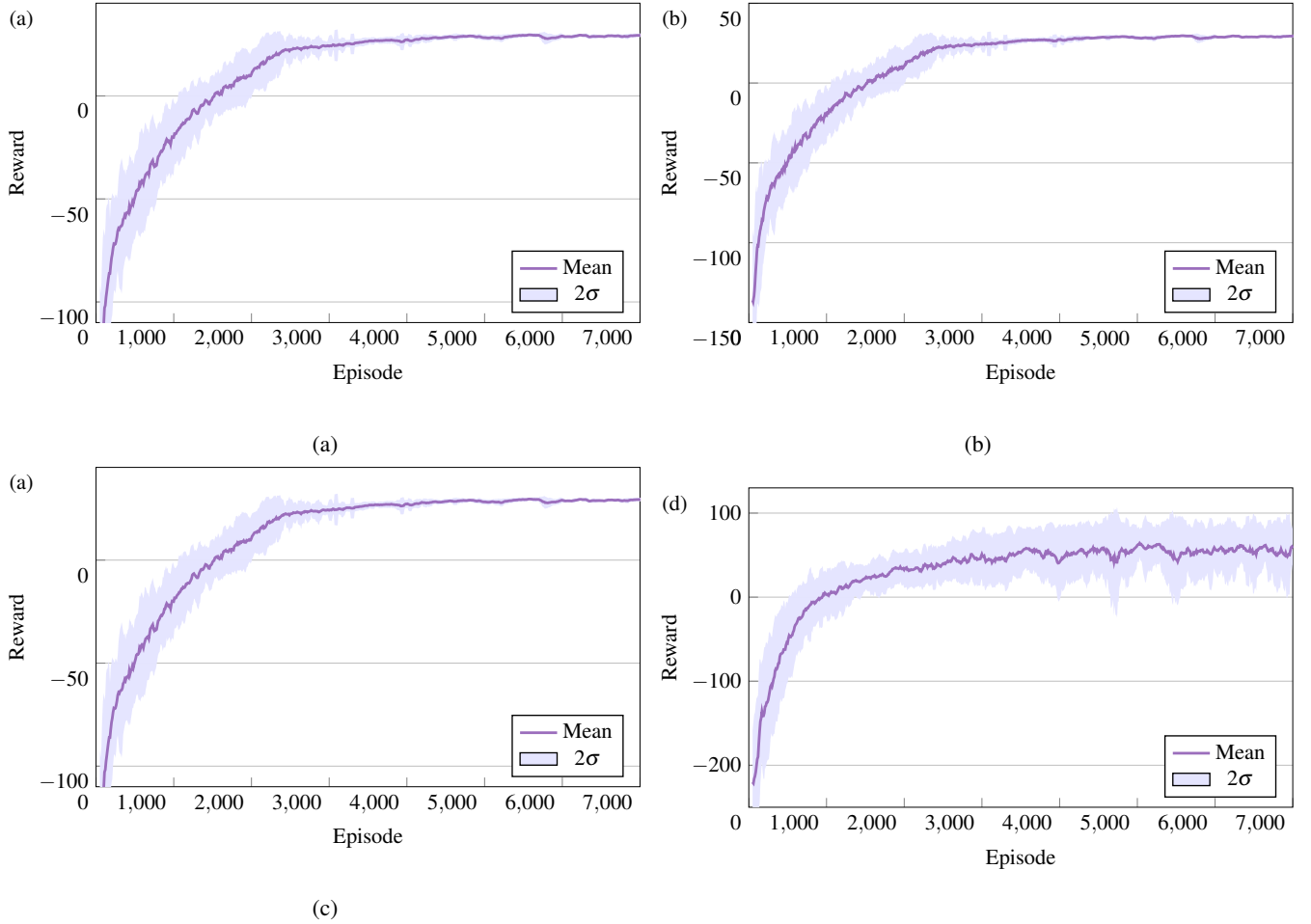


FIG. 6: Cumulative reward for the DRL training process using the SAC algorithm. The solid curves depict the mean reward at each episode, with the shaded area representing a range of two standard deviations (2σ) from the mean. (a) $Re = 100$; (b) $Re = 200$; (c) $Re = 300$; (d) $Re = 400$.

At higher Re , similar learning trajectories are evident in Figs. 6(b) and 6(c) and Section III B, where rewards in each instance converge over time. This consistency underscores the remarkable proficiency of DRL agents in acquiring effective control methods across a spectrum of Re . Beyond this commonality, it is observable that the extent of variation within each reward trajectory differs. With an increase in the Re , the level of fluctuation intensifies. Specifically, at $Re = 400$, the variability remains comparatively large at a notable magnitude even while the average reward stabilizes. This extensive fluctuation in the reward signal at higher Re suggests that more complex flow dynamics present greater challenges, necessitating more sophisticated control strategies.

Fig. 7 depicts the continuous adjustment of jet velocity by the SAC agent during its interaction with the CFD environment, across $Re = 100, 200, 300$, and 400 , respectively. In Fig. 8, the consequent variations in C_D and C_L are presented. The reference baseline is included to enable comparisons between the flow conditions before and after the implementation of control. The initial phase of control is characterized by a pronounced surge in jet velocity, which subsequently undergoes a gradual decline before achieving stabilization near a negligible value. Under the influence of the synthetic jet, both the C_D and C_L fluctuate significantly as well, then sharply decrease and eventually reach a stable equilibrium. Notably, across all conducted experiments, there is a substantial reduction in the C_D , while the C_L approaches a value near zero. Even at higher Re where the baseline flow is more chaotic, the DRL algorithm succeeds in devising effective flow control strategies that mitigate drag and lift fluctuations. The near-zero magnitude of the jet velocity also implies that minimal energy is required to sustain the stabilized flow for long-duration drag reduction.

We summarize the results of active flow control performed at multiple Reynolds numbers in Table. II. The quantities of interest include the reduction ratio of time-averaged drag, time-averaged lift, time-averaged action values, and the suppression of vortex shedding after control. We delay the discussion of vortex shedding suppression to Section III C. We also include published results from existing literature. For a Re of 100, our study achieves approximately 14.4% reduction in average drag, surpassing

