JM PAPERS

Deep reinforcement learning finds a new strategy for vortex-induced vibration control

Feng Ren[1,2,†], **Chenglei Wang**[3], **Jian Song**[1] **and Hui Tang**[3,4,†]

[1]School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, PR China

[2]Innovation Center of NPU in Chongqing, Northwestern Polytechnical University, Chongqing 401100, PR China

[3]Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong, PR China

[4]The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen 518057, PR China

As a promising machine learning method for active flow control (AFC), deep reinforcement learning (DRL) has been successfully applied in various scenarios, such as the drag reduction for stationary cylinders under both laminar and weakly turbulent conditions. However, current applications of DRL in AFC still suffer from drawbacks including excessive sensor usage, unclear search paths and insufficient robustness tests. In this study, we aim to tackle these issues by applying DRL-guided self-rotation to suppress the vortex-induced vibration (VIV) of a circular cylinder under the lock-in condition. With a state space consisting only of the acceleration, velocity and displacement of the cylinder, the DRL agent learns an effective control strategy that successfully suppresses the VIV amplitude by 99.6 %. Through systematic comparisons between different combinations of sensory-motor cues as well as sensitivity analysis, we identify three distinct stages of the search path related to the flow physics, in which the DRL agent adjusts the amplitude, frequency and phase lag of the actions. Under the deterministic control, only a little forcing is required to maintain the control performance, and the VIV frequency is only slightly affected, showing that the present control strategy is distinct from those utilizing the lock-on effect. Through dynamic mode decomposition analysis, we observe that the growth rates of the dominant modes in the controlled case all become negative, indicating that DRL remarkably enhances the system stability. Furthermore, tests involving various Reynolds numbers and upstream perturbations confirm that the learned control strategy is robust. Finally, the present study shows that DRL is capable of controlling VIV with a very small number of sensors, making it effective, efficient, interpretable and robust. We anticipate that DRL could provide a general framework for AFC and a deeper understanding of the underlying physics.

† Email addresses for correspondence: renfeng@nwpu.edu.cn, h.tang@polyu.edu.hk

---

1. Introduction

Active flow control (AFC) is an important research area in the field of fluid mechanics, which involves a fluid system being purposely altered by actuators through the exertion of a certain amount of energy input (Collis *et al.* 2004; Choi, Jeon & Kim 2008). Compared with passive control methods that usually involve geometrical modifications, AFC is adaptive and hence can realize more effective control over a much wider operation range. Depending on whether selected signals from the system output are fed back to regulate the actuator(s), AFC can be either open loop or closed loop. Closed-loop control can adjust actuation(s) using feedback signals from sensors, and therefore can automatically operate over a much wider range. The core task of closed-loop control is to design the controller. However, in flow-related systems, the fluid motion is governed by the nonlinear Navier–Stokes equations, and the systems usually involve multiscale and multimodal features, it is thus challenging to judge how the system evolve with certain types of control input. Therefore, the use of classical control theory to design the control law is often unfeasible in AFC. To determine a control law, researchers have developed model-based methods that usually involve certain simplifications, such as linear models (Kim & Bewley 2007), stochastic models (Brackston *et al.* 2016) and reduced-order models (Rowley & Dawson 2017), which have been successfully employed in AFC problems.

Take a classical control problem, vortex-induced vibration (VIV) control, as an example. As one of the most typical problems in both scientific study and engineering applications, VIV occurs when a flow passes a blunt body and asymmetrical vortex shedding appears (Williamson & Govardhan 2004). When the vortex shedding frequency matches the natural frequency of the mass–spring system, more challenging lock-in phenomena occur (Zhang *et al.* 2015), which can be induced by two different mechanisms – resonance and flutter. Even under a subcritical Reynolds number ($Re < 47$), when the flow past a stationary cylinder is still steady and stable, VIV can occur (Mittal & Singh 2005). In addition to understanding the complex underlying physics, suppression of these unwanted phenomena is also of great interest, and AFC is a good solution to this (Hong & Shah 2018).

Until now, most AFC approaches for suppressing VIV have been open loop. For example, Wang *et al.* (2016*a*, 2017*a*,*b*) applied open-loop control strategies to explore lock-on phenomena in VIV control, with a pair of synthetic jets being used to interfere with the vortex shedding process. Utilizing the thermal effect, Wan & Patnaik (2016) studied the effect of body heating on the VIV of a circular cylinder with various mass ratios and structural stiffnesses. Muddada & Patnaik (2017) used moving-surface boundary-layer control, in which two small cylinders are deployed near the separation point of the main cylinder and their rotational velocities are adjusted to enforce a certain momentum injection. It is also noteworthy that as a simple yet powerful tool for VIV control, rotary control has been utilized in both numerical and experimental environments. For example, Bourguet & Lo Jacono (2014) studied the effect of rotation rate on the VIV of a circular cylinder at $Re = 100$, and found that symmetry breaking due to rotary control influenced the higher harmonic fluid force components as well as the phase difference between the fluid force and structural vibration. Du & Sun (2015) used rotary control to suppress the VIV of a circular cylinder, and their results showed that the VIV can be suppressed in the 'lock-on' region, where the vortex shedding frequency is locked to the forcing