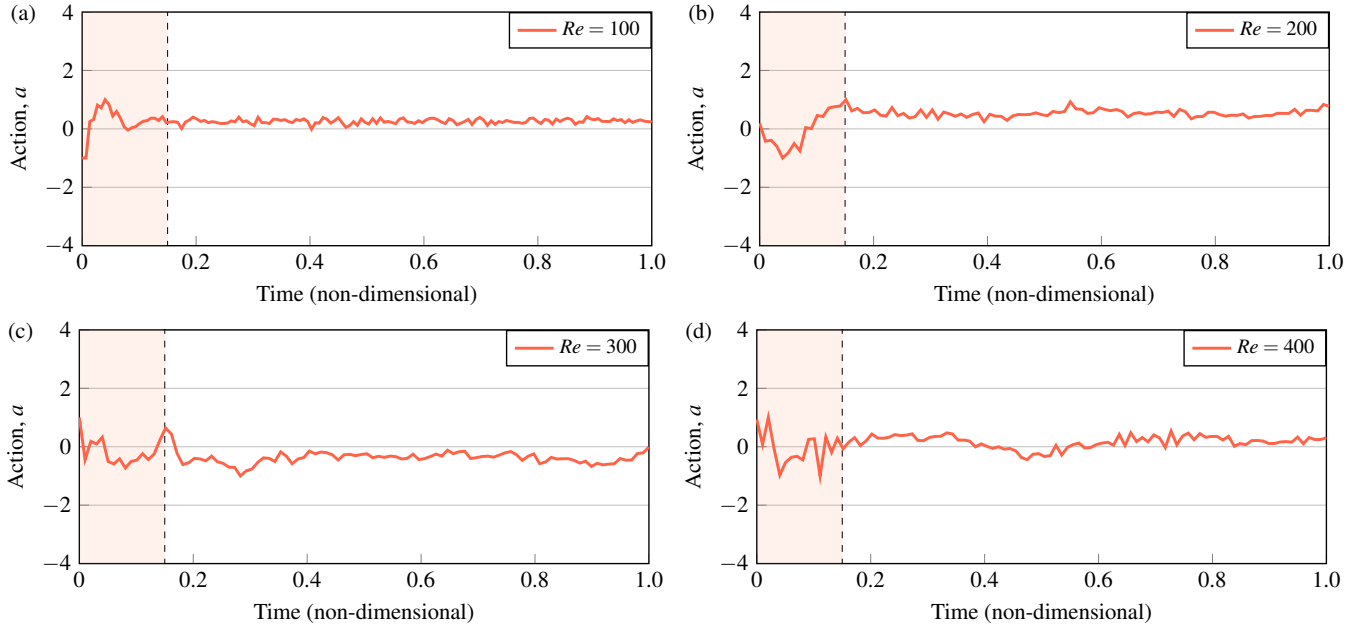


FIG. 7: The value of the action generated during the interaction between the agent and the environment is the value of the mass flow rate of the synthetic jet. (a) $Re = 100$; (b) $Re = 200$; (c) $Re = 300$; (d) $Re = 400$.

TABLE II: Summary of results for our work and related research on AFC applied to square cylinder using DRL.

Re	Jet Location	$\bar{C}_{D,Baseline}$	$\bar{C}_{D,Controlled}$	Drag reduction rate (%)	$\bar{C}_{L,Controlled}$	Action, $\bar{a}$	Vortex shedding suppression	Reference
100	Trailing	1.549	1.337	13.7	-	-	NO	Wang <i>et al.</i> ³⁹
	Leading	1.476	1.370	7.2	-	-	NO	Chen <i>et al.</i> ⁵⁰
	Middle	1.476	1.440	2.4	-	-	NO	Chen <i>et al.</i> ⁵⁰
	Trailing	1.476	1.280	13.3	-	-	YES	Chen <i>et al.</i> ⁵⁰
100	Trailing	1.549	1.325	14.4	0.0013	0.2675	YES	Present study
200	Trailing	1.633	1.203	26.4	-0.0101	0.1180	YES	
300	Trailing	1.854	1.133	38.9	-0.0240	-0.3655	YES	
400	Trailing	1.964	1.041	47.0	-0.0065	0.05905	YES	

a. Jet Location: synthetic jets are located on the lateral or top/bottom sides of the square cylinder. Specifically, 'Trailing' refers to the rear corners; 'middle' refers to the middle position between front and rear sides; and 'leading' refers to the front corners of the square cylinder.

b. $\bar{C}_{D,Baseline}$ represents the time-averaged drag of the square cylinder after the flow field stabilizes without AFC. $\bar{C}_{D,Controlled}$, $\bar{C}_{L,Controlled}$, and $\bar{a}$ denote the time-averaged values under stable conditions after AFC is applied, excluding the initial transient phase.

c. "Vortex shedding suppression" indicates whether synthetic jets control can fully prevent vortex shedding in the wake of the square cylinder. Please refer to Fig. 11 for flow visualizations.

the drag reduction reported by Wang *et al.*³⁹ and Chen *et al.*⁵⁰. It is worth mentioning that Chen *et al.*⁵⁰ found trailing jets to be more effective than leading jets, and they were able to achieve vortex shedding suppression with trailing jet control. At Re values of 200, 300, and 400, the AFC strategy implemented in our study achieves drag reduction rates of 26.4%, 38.9%, and 47.0% respectively, while maintaining the lift coefficient near zero. Furthermore, the action value of the synthetic jet evidences the strategy's compliance with the low-energy consumption requirements inherent in AFC technologies.

In Fig. 9, we compare the PSD in the aerodynamic coefficients before and after flow control. In the baseline scenario without control, the PSD shows prominent peaks at specific frequencies near $fD/U = 0.142$, indicating the periodic nature of the fluid vibrations and the corresponding dominant frequencies. Conversely, upon the implementation of flow control, these prominent peaks are either remarkably attenuated or completely vanish, suggesting that the control strategy has significantly altered the oscillatory behavior of the fluid and effectively suppressed periodic fluctuations. At Re of 100 and 200, the previously significant peaks in the PSD graphs are eliminated after control, implying that the flow control measures have disrupted the periodic fluctuations of the C_L . For Re of 300 and 400, multiple significant frequency peaks are evident in the baseline state without control. However, after the application of flow control, all characteristic frequencies are eliminated as well, indicating that the DRL-based control strategy has a mitigating effect on multiple frequencies at higher Re . This highlights the success and robustness of the SAC agent in identifying relevant frequencies and deploying intelligent control strategies to mitigate drag and

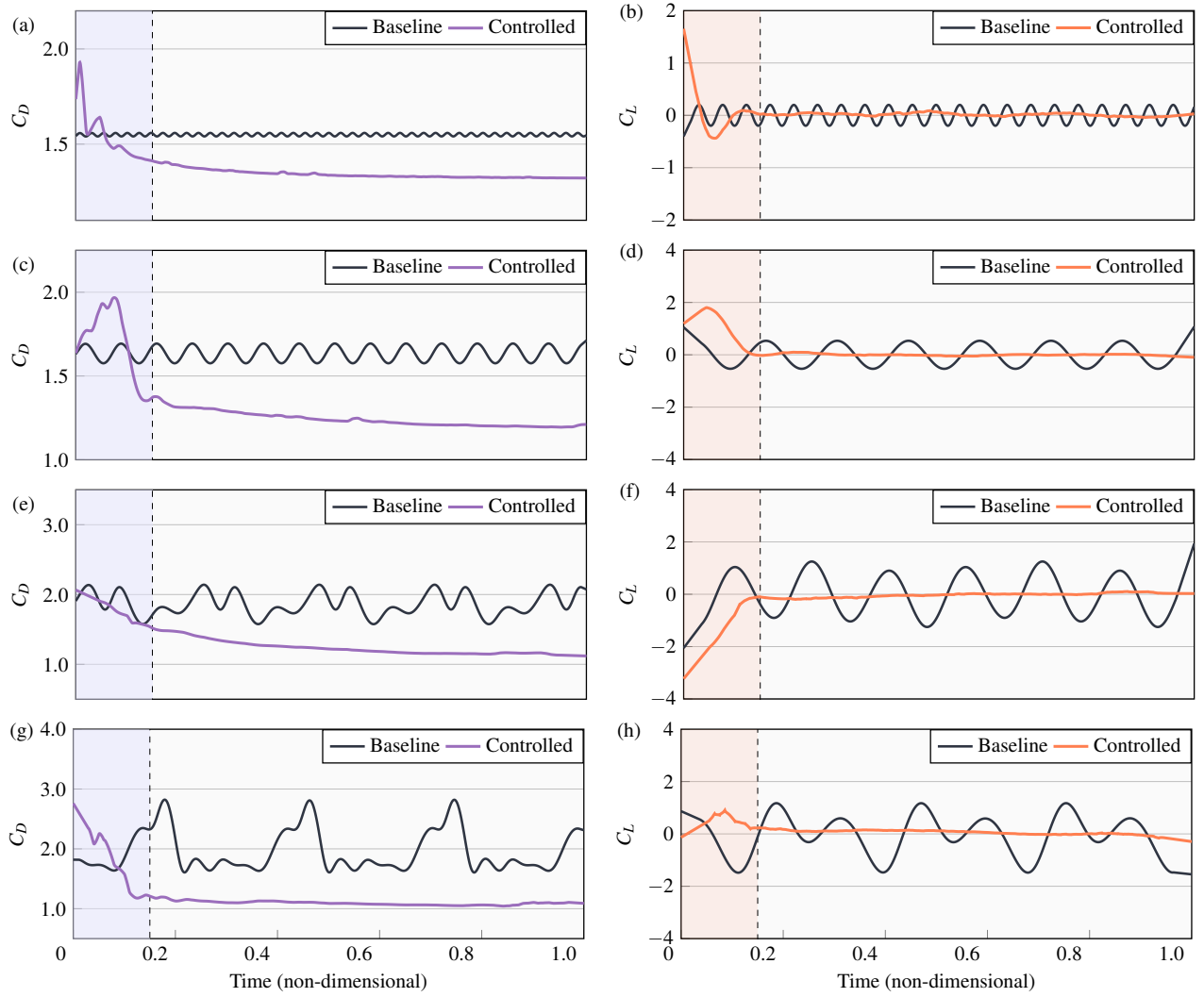


FIG. 8: Comparative diagram of the fluid force coefficients before and after flow control. The left side is C_D and the right side is C_L . (a) and (b) $Re = 100$; (c) and (d) $Re = 200$; (e) and (f) $Re = 300$; (g) and (h) $Re = 400$.