frequency. With a proper forcing frequency and rotational velocity, the VIV amplitude can be suppressed to less than 1 % of the cylinder diameter. Wong *et al.* (2018) investigated the flow-induced vibration of a sinusoidally rotating cylinder at a low mass ratio and observed that the vibration frequency may be locked into the forcing rotary oscillation frequency or its one-third subharmonic. The above open-loop control attempts provide basic knowledge about how to mediate the fluid–structure interaction (FSI) process, especially the corresponding physical mechanism (Du & Sun 2015). This can be used either to make the vortex shedding frequency deviate from the resonance frequency or to suppress the strength of vortex shedding. In order to achieve the best performance, parameter optimization is usually utilized (Ghraieb *et al.* 2021), which also requires huge cost of computational resources. From the perspective of control theory, open-loop control suffers from varying or perturbed systems, and thus can usually work efficiently only under limited parametric conditions.

To better facilitate VIV control, researchers have also adopted closed-loop control strategies. However, in contrast with the extensive open-loop AFCs, closed-loop controls are quite rare, and mainly apply the classical proportional integral–differential (PID) control scheme or its variants. For example, Zhang, Cheng & Zhou (2004) used PID control for VIV suppression under the resonance condition, with a combination of a hot wire sensor and a structural oscillation sensor providing feedback signals to achieve the best VIV control performance, i.e. a VIV amplitude suppression of 82 % and drag reduction of 35 %. Wang *et al.* (2016*b*) used PID control for VIV suppression, with the flow being mediated by two pairs of windward-suction–leeward-blowing actuators. The results suggest that although the selected control strategies can reach the VIV suppression target, the control performance depends significantly on the choice of control parameters. Vicente-Ludlam, Barrero-Gil & Velazquez (2018) investigated closed-loop VIV control in an experimental environment, and a rotation law proportional to the cylinder's displacement was found to enhance the vibration while another rotation law proportional to the cylinder's transverse velocity could reduce the vibration amplitude. This work suggested that different sensors can play distinct roles in VIV control. Based on classical control theory, some improved closed-loop control schemes have also been successfully proposed, which usually introduce fuzzy, adaptive methods or the sliding mode scheme, e.g. fuzzy PID (Raibaudo *et al.* 2020) or the fuzzy sliding mode (Hasheminejad *et al.* 2014; Lou *et al.* 2021). Furthermore, based on the linear quadratic regulator, a proportional control scheme with a time delay that embeds more flow physics has been proposed by Yao & Jaiman (2017) for low Reynolds number scenarios. In that study, researchers developed a reduced-order model for controlled wake flows with VIV using the eigensystem realization algorithm.

In addition to the above closed-loop control methods, which have mainly been built upon classical control theory, researchers have been exploring machine-learning-based methods to find more efficient control strategies in recent years. For instance, Ren, Wang & Tang (2019) applied genetic programming (GP) to adjust a pair of blowing/suction jets to control VIV and successfully suppressed the VIV amplitude by 94.2 %. Furthermore, GP-based control is also superior in terms of energy costs and robustness under larger Reynolds numbers. Recently, Zheng *et al.* (2021) applied another machine learning approach, deep reinforcement learning (DRL), to VIV control and reduced the VIV amplitude by 82.7 % in a setting where 152 velocity sensors were deployed both around the cylinder and in the wake to provide feedback signals and moved synchronously with the vibrating cylinder. Although this was a good attempt, the final VIV control performance needs to be greatly

improved. Moreover, the control set-up is quite complicated, especially for the sensor array, limiting its feasibility in realistic situations.

As one of the most promising methods for AFC and as a hot research area in the booming artificial intelligence (AI) field, DRL has an advantageous environment-interaction feature, which has made it successful in many very challenging tasks such as playing Go games (Silver *et al.* 2016, 2017) and adversarial computer games (Mnih *et al.* 2019), learning the gliding of birds (Reddy *et al.* 2016) and the swimming of fish (Verma, Novati & Koumoutsakos 2018). The feasibility of DRL in some AFC problems has also been verified (Brunton, Noack & Koumoutsakos 2020; Rabault *et al.* 2020; Ren, Hu & Tang 2020). For example, Rabault *et al.* (2019) achieved drag reduction for a circular cylinder in the laminar flow regime. Paris, Beneddine & Dandois (2021) also studied the typical drag reduction problem and explored how to reduce the state dimensions of the DRL system. Li & Zhang (2022) performed a global stability analysis and sensitivity analysis of the above flow problem. In our prior study (Ren, Rabault & Tang 2021*a*), DRL was applied to the same problem in a more challenging situation where the flow exhibits weak turbulence, confirming that DRL can find effective strategies in chaotic flow systems. A more recent study by Sonoda *et al.* (2023) further extends the DRL-guided AFC to the fully turbulent regime, which is significant progress towards controlling turbulence. In a similar flow system and using an almost identical sensor layout to those in Rabault *et al.* (2019) and Ren *et al.* (2021*a*), Mei *et al.* (2021) applied DRL to enhance VIV for energy harvesting. Fan *et al.* (2020) performed DRL-guided AFC in an experimental environment targeting reducing the drag of a circular cylinder. In our more recent work (Ren, Wang & Tang 2021*b*) focusing on hydrodynamic stealth, the learned strategy in a stationary cylinder scenario was transferred to the VIV scenario, confirming that DRL is a good option for VIV control. Castellanos *et al.* (2022) also performed an analysis of the exploration path of DRL-guided control and made comparisons with GP-based AFC based on cluster-based network modelling that was developed by Fernex, Noack & Semaan (2021) and Li *et al.* (2021), which can project individual controllers onto a low-dimensional orthogonal domain and group them in clusters. All these works are important contributions towards fully qualifying DRL as an effective control algorithm for applications related to fluid mechanics. However, as far as we can see, some general issues associated with DRL-guided AFC still need to be addressed, including excessive sensor usage, unclear search paths and insufficient robustness tests.

Hence, in this study, we explore the effectiveness, efficiency, interpretability and robustness of DRL-guided AFC in suppressing the VIV of a circular cylinder. We highlight the use of simple yet efficient sensors so as to facilitate realistic applications, as well as the interpretability of the search paths of the DRL agent with different types of feedback signals so as to reveal how the AI agent 'thinks' during the random exploration process. In addition, we confirm the control robustness in terms of the Reynolds number effect and perturbations from the upstream flow. The DRL control strategy and physical insights drawn from this study can shed some light on potential DRL-guided control schemes for more complicated AFC problems.

2. Problem description and methodology

2.1. *Problem description*

Vortex-induced vibration occurs when asymmetrical vortex pairs are shed from a bluff body immersed in a fluid flow. Figure 1 shows a schematic of this FSI problem, where the velocity of the uniform incoming flow is U_0. A cylinder of diameter D_0 is connected
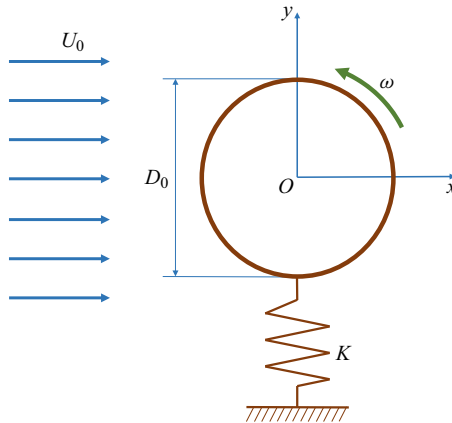
Figure 1. A schematic diagram of the FSI system. The cylinder of diameter D_0 is connected to a spring with stiffness K. The upstream flow has a uniform velocity U_0. The system damping is neglected.

to a spring along the transverse direction while being fixed in the streamwise direction. The Reynolds number is defined as $Re = U_0 D_0 / \nu$ and is fixed as 100, unless otherwise mentioned, where ν is the kinematic viscosity of the fluid.

Ignoring the damping effect of the FSI system, we can express the transverse motion of the cylinder as

$$m\ddot{y} = -Ky + F_L, \tag{2.1}$$

where m is the mass of the cylinder, $\ddot{y}$ is the acceleration, y is the displacement, K is the stiffness of the spring and F_L is the transverse force exerted on the cylinder by the surrounding fluid. In addition to Re, the FSI system is governed by two dimensionless parameters, i.e. the mass ratio m^* and the reduced velocity U_R:

$$m^* = \frac{m}{\rho_0 D_0^2}, \quad U_R = \frac{U_0}{f_N D_0}, \tag{2.2a,b}$$

where ρ_0 is the fluid density and $f_N = \sqrt{K/m}/2\pi$ is the natural frequency of the mass–spring system in vacuum. In this study, we focus on the parameter combination of $m^* = 2$ and $U_R = 5$, at which the VIV is locked in and is more difficult to control (Du & Sun 2015; Zhang *et al.* 2015).

The hydrodynamic forces, the drag F_D and the lift F_L, are normalized by $\frac{1}{2}\rho_0 U_0^2 D_0$. The lift force can be further decomposed into two components, the potential force F_A due to the added-mass effect and the vortex force F_V. Generally, we have $F_A = -C_a m_d \ddot{y}$, where $m_d = \frac{1}{4}\pi \rho_0 D_0^2$ is the displaced fluid mass and C_a is the coefficient. Although an empirical value of $C_a = 1.0$ had been frequently used for circular cylinders, as suggested by Govardhan & Williamson (2000), this value has been proven to vary with the reduced velocity and rotation rate (Bourguet & Lo Jacono 2014). In this study, C_a is determined using the measured vibration frequency, based on the relationship between the actual VIV frequency and total mass, i.e.

$$f_{viv} = 2\pi \sqrt{\frac{K}{m + C_a m_d}}. \tag{2.3}$$

In addition to F_V and F_A, the cylinder is also subjected to the elastic force F_E exerted by the spring. After being normalized by $\frac{1}{2}\rho_0 U_0^2 D_0$, they are then denoted as C_V, C_A and

990 A7-5

C_E, respectively. The first two forces vary linearly with the kinematic variables, while the vortex force describes the hydrodynamic behaviour. Substituting their definitions into (2.1) and performing the normalization, one can calculate C_V as

$$C_V = \left(\frac{\pi}{2}C_a + 2m^*\right)\ddot{y}^* + 8\pi^2 \frac{m^*}{U_R^2}y^*, \tag{2.4}$$

where $\ddot{y}^* = \ddot{y}D_0/U_0^2$ and $y^* = y/D_0$ are the normalized transverse acceleration and displacement, respectively.