regulate lift fluctuations. The capability to react to multi-scale frequencies provides promising aspects of the adaptability of DRL in chaotic and turbulent AFC scenarios.

Lastly, Fig. 10 illustrates the comparison between the time-averaged pressure distributions on the surface of the square cylinder under both controlled and uncontrolled flow scenarios. The square cylinder possesses a side length denoted by D , with its center located at the origin of the coordinate system. The specific points on the square cylinder are designated as follows: point A at $(-D/2, -D/2)$, point B at $(D/2, -D/2)$, point C at $(D/2, D/2)$, and point D at $(-D/2, D/2)$. In Fig.10 (a), At $Re = 100$, the control strategy deployed by the SAC agent has a negligible effect on the pressure distribution along the AD edge (the windward face) of the square cylinder. Conversely, it substantially affects the pressure distribution on both lateral sides (AB and CD) and the leeward face (BC) of the cylinder, leading to a marked decrease in the average pressure at these sites following the implementation of control measures. Similarly, Fig.10 (b) illustrates that at $Re = 200$, the changes in the average pressure distribution around the square cylinder before and after control are essentially consistent with those observed at $Re = 100$. Notably, at $Re = 200$, the reduction in pressure distribution on both lateral sides and the leeward face of the cylinder is even more pronounced. In Fig.10 (c) and Fig.10 (d), under conditions of $Re = 300$ and $Re = 400$ the flow control strategy continues to exert minimal influence on the average pressure distribution on the windward face of the square cylinder. Similar to observations at $Re = 100$ and $Re = 200$, the impact on the pressure distribution on both lateral sides and the leeward face of the cylinder is more significant. Notably, the influence of flow control on the pressure distribution on the leeward face is greater than that on the lateral sides, with the most pronounced effect occurring near the middle of the leeward face. Overall, this set of comparisons illustrate that the majority of drag reduction is achieved via pressure forces, which highlights the significance of suppressing vortical structures for optimal performance.

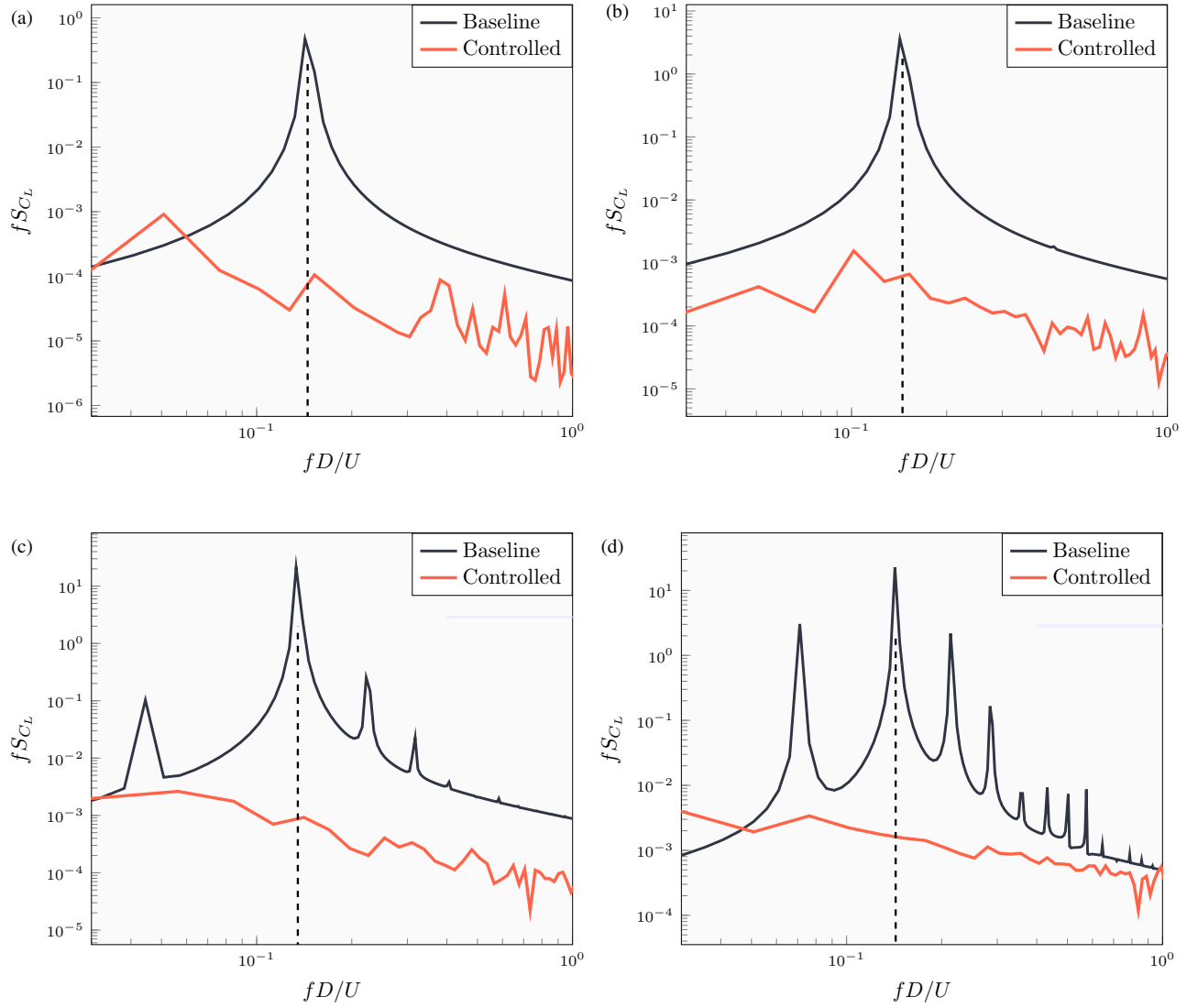


FIG. 9: A comparative diagram of the PSD of C_L before and after flow control. (a) $Re = 100$; (b) $Re = 200$; (c) $Re = 300$; (d) $Re = 400$.

C. Detailed flow field comparison before and after control

As reflected by various statistical measures, the DRL agent is capable of finding optimal control strategies for reducing the overall drag. To elucidate the underlying control mechanism, a meticulous examination of the flow field is imperative. Here, we aim to decode the interaction dynamics between the synthetic jet actuation and the surrounding flow to facilitate an understanding of the DRL algorithm in AFC problems.

Fig. 11 illustrates the velocity contours at three pivotal moments during the flow control process around the square cylinder for four different Re . t_0 marks the commencement of control, representing the velocity field of the uncontrolled flow. t_1 corresponds to an intermediate moment during the control process, and t_2 displays velocity field of the controlled flow. This sequential representation serves to visually capture the evolution of the flow field from its initial state, through the application of control strategies, to the final controlled state, thereby providing insights into the effectiveness and impact of the control measures implemented.