To mediate the flow, we apply rotary motion to the cylinder that is adjustable in real time, as in figure 1. Here, the angular velocity ω can be either positive (corresponding to the anticlockwise direction) or negative (corresponding to the clockwise direction). The dimensionless rotational velocity is then defined as $\omega^* = \omega D_0/U_0$ and is limited to a range of $[-1, 1]$. In this sense, the tangential velocity on the edge of the cylinder would not exceed $U_0/2$. Compared with prior studies where the open-loop rotary control is adopted, such as the work of Du & Sun (2015), the prescribed range of the rotational amplitude is relatively small so as to provide efficient control. To quantify the control energy, we introduce the momentum coefficient to characterize the kinetic energy of the cylinder undergoing rotary motion, as well as the power coefficient to characterize the power required to drive the rotations. These are formulated as

$$C_\mu = \frac{\frac{1}{2}I\omega^2}{\frac{1}{2}mU_0^2} = \frac{1}{8}\omega^{*2}, \quad C_P = -\frac{1}{2}C_M\omega^*, \tag{2.5a,b}$$

where I represents the moment of inertia of the cylinder and we assume that its mass is evenly distributed. Here C_M represents the torque exerted on the rotating cylinder, normalized by $\frac{1}{2}\rho_0 U_0^2 D_0^2$.

In the closed-loop AFC system, we choose the kinematic variables along the cross-flow direction as feedback signals, including the displacement y, the velocity $\dot{y}$ and the acceleration $\ddot{y}$ of the cylinder. Unlike the velocity sensor array used in prior studies targeted at reducing drag (Rabault *et al.* 2019; Tang *et al.* 2020; Ren *et al.* 2021*a*) or suppressing VIV (Zheng *et al.* 2021), our choice of feedback signals is easy to realize in experimental or real engineering applications.

For the specified actuation and sensors, the control law can be generally expressed as

$$\omega_{t+1}^* = f(y_t^*, \dot{y}_t^*, \ddot{y}_t^*), \tag{2.6}$$

where the subscripts 't' and '$t+1$' mean that the control is conducted in a real-time manner. Here '$t+1$' is the actuation instant followed by the time step when the feedback signals are sampled. Here $\dot{y}^*$ is the velocity normalized by U_0. In this study, the control law f is modelled in the form of an 'actor' network, which is then determined by the DRL agent. A detailed introduction to the DRL-base control strategy will be given in § 2.3.

Targeting at the suppression of the VIV, we introduce the following reward:

$$r = \langle -|y^*|\rangle_A, \tag{2.7}$$

where $\langle\cdot\rangle_A$ denotes the average over the duration of one actuation. With the relatively strict momentum constraint in this study, the extra penalization terms related to energy cost that have been commonly adopted in prior studies are unnecessary here. The results in subsequent sections will show that the present reward definition is sufficient for the VIV suppression target. Therefore, the objective of this study is to seek the optimal control strategy generalized in (2.6) so as to achieve the maximum reward expressed in (2.7).
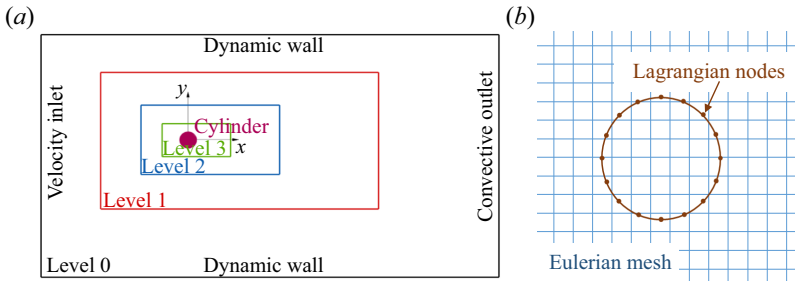
Figure 2. (*a*) The computational domain with four-level mesh refinement (not in scale). (*b*) A schematic diagram of the immersed boundary method to model the moving cylinder as Lagrangian nodes distributed with equal spacing.

Re	Case	Mesh resolution	Fixed cylinder			VIV cylinder	
		$D_0/\delta x$	C_D	C_L	St_1	y_A	St_2
100		32	1.396 ± 0.010	± 0.347	0.164	± 0.550	0.181
	Present	64	1.383 ± 0.010	± 0.342	0.166	± 0.539	0.183
		128	1.375 ± 0.010	± 0.339	0.166	± 0.534	0.184
	Russell & Wang (2003)	—	1.38 ± 0.007	± 0.332	0.169	—	—
	Choi *et al.* (2007)	—	1.34 ± 0.011	± 0.335	0.164	—	—
	Wang *et al.* (2016*a*)	60	—	± 0.337	0.169	± 0.545	0.186
	Ren *et al.* (2019)	64	1.374 ± 0.011	± 0.338	0.169	± 0.522	0.186
300		32	1.376 ± 0.085	± 0.915	0.164	± 0.496	0.197
	Present	64	1.385 ± 0.088	± 0.931	0.166	± 0.504	0.202
		128	1.381 ± 0.087	± 0.928	0.166	± 0.503	0.204
		256	1.378 ± 0.090	± 0.924	0.166	± 0.502	0.204

Table 1. Comparisons with prior studies and mesh convergence results. The Strouhal numbers St_1 and St_2 are the vortex shedding frequencies in the fixed and VIV scenarios, respectively, normalized by T_0^{-1}. Here y_A represents the amplitude of the cross-flow displacement.