Starting at $Re = 100$, Fig. 11(a) displays the wake flow around the square cylinder without any flow control interventions. The flow regime around the square cylinder is characterized by significant flow separation around the body and a von Kármán vortex street within the wake region. The unsteady vortex shedding is periodic with a single shedding frequency. When jet control is

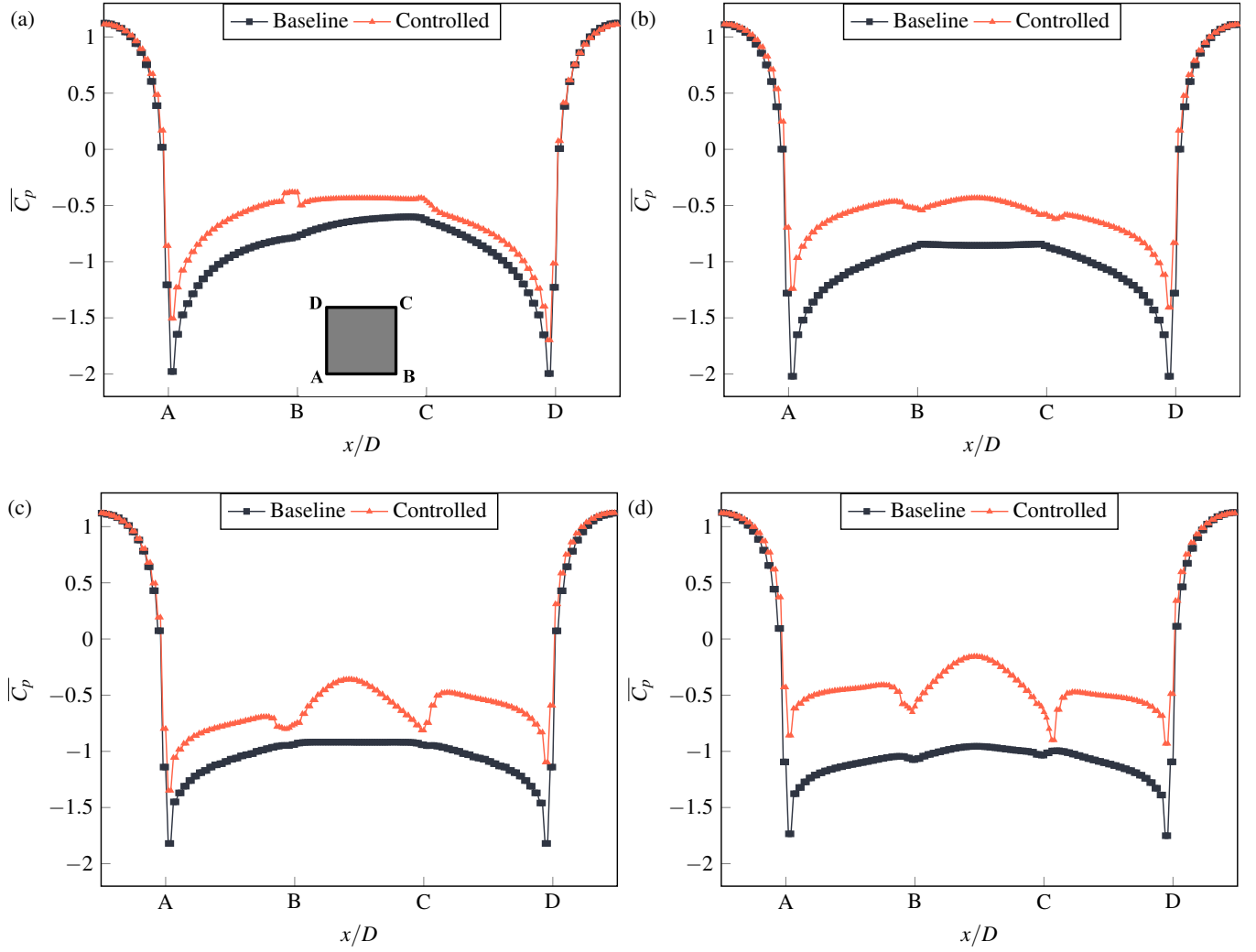


FIG. 10: Illustration of the training process: profiles of the average pressure coefficient. (a) $Re = 100$; (b) $Re = 200$; (c) $Re = 300$; (d) $Re = 400$.

introduced (as seen in Fig. 11(b)), by simultaneous blowing and suction, it substantially elongates the separation zone behind the square cylinder. As a result, the onset of vortex shedding is delayed to a position further downstream than observed in the uncontrolled case. By Fig. 11(c), the jet control has fully influenced the wake flow, and the length of the separation zone has developed to its maximum extent. There is minimal vortex shedding observed in the far downstream of the wake flow, which is indicative of a more stable flow regime.

As we increase the Re , the flow fields in Figs. 11(d), 11(g) and 11(j) demonstrate a transition towards increasingly unstable and complex vortex shedding behavior. The extent of fluctuation amplifies significantly. Specifically, at Re of 300 and 400, the base flow (at t_0) is characterized by chaotic vortex shedding with strong asymmetry. This can be quantified by the multi-frequency PSD analysis in Fig. 5. Nonetheless, the introduction of synthetic jets influences the wake in a similar fashion across all examined Re at t_1 , as shown in Figs. 11(e), 11(h) and 11(k). In each instance, recirculation zones emerge behind the square body, which pushed the vortical structures previously shed downstream. In Figs. 11(f), 11(i) and 11(l) where the recirculation zones are given time to stabilize, the von Kármán vortex street is absent. The flow is characterized by a marked reduction in unsteady behavior, signifying the efficacious stabilization through the deployment of DRL-based flow control strategies.

We would like to highlight that, to the best of our knowledge, the successful suppression of vortex shedding demonstrated in this wide range of Re has not been observed in existing literature. At $Re = 100$, Wang *et al.* and Chen *et al.* showed successful drag reduction through DRL-based AFC control, but did not achieve the desired effect of suppressing vortex shedding in most cases^{39,50}. Notably, in the work by Chen *et al.*, only when the jet actuator was positioned at the rear corners of the square cylinder could the vortices in the wake field of the square cylinder be stabilized, making it the only test case to date capable