2.2. *Numerical environment*

In this study, unsteady computational fluid dynamics (CFD) simulations are conducted to provide training data for the DRL-guided AFC and to assess the performance of the learned control strategy. In all simulations, the fluid is treated as incompressible and Newtonian. We adopt the lattice Boltzmann method to numerically solve the spatiotemporally evolving FSI process. In this method, we use the He–Luo model (He & Luo 1997) to ensure fluid incompressibility and the multirelaxation time scheme (D'Humières *et al.* 2002) to enhance the numerical stability.

Figure 2 shows a schematic of the computational domain, the grid partition and the boundary conditions. In the subsequent training stage, the size of the computational domain is $64D_0 \times 32D_0$, and the circular cylinder is initially located at the centreline, $20D_0$ downstream of the inlet. To allow the dynamic mode decomposition (DMD) analysis in the near-wall wake region, the computational domain is extended to $128D_0 \times 64D_0$ in the subsequent deterministic control stage. The multiblock grid partition method (Yu, Mei & Wei 2002) is utilized to balance the computational accuracy and efficiency. We adopt a four-level grid refinement, where the mesh resolution is doubled from level 0 to level 3. Around the cylinder, the mesh resolution of the finest block (level 3) is shown in table 1.

The inlet velocity is set as $U_0 = 0.04c$, where c is the lattice speed, corresponding to a time step $\delta t = T_0/1600$, where $T_0 = D_0/U_0$ is set as the reference time period. With the He–Luo incompressible model, as well as a small Mach number $Ma = 0.069$, the effect of fluid incompressibility is marginal.

We apply the Dirichlet boundary condition at the inlet and top/bottom walls, which is achieved via a modified bounce-back scheme with momentum exchange (Ladd 1994). The convective flow condition, i.e. $\partial_t \phi + U_0 \partial_x \phi = 0$, is utilized at the outlet to allow the vortices to smoothly cross the boundary with the least reflection (Fakhari & Lee 2014), where ϕ denotes either component of the velocity. Similar grid partitions and boundary set-ups have been extensively used in our previous studies (Ren *et al.* 2019, 2021*b*; Wang *et al.* 2016*a*, 2017*a*,*b*, 2016*b*).

Unlike the curved boundary treatment in our prior studies (Ren *et al.* 2019, 2021*b*), we model the moving cylinder using the multidirect forcing immersed boundary method (Wang, Fan & Luo 2008). In this method, the moving boundary is modelled with a group of Lagrangian nodes, where the no-slip and no-penetration conditions of these boundary nodes are transformed into body forces exerted on the surrounding fluid nodes. The Euler–Lagrange mapping process is iterated five times to reduce numerical errors. The Lagrangian node spacing is fixed at $\pi\delta x/4$, to ensure that the Lagrangian node spacing is smaller than or close to the Eulerian grid spacing. The hydrodynamic forces and torque exerted on the cylinder are then integrated based on the Eulerian forces. The kinematic equation, i.e. (2.1), is solved using a third-order total variation diminishing Runge–Kutta method (Gottlieb & Shu 1996), so as to accurately obtain the cylinder's instantaneous velocity and displacement. Note that the internal mass effect when solving the FSI process is calculated using the rigid body approximation (Feng & Michaelides 2009), as has been well verified by Suzuki & Inamuro (2011).

Table 1 lists the mesh convergence results and comparisons with data from prior studies, where C_D is the drag coefficient and C_L is the lift coefficient. At $Re = 100$, since the mesh resolution with $D_0/\delta x = 32$ provides a better choice to balance the computational efficiency and accuracy, we apply this set-up in the subsequent DRL training stage. For the deterministic control, we choose a finer mesh resolution with $D_0/\delta x = 64$ for the sake of accuracy. For cases in § 3.6, where we also have $Re = 200$ and $Re = 300$, we choose the mesh resolution of $D_0/\delta x = 64$ for both the training stage and the deterministic control.

To satisfy the strict requirements of DRL-guided AFC in terms of both computational accuracy and efficiency, we apply an in-house graphics processing unit (GPU)-accelerated solver (Ren *et al.* 2018), on a hardware platform that includes an Intel Xeon E5-2678 central processing unit and an NVIDIA Titan V GPU. This solver greatly shortens the time required for the DRL agent to complete one trial, which is less than 2 min in the following training stage using the coarse mesh and approximately 15 min for one deterministic run.

2.3. *Deep reinforcement learning guided control*

Unlike conventional approaches for constructing a controller, which rely heavily on prior knowledge of the system, the DRL agent does not have any knowledge of fluid dynamics and thus is regarded as a model-free approach. In order to learn the control strategy, we apply proximal policy optimization (Schulman *et al.* 2017) as the DRL agent, which is advantageous for determining continuous actions.

In the DRL framework, effective control strategies are learned through interactions between the DRL agent and the VIV environment. In the beginning, the DRL interacts with the VIV environment through randomized actions. Through trial and error, it learns how to exert a certain action when the system is in a certain state. In the meantime, it learns
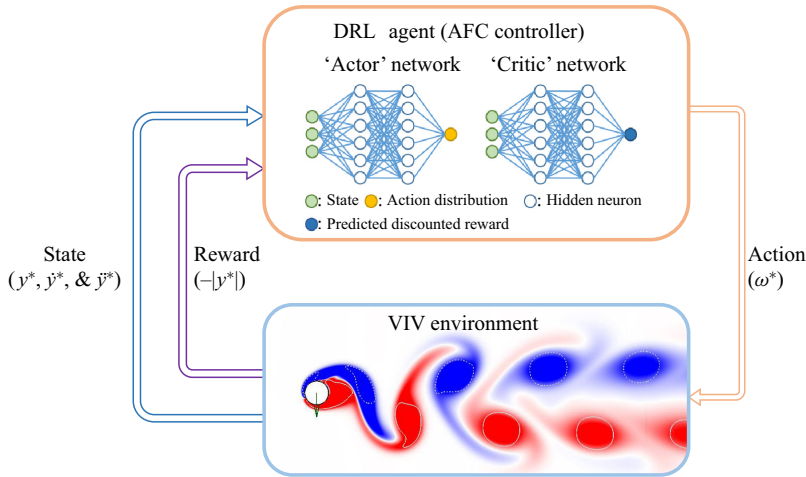
Figure 3. A schematic diagram of the DRL loop used in the present study. The DRL agent adopts two independent neural networks for decision-making ('actor') and reward evaluation ('critic'). In the VIV environment, the instantaneous vorticity field is shown, together with grey lines to identify vortex structures based on the λ_{ci} criterion (Zhou *et al.* 1999).

how to predict the long-term reward with this state information. As shown in figure 3, the state of the environment is represented by sensory-motor cues, and the action is the rotary motion imposed on the cylinder. The reward defined in (2.7) is then evaluated and fed to the agent, providing a baseline with which the agent can learn how to evaluate the control performance, so as to encourage it to develop a better controller that achieves a higher reward.

Two sets of artificial neural networks are used in the agent, generally named the 'critic' and 'actor' networks. As depicted in figure 3, both networks use the state as their input. On the output side, the 'critic' network estimates the long-term discounted reward, whereas the 'actor' network models the control strategy. The mathematical expression for the objective of the 'critic' network is

$$J_{critic} = \hat{E}_t(-\hat{A}_t^2), \tag{2.8}$$

where the advantage function $\hat{A}_t = R_t - V_\Theta(s_t)$ determines the gain between the real long-term performance $R_t = \sum_{t'>t} \gamma^{t'-t} r_{t'}$ and that predicted by the 'critic' network V_Θ for a given s_t. Herein, the parameter γ is a discount factor close to unity, so the long-term reward is preferred. In contrast, the objective of the 'actor' network is

$$J_{actor} = \hat{E}_t[\min(q_t(\Theta)\hat{A}_t, \text{clip}(q_t(\Theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_t)], \tag{2.9}$$

where $q_t(\Theta) = \pi_\Theta(a_t|s_t)/\pi_{old}(a_t|s_t)$ is the ratio of the probability of the current policy π_Θ adopting action a_t according to state s_t to the probability of the previous policy π_{old}. The clip term inside (2.9) indicates that $q_t(\Theta)$ is constrained to an interval $[1 - \epsilon, 1 + \epsilon]$, where ϵ is the clipping rate needed to avoid an excessively large policy update. By using this proximal policy optimization-based DRL framework, it is expected that the agent can learn effective control strategies through stochastic trial and error. More details on this framework can be found in our prior studies (Ren *et al.* 2021*a,b*). The hyperparameters used in this study are listed in table 2. The adaptive moment estimation optimizer

Parameter	Value/Method	Comment/Reference
Network architecture (actor and critic)	50×50	Two fully connected layers
Actuation function	Swish	Ramachandran, Zoph & Le (2017)
Actuations per T_0	10	—
Length of each episode	$100T_0$	—
Optimizer	Adam	Kingma & Ba (2014)
Action distribution	Beta	Chou, Maturana & Scherer (2017)
Discount factor (actor and critic)	0.97	Empirical value, $0.9 \sim 0.999$
Learning rate (actor)	3×10^{-4}	Empirical value, $10^{-4} \sim 10^{-2}$
Learning rate (critic)	4×10^{-4}	Empirical value, $10^{-4} \sim 10^{-2}$
Clipping rate	0.2	Schulman *et al.* (2017)

Table 2. Hyperparameters used in the DRL during the training stage.

(abbreviated as 'Adam') is used when updating the parameters of both neural networks, with relatively small learning rates being used to help the training converge more stably.