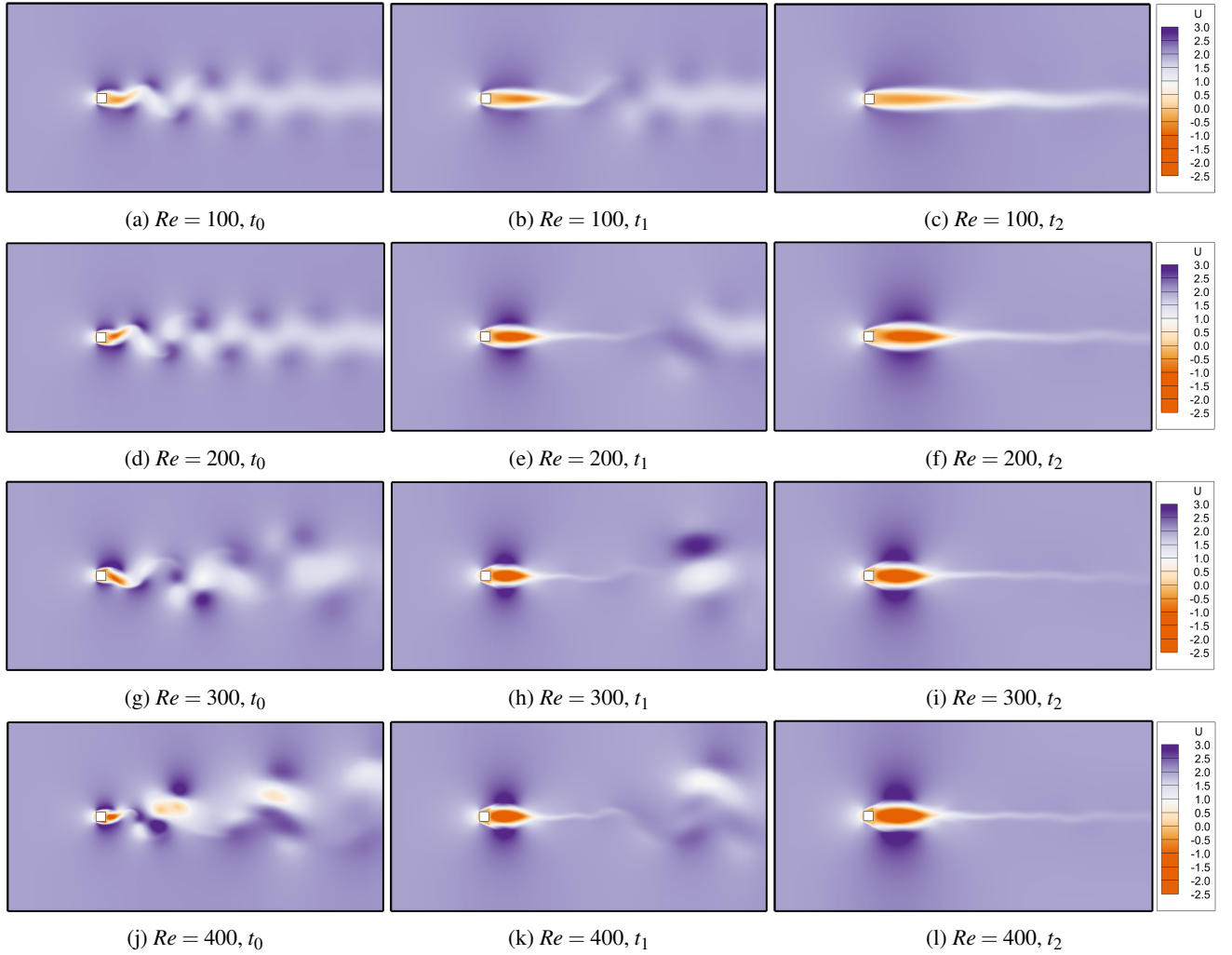


FIG. 11: Instantaneous velocity contours for a range of Reynolds numbers ($Re = 100, 200, 300$, and 400) captured at successive temporal instances. Here, t_0 denotes the initial time, illustrating the uncontrolled velocity field. t_1 corresponds to the controlled state during the implementation of the DRL-based flow control algorithm. t_2 conveys the velocity field following the cessation of control, thereby reflecting the results of the DRL algorithm's application. (a) $Re = 100$, at t_0 . (b) $Re = 100$, at t_1 . (c) $Re = 100$, at t_1 . (d) $Re = 200$, at t_0 . (e) $Re = 200$, at t_1 . (f) $Re = 200$, at t_1 . (g) $Re = 300$, at t_0 . (h) $Re = 300$, at t_1 . (i) $Re = 300$, at t_1 . (j) $Re = 400$, at t_0 . (k) $Re = 400$, at t_1 . (l) $Re = 400$, at t_1 .

of completely controlling vortex shedding in the square cylinder wake. Other related studies include those by Yan *et al.*⁵¹ at $Re = 500, 1000$, and 2000 , and Yan *et al.*⁵² at $Re = 1000$, which demonstrate promising results in drag and lift control around the square cylinder. However, vortex shedding still persists in the wake region under control. In our work, we have showcased the robustness of DRL-based frameworks in devising effective flow control With successful elimination of vortex shedding. A flow around square cylinders with stable wake not only gives many desirable engineering features such as greater drag reduction and minimal lateral vibration, but also paves a way for us to study its control mechanism.

To understand the underlying mechanism of flow control provided by the DRL agent, we investigate the instantaneous streamlines at $Re = 100$ and $Re = 400$ for the uncontrolled baseline and the controlled state. In Fig. 12(a), we observe the formation of a single vortical structure shedding from the square cylinder. A detailed examination in the vicinity of the body shows that, as flow approaches the square body, it separates at the two corners, reattaches along the two lateral sides, and ultimately separates again, which results in the formation of the vortex at the rear of the body. In the absence of control, the vortex shedding process is inherently unsteady and alternates between the two sides, leading to the formation of a vortex street. In the controlled state, Fig. 12(b) shows the emergence of an elongated recirculation zone with a pair of symmetric Föpl vortices⁶⁴ in equilibrium. With the zoom-in view, it is evident that both separation bubbles on the two lateral sides of the square cylinder are still present, but the strategic placement of two jets towards the cylinder's rear significantly alters the flow dynamics. At this particular instant, the suction at Γ_1 and blowing at Γ_2 provides a more symmetrical flow pattern around the two rear corners. This adjustment

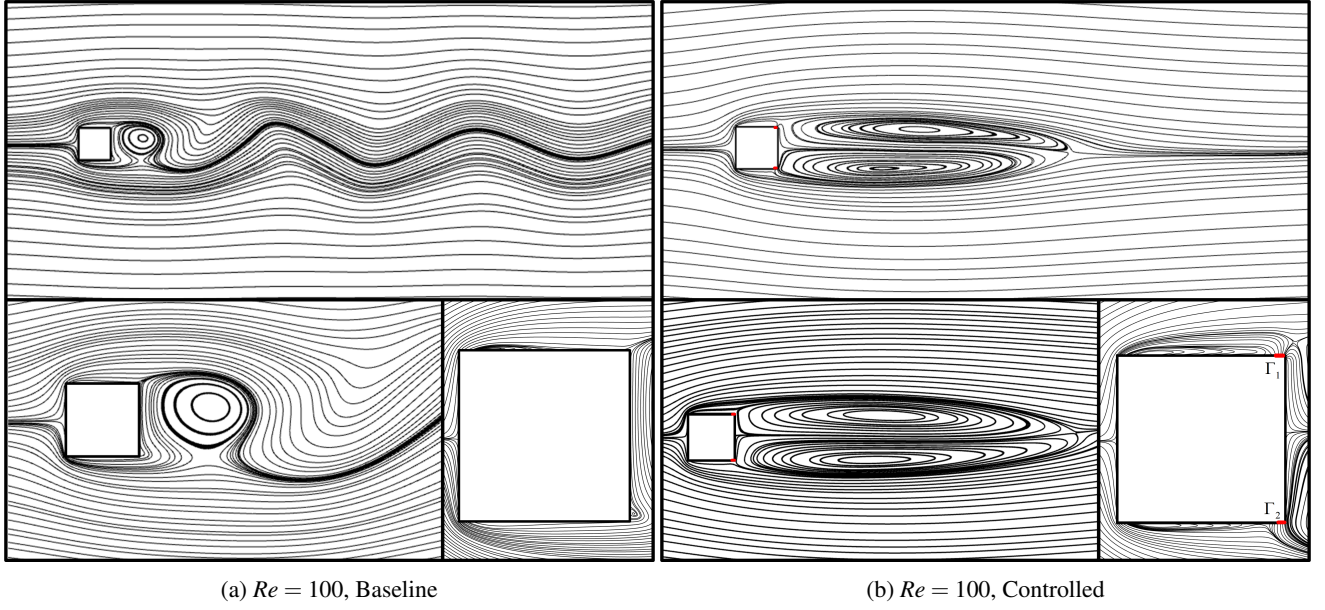


FIG. 12: Streamline comparison of uncontrolled flow (Baseline flow) and controlled flow at $Re = 100$. The upper part of the figure is the wake field of the square cylinder, and the lower part is a detailed view of the flow field around the square cylinder, showing the streamlines downstream of and around the square cylinder.

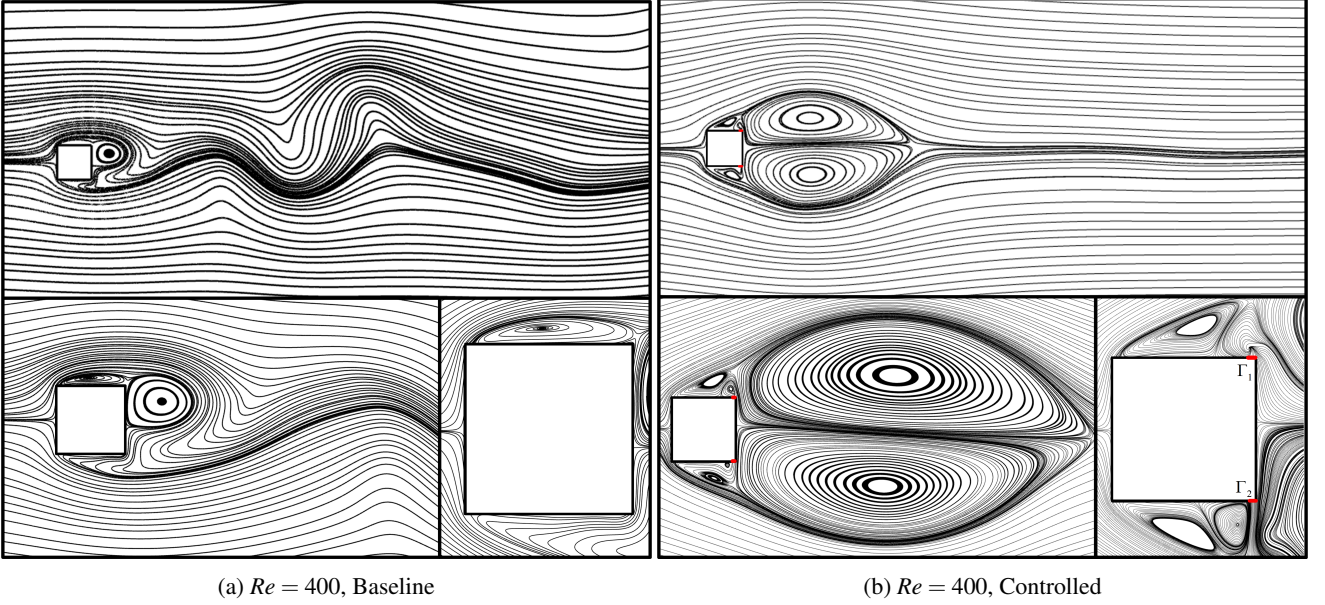


FIG. 13: Streamline comparison of uncontrolled flow (Baseline flow) and controlled flow at $Re = 400$.

fosters the development and stabilization of a recirculation zone, highlighting the DRL agent's capacity to manipulate flow characteristics through precise interventions.

At $Re = 400$, the intervention due to jet actuation is more evident. Fig. 13(a) shows the vortex shedding streamline pattern. A significantly thicker separation bubble can be observed on one side of the square due to the increased Re . The implementation of active flow control yields a recirculation zone as depicted in Fig. 13(b), which, although shorter and thicker compared to that observed in Fig. 12(b), still produces nearly horizontal streamlines in the wake. This indicates that the Föpl vortices have achieved a state of equilibrium. However, a zoom-in view reveals a different control mechanism at play. Instead of directly interrupting the separation bubbles on the sides, the two synthetic jets generate small vortical structures that act as barriers, preempting the upstream propagation of separation bubbles towards the sharp rear corners of the square cylinder. This approach

also facilitates the concurrent development of separation bubbles on both sides of the cylinder. Such symmetry is instrumental in the stable emergence of Föppl vortices directly behind the square cylinder, effectively precluding the formation of vortex shedding.

It is noteworthy that no explicit information pertaining to the control mechanism was provided throughout the training phase of the SAC agent. The only constraint was the limitation on the magnitude of the jet flow for energy efficiency. Nevertheless, the DRL algorithm demonstrated remarkable proficiency in identifying an optimal control strategy, which is tailored to the specific characteristics of the flow. This adaptability underscores the robustness of the DRL framework in addressing AFC problems. Such autonomy is particularly valuable in scenarios where the flow dynamics are complex and highly variable, as it allows the algorithm to adapt its control strategy to optimize performance across a broad spectrum of conditions.

IV. CONCLUSIONS

The current study methodically investigates the application of DRL-based AFC through the actuation of synthetic jets on a confined squared cylinder. We leverage the SAC algorithm to precisely control the mass flow rate of synthetic jets located on both the upper and lower sides of a squared cylinder. By altering the formation and shedding of vortices as the fluid navigates past the squared cylinder, the control strategy informed by DRL successfully mitigates lift and reduces drag.

- Initially, computational analysis was performed on the flow around the confined squared cylinder at $Re = 100, 200, 300,$ and 400 , encompassing both transitional phases and the establishment of stable periodic vortex shedding stages. Under the flow conditions at $Re = 100$ and 200 , the wake behind the squared cylinder evolved through a transitional phase into a state of stable periodic vortex shedding. Spectral analysis results indicated that at these two Re , the vortex shedding frequency associated with the C_L displayed characteristics of mono-frequency. In contrast, the flow configurations at $Re = 300$ and $Re = 400$ resulted in irregular and unstable vortex shedding behind the square cylinder. Spectral analysis revealed that under these flow conditions, the vortex shedding frequency was dominated by multiple frequencies, diverging from the singular frequency shedding observed at lower Re . The dominance of multiple frequencies in vortex shedding around the squared cylinder implies a more complex flow pattern downstream of the squared cylinder, thereby escalating the challenge of implementing flow control through DRL algorithms.
- We utilize the SAC algorithm for active flow control with synthetic jets on a square cylinder. Results show that DRL agents exhibit proficiency in acquiring effective control methods across a range of Re , with higher Re presenting greater challenges and variability in reward trajectories. The SAC-based flow control results show that under four different Re flow configurations, the C_L quickly drops to a level close to zero after the control is initiated and stabilizes at a level close to zero. Simultaneously, the shedding frequency of vortices decreases, eliminating the prominent peaks observed in spectral analysis under baseline flow conditions. The C_D undergo significant oscillations during synthetic jet activation before quickly decreasing and stabilizing near their minimum values. The previously observed dominant single- or multi-frequency modes within the baseline flow are disrupted and eliminated, indicating the effective removal of regular or chaotic vortex shedding phenomena by the synthetic jet. Quantification of average C_D demonstrates a significant reduction at Re of $100, 200, 300,$ and 400 by approximately $14.4\%, 26.4\%, 38.9\%,$ and 47.0% , respectively. With the aim of reducing resistance and suppressing lift, we hope that the mass flow rate of the jet used can remain at a relatively low level to ensure that the control strategy meets energy-saving requirements. The strategy provided by SAC has met our expectations in this regard.
- An in-depth analysis of the flow field pre and post control is conducted to enhance understanding of the agent's control strategy. From the velocity contour plots under both uncontrolled and controlled flow conditions at $Re = 100, 200, 300,$ and 400 , it is observed that in the wake of the uncontrolled flow around the squared cylinder at $Re = 100$ and 200 , there is a regular, alternating vortex shedding phenomenon. However, at $Re = 300$ and 400 , the vortex shedding as the fluid flows past the squared cylinder becomes irregular and asymmetric. After the implementation of AFC, the periodic shedding of vortices in the wake of the square cylinder is effectively suppressed. Importantly, the irregular and unsteady vortex shedding phenomena are also successfully mitigated. The streamline contour maps clearly demonstrate that under the influence of the synthetic jets, the recirculation region downstream of the square cylinder is significantly elongated and expanded. The establishment of Föppl vortices in equilibrium indicate that the resultant flow is stable and steady. A close examination of the flow around the square body reveals different control mechanisms at varying Re . This demonstrates the robustness of the DRL-based AFC technique in devising catered control strategies for distinct flow field to adapt for optimal performance.