Like many other machine learning applications that generally involve an exploration–exploitation process, DRL-guided AFC can also be divided into a training stage and a deterministic control stage. At the beginning of the training stage, the parameters of both neural networks are randomly initialized. During the training, the CFD simulation of each episode lasts for $100T_0$, longer than 18 vortex shedding periods of the uncontrolled case. An episode is a complete run with a reinitialized flow environment using a fully developed flow field and a control strategy updated according to the state-action–reward chain obtained in the last run. In each episode, states are extracted 10 times within one T_0, and the actions are also enforced at the same pace. Furthermore, to enlarge the search space, actions are sampled from the bias-free Beta distribution (Chou *et al.* 2017).

After the training reaches convergence, the deterministic control is performed, where the neural networks stop updating and only the 'actor' network is utilized to determine the action(s) with the given state data. Moreover, each action is no longer sampled from the aforementioned probabilistic distribution but is directly assigned its mathematical expectation to avoid randomness. To reach good convergence, the deterministic control lasts for $200T_0$, and more actuations (100 times within one T_0) are sampled to perform smoother actuations.

3. Results and discussion

3.1. *Interpretable learning with different state information*

In a control system, the choice of sensors usually plays a crucial role. It is thus essential to determine the impact of each sensor as well as their combinations, so as to identify the most influential sensors. First, we explore the effects of different combinations of kinematic variables that provide feedback signals. Using the same hyperparameters listed in table 2 but different random seeds, we performed six groups of trainings and plotted the learning curves in figure 4, where the Reynolds number is all fixed as 100. Here $\langle \cdot \rangle$ indicates an average over one episode. Because the training process involves a certain level of randomness, we conducted three trials for each training in figures 4(*a*) and 4(*b*) and observed quite similar learning trends. For the cases in figure 4(*c*) that will be the focus of the remaining sections, we performed five trials. Therefore, figure 4 consists of 23 training trials, i.e. 46 000 individual CFD cases in total.
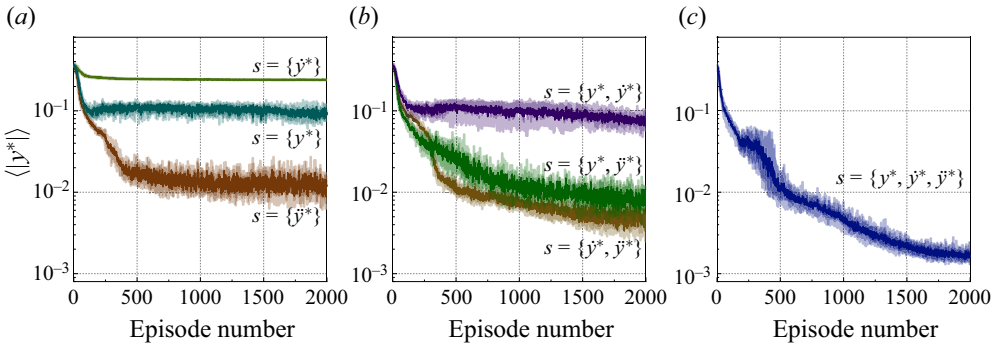
Figure 4. Training processes using all six types of combinations of sensory-motor cues as the state space. In each training process, three or five independent trials are performed. The associated performances are shown with translucent lines, while the mean data are shown as thick solid lines.

In most trials, the learning starts with a rapidly decreasing trend and then proceeds with a gradually converging trend. All the learning curves eventually reach good convergence. Taking figure 4(c) as an example, the DRL reaches a level of $\langle|y^*|\rangle \approx 0.01$ roughly at the 500th episode, then progresses rather slowly beyond this point. Although the learning curve sometimes appears to show a non-decreasing trend, the agent can always escape from the locally optimal strategy and find the right search direction again after some unsuccessful trials. This fact reflects the self-adaptivity and reliability of the DRL agent in exploring nonlinear systems.

By comparing the learning curves of different combinations of sensors, it can be observed that using the velocity sensor alone leads to the worst control performance, while the combination of all three kinematic variables as sensors provides the best-converged strategy, where $\langle|y^*|\rangle$ eventually reaches a level of approximately 0.002, which is a remarkable level of VIV suppression. In addition, the $\ddot{y}^*$ sensor alone can lead to a VIV suppression to approximately 0.01, yielding a less effective but simpler alternative for the control.