In conclusion, across various Re , the deployment of intelligent agents trained via the SAC algorithm has facilitated the learning of a control strategy capable of dynamically adjusting the mass flow rate of synthetic jets. This strategy significantly diminishes lift around the squared cylinder, substantially reduces drag, and suppresses the generation and shedding of vortices within the

complex, multi-frequency dominated wake. These research outcomes underscore the adeptness of DRL in controlling intricate, multi-frequency dominated vortex shedding phenomena, and highlight the robustness of *Re* to AFC techniques surrounding a confined squared cylinder based on DRL.

ACKNOWLEDGMENTS

The authors thank Dr. Jean Rabault (University of Oslo, Oslo, Norway) and Mr. Qiulei Wang (The University of Hong Kong, Hong Kong SAR, China) for making their open-source codes for deep reinforcement learning and numerical simulation available online at <https://github.com/jerabault29/Cylinder2DFlowControlDRLParallel>⁶⁵, and <https://github.com/venturi123/DRLinFluids>⁶⁶.

DECLARATION OF INTERESTS

The authors report no conflict of interest.

AUTHOR ORCIDS

Wang Jia <https://orcid.org/0009-0008-2786-397X>

Hang Xu <https://orcid.org/0000-0003-4176-0738>.

REFERENCES

- ¹R. Bhatt and M. M. Alam, "Vibrations of a square cylinder submerged in a wake," *Journal of Fluid Mechanics* **853**, 301–332 (2018).
- ²H. Bai and M. M. Alam, "Dependence of square cylinder wake on Reynolds number," *Physics of Fluids* **30**, 015102 (2018).
- ³C.-B. C. Dong-Hyeog Yoon, Kyung-Soo Yang, "Flow past a square cylinder with an angle of incidence," *Physics of Fluids* **22**, 043603 (2010).
- ⁴S. Sen, S. Mittal, and G. Biswas, "Flow past a square cylinder at low reynolds numbers," *International Journal for Numerical Methods in Fluids* **67**, 1160–1174 (2011).
- ⁵M. Zhao, L. Cheng, and T. Zhou, "Numerical simulation of vortex-induced vibration of a square cylinder at a low Reynolds number," *Physics of Fluids* **25**, 023603 (2013).
- ⁶M. Rastan, A. Sohankar, C. Doolan, D. Moreau, E. Shirani, and M. M. Alam, "Controlled flow over a finite square cylinder using suction and blowing," *International Journal of Mechanical Sciences* **156**, 410–434 (2019).
- ⁷D. Gao, H. Meng, Y. Huang, G. Chen, and W.-L. Chen, "Active flow control of the dynamic wake behind a square cylinder using combined jets at the front and rear stagnation points," *Physics of Fluids* **33**, 047101 (2021).
- ⁸J. Rabault, F. Ren, W. Zhang, H. Tang, and H. Xu, "Deep reinforcement learning in fluid mechanics: A promising method for both active flow control and shape optimization," *Journal of Hydrodynamics* **32**, 234–246 (2020).
- ⁹F. Ren, H.-b. Hu, and H. Tang, "Active flow control using machine learning: A brief review," *Journal of Hydrodynamics* **32**, 247–253 (2020).
- ¹⁰S. L. Brunton, B. R. Noack, and P. Koumoutsakos, "Machine learning for fluid mechanics," *Annual Review of Fluid Mechanics* **52**, 477–508 (2020).
- ¹¹C. Vignon, J. Rabault, and R. Vinuesa, "Recent advances in applying deep reinforcement learning for flow control: Perspectives and future directions," *Physics of Fluids* **35**, 031301 (2023).
- ¹²F. Xie, C. Zheng, T. Ji, X. Zhang, R. Bi, H. Zhou, and Y. Zheng, "Deep reinforcement learning: A new beacon for intelligent active flow control," *Aerospace Research Communications* **1** (2023), 10.3389/arc.2023.11130.
- ¹³J. Rabault and A. Kuhnle, "Deep reinforcement learning applied to active flow control," in *Data-Driven Fluid Mechanics: Combining First Principles and Machine Learning*, edited by M. A. Mendez, A. Ianiro, B. R. Noack, and S. L. Brunton (Cambridge University Press, 2023) pp. 368–390.
- ¹⁴P. Garnier, J. Viquerat, J. Rabault, A. Larcher, A. Kuhnle, and E. Hachem, "A review on deep reinforcement learning for fluid mechanics," *Computers & Fluids* **225**, 104973 (2021).
- ¹⁵R. Vinuesa, O. Lehmkuhl, A. Lozano-Durán, and J. Rabault, "Flow control in wings and discovery of novel approaches via deep reinforcement learning," *Fluids* **7**, 62 (2022).
- ¹⁶S. Razdan and S. Shah, "Optimization of fluid modeling and flow control processes using machine learning: A brief review," in *Advances in Mechanical Engineering and Material Science*, Lecture Notes in Mechanical Engineering, edited by K. Popat, S. Kanagaraj, P. Sreekanth, and V. Kumar (Springer, Singapore, 2022).
- ¹⁷J. Viquerat, P. Meliga, A. Larcher, and E. Hachem, "A review on deep reinforcement learning for fluid mechanics: An update," *Physics of Fluids* **34**, 111301 (2022).
- ¹⁸S. L. Brunton and B. R. Noack, "Closed-Loop Turbulence Control: Progress and Challenges," *Applied Mechanics Reviews* **67**, 050801 (2015).
- ¹⁹M. Batta, "Machine learning algorithms-a review," *International Journal of Science and Research (IJSR)* **9**, 381–386 (2020).
- ²⁰I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science* **2** (2021), 10.1007/s42979-021-00592-x.
- ²¹J. Burrell, "How the machine 'thinks': Understanding opacity in machine learning algorithms," *Big Data & Society* **3** (2016), 10.1177/2053951715622512.
- ²²V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature* **518**, 529–533 (2015).
- ²³D. Silver, J. Schrittwieser, K. Simonyan, *et al.*, "Mastering the game of go without human knowledge," *Nature* **550**, 354–359 (2017).
- ²⁴L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research* **4**, 237–285 (1996).

- ²⁵K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, “Deep reinforcement learning: A brief survey,” *IEEE Signal Processing Magazine* **34**, 26–38 (2017).
- ²⁶L. Guastoni, J. Rabault, P. Schlatter, R. Vinuesa, and H. Azizpour, “Discovering drag reduction strategies in wall-bounded turbulent flows using deep reinforcement learning,” in *ICLR 2023 Workshop on Physics for Machine Learning* (2023).
- ²⁷L. Guastoni, J. Rabault, P. Schlatter, H. Azizpour, and R. Vinuesa, “Deep reinforcement learning for turbulent drag reduction in channel flows,” *The European Physical Journal E* **46**, 27 (2023).
- ²⁸Y.-Z. Wang, X.-J. He, Y. Hua, Z.-H. Chen, W.-T. Wu, and Z.-F. Zhou, “Closed-loop forced heat convection control using deep reinforcement learning,” *International Journal of Heat and Mass Transfer* **202**, 123655 (2023).
- ²⁹E. Hachem, H. Ghraieb, J. Viquerat, A. Larcher, and P. Meliga, “Deep reinforcement learning for the control of conjugate heat transfer,” *Journal of Computational Physics* **436**, 110317 (2021).
- ³⁰Y. Cheng, Y. Huang, B. Pang, and W. Zhang, “Thermalnet: A deep reinforcement learning-based combustion optimization system for coal-fired boiler,” *Engineering Applications of Artificial Intelligence* **74**, 303–311 (2018).
- ³¹X. Zhan, H. Xu, Y. Zhang, X. Zhu, H. Yin, and Y. Zheng, “Deepthermal: Combustion optimization for thermal power generating units using offline reinforcement learning,” *Proceedings of the AAAI Conference on Artificial Intelligence* **36**, 4680–4688 (2022).
- ³²E. Amico, G. Cafiero, and G. Iuso, “Deep reinforcement learning for active control of a three-dimensional bluff body wake,” *Physics of Fluids* **34**, 105126 (2022).
- ³³C. Lee, J. Kim, D. Babcock, and R. Goodman, “Application of neural networks to turbulence control for drag reduction,” *Physics of Fluids* **9**, 1740–1747 (1997).
- ³⁴T. Lee, J. Kim, and C. Lee, “Turbulence control for drag reduction through deep reinforcement learning,” *Physical Review Fluids* **8**, 024604 (2023).
- ³⁵L. B. Gerben Beintema, Alessandro Corbetta and F. Toschi, “Controlling rayleigh-bénard convection via reinforcement learning,” *Journal of Turbulence* **21**, 585–605 (2020).
- ³⁶J. Rabault, M. Kuchta, A. Jensen, U. Réglade, and N. Cerardi, “Artificial neural networks trained through deep reinforcement learning discover control strategies for active flow control,” *Journal of Fluid Mechanics* **865**, 281–302 (2019).
- ³⁷J. Li and M. Zhang, “Reinforcement-learning-based control of confined cylinder wakes with stability analyses,” *Journal of Fluid Mechanics* **932**, A44 (2022).
- ³⁸R. Paris, S. Beneddine, and J. Dandois, “Robust flow control and optimal sensor placement using deep reinforcement learning,” *Journal of Fluid Mechanics* **913**, A25 (2021).
- ³⁹Q. Wang, L. Yan, G. Hu, C. Li, Y. Xiao, H. Xiong, J. Rabault, and B. R. Noack, “DRLinFluids: An open-source Python platform of coupling deep reinforcement learning and OpenFOAM,” *Physics of Fluids* **34**, 081801 (2022).
- ⁴⁰H. Tang, J. Rabault, A. Kuhnle, Y. Wang, and T. Wang, “Robust active flow control over a range of Reynolds numbers using an artificial neural network trained through deep reinforcement learning,” *Physics of Fluids* **32**, 053605 (2020).
- ⁴¹N. Heess, D. TB, S. Sriram, J. Lemmon, J. Merel, G. Wayne, Y. Tassa, T. Erez, Z. Wang, S. M. A. Eslami, M. Riedmiller, and D. Silver, “Emergence of locomotion behaviours in rich environments,” (2017), arXiv:1707.02286 [cs.AI].
- ⁴²F. Ren, J. Rabault, and H. Tang, “Applying deep reinforcement learning to active flow control in weakly turbulent conditions,” *Physics of Fluids* **33**, 037121 (2021).
- ⁴³W. Jia and H. Xu, “Optimal parallelization strategies for active flow control in deep reinforcement learning-based computational fluid dynamics,” (2024), arXiv:2402.11515 [cs.LG].
- ⁴⁴A. Sohankar, C. Norberg, and L. Davidson, “Low-reynolds-number flow around a square cylinder at incidence: study of blockage, onset of vortex shedding and outlet boundary condition,” *International Journal for Numerical Methods in Fluids* **26**, 39–56 (1998).
- ⁴⁵A. K. Saha, “Effect of transitions on flow past a square cylinder at low reynolds number,” *Journal of Engineering Mechanics* **135**, 839–851 (2009).
- ⁴⁶S. Sen, S. Mittal, and G. Biswas, “Steady separated flow past a circular cylinder at low reynolds numbers,” *Journal of Fluid Mechanics* **620**, 89–119 (2009).
- ⁴⁷D. J. Tritton, “Experiments on the flow past a circular cylinder at low reynolds numbers,” *Journal of Fluid Mechanics* **6**, 547–567 (1959).
- ⁴⁸A. Saha, G. Biswas, and K. Muralidhar, “Three-dimensional study of flow past a square cylinder at low reynolds numbers,” *International Journal of Heat and Fluid Flow* **24**, 54–66 (2003).
- ⁴⁹J. Park, K. Kwon, and H. Choi, “Numerical solutions of flow past a circular cylinder at reynolds numbers up to 160,” *Ksme International Journal* **12**, 1200–1205 (1998).
- ⁵⁰W. Chen, Q. Wang, L. Yan, G. Hu, and B. R. Noack, “Deep reinforcement learning-based active flow control of vortex-induced vibration of a square cylinder,” *Physics of Fluids* **35**, 053610 (2023).
- ⁵¹L. Yan, Y. Li, G. Hu, W.-I. Chen, W. Zhong, and B. R. Noack, “Stabilizing the square cylinder wake using deep reinforcement learning for different jet locations,” *Physics of Fluids* **35**, 115104 (2023).
- ⁵²L. Yan, Y. Li, B. Liu, and G. Hu, “Aerodynamic force reduction of rectangular cylinder using deep reinforcement learning-controlled multiple jets,” *Physics of Fluids* **36**, 025169 (2024).
- ⁵³H. Jasak, “OpenFOAM: Open source CFD in research and industry,” *International Journal of Naval Architecture and Ocean Engineering* **1**, 89–94 (2009).
- ⁵⁴H. Jasak, A. Jemcov, Z. Tukovic, *et al.*, “Openfoam: A c++ library for complex physics simulations,” in *International Workshop on Coupled Methods in Numerical Dynamics*, Vol. 1000 (2007) pp. 1–20.
- ⁵⁵T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, “Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor,” in *Proceedings of the 35th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 80, edited by J. Dy and A. Krause (PMLR, 2018) pp. 1861–1870.
- ⁵⁶T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel, and S. Levine, “Soft actor-critic algorithms and applications,” (2019), arXiv:1812.05905 [cs.LG].
- ⁵⁷M. L. Puterman, “Markov decision processes,” *Handbooks in operations research and management science* **2**, 331–434 (1990).
- ⁵⁸M. van Otterlo and M. Wiering, “Reinforcement learning and markov decision processes,” in *Reinforcement learning: State-of-the-art*, Vol. 12 (Springer, 2012) pp. 3–42.
- ⁵⁹Y. Mao, S. Zhong, and H. Yin, “Active flow control using deep reinforcement learning with time delays in markov decision process and autoregressive policy,” *Physics of Fluids* **34**, 053602 (2022).
- ⁶⁰B. D. Ziebart, *Modeling Purposeful Adaptive Behavior with the Principle of Maximum Causal Entropy*, Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, PA (2010).
- ⁶¹T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel, and S. Levine, “Soft actor-critic algorithms and applications,” (2019), arXiv:1812.05905 [cs.LG].
- ⁶²J. Rabault and A. Kuhnle, “Accelerating deep reinforcement learning strategies of flow control through a multi-environment approach,” *Physics of Fluids* **31**, 094105 (2019).

- ⁶³J. Weng, H. Chen, D. Yan, K. You, A. Duburcq, M. Zhang, Y. Su, H. Su, and J. Zhu, “Tianshou: A highly modularized deep reinforcement learning library,” *Journal of Machine Learning Research* **23**, 1–6 (2022).
- ⁶⁴L. Föppl, “Far movement behind a circular cylinder,” *Meeting Reports*: 1913,1 (1913).
- ⁶⁵J. Rabault and A. Kuhnle, “Cylinder2dflowcontroldrparallel,” GitHub. <https://github.com/jerabaul29/Cylinder2DFlowControlDRLParallel> (2019).
- ⁶⁶W. C. Qiulei Wang, Lei Yan, “Drinfluids,” GitHub. <https://github.com/venturi123/DRLinFluids> (2022).