One question that comes to mind is how the DRL agent searches the optimal strategy in one single training. In other words, how does the AI agent think and make decisions during the exploration process? To answer these questions, we select four representative training trials with different state spaces from figure 4 and plot the scatters of the mean drag, vortex lift fluctuation and cross-flow displacement against the AFC forcing strength in figure 5. Although, as we have mentioned, the search path involves a certain level of randomness, it is surprising to see that figure 5 reveals some clear search paths. In particular, by comparing the search paths of $\langle|y^*|\rangle$ in figure 5(a iii), one can observe that the training process can be roughly divided into three stages. In Stage I, the DRL agent explores a strategy that applies larger and larger forcing strength, which does gradually improve the control performance before hitting a plateau at approximately the 150th episode, when $\langle|y^*|\rangle$ becomes close to 0.08, with the fluctuation of ω^* approaching its saturated value as can be more clearly observed in figure 7(b). In Stage II, a sudden turn appears in the $\langle|C_D|\rangle \sim \langle\omega^{*2}\rangle^{1/2}$ trajectory as well as the $\langle|y^*|\rangle \sim \langle\omega^{*2}\rangle^{1/2}$ trajectory, occurring roughly at the 150th–200th episode. Unlike in Stage I, although the forcing strength in Stage II gradually decreases, the control performance keeps improving, and reaches a level of $\langle|y^*|\rangle \approx 0.01$ at the end of Stage II around the 1000th episode. The subsequent Stage III experiences a slower converging stage, in which the DRL agent carefully adjusts its strategy. Eventually, the DRL agent achieves a strategy that almost
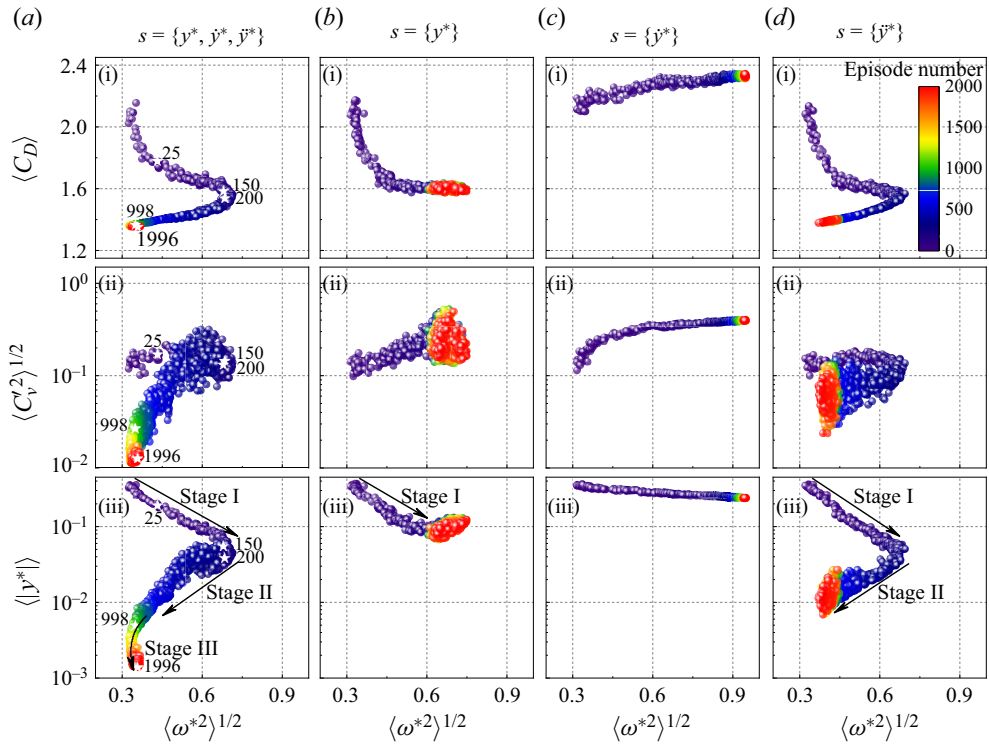
Figure 5. Trajectories of the (*a*) mean drag, (*b*) r.m.s. of the vortex force and (*c*) absolute value of the transverse displacement against the AFC forcing strength during training. Four training processes with different combinations of sensory-motor cues are shown in (*a–a*). The scattered points are coloured with the episode number, ranging from 1 to 2000. Five representative cases in subpanels of (*a*) are marked with white star.

completely suppresses the VIV. The $\langle C_D \rangle \sim \langle \omega^{*2} \rangle^{1/2}$ curve shows a similar trend to that of the $\langle |y^*| \rangle \sim \langle \omega^{*2} \rangle^{1/2}$ curve, and the close relationship between $\langle C_D \rangle$ and $\langle |y^*| \rangle$ is further confirmed by figure 6(*a*). As the main hydrodynamic source that drives the VIV, the root-mean-square (r.m.s.) of the fluctuation part of C_V, i.e. C_V', shows a gradually decreasing trend, except for scatter at the very beginning. From the trajectories in figure 5(*a*), one can see that the DRL agent adopts different strategies in the exploration process, and the fact that changing tactics when proceeding along one direction can no longer improve the performance implies that the decision making of the DRL agent is adaptive.

However, one can only find a similar Stage I in figure 5(*b* iii) and in the first two stages in figure 5(*d* iii). In contrast, figure 5(*c* iii) only shows a gradually decreasing trend and eventually reaches a high level of vibration displacement, while exerting increasing AFC forcing and causing increased drag. Recalling figure 4(*a*), this case is actually a failure. The trajectories illustrated in figure 5 reflect that properly selected sensors play a crucial role in DRL-guided AFC, and can lead to distinct exploration paths during training. Importantly, the revealing of distinct stages implies that while exploring the VIV environment, the DRL agent is smart and capable enough to make radical adjustments to its control strategy when the existing learned strategies cannot make further contributions, which we believe is related to the underlying physics of the system being controlled.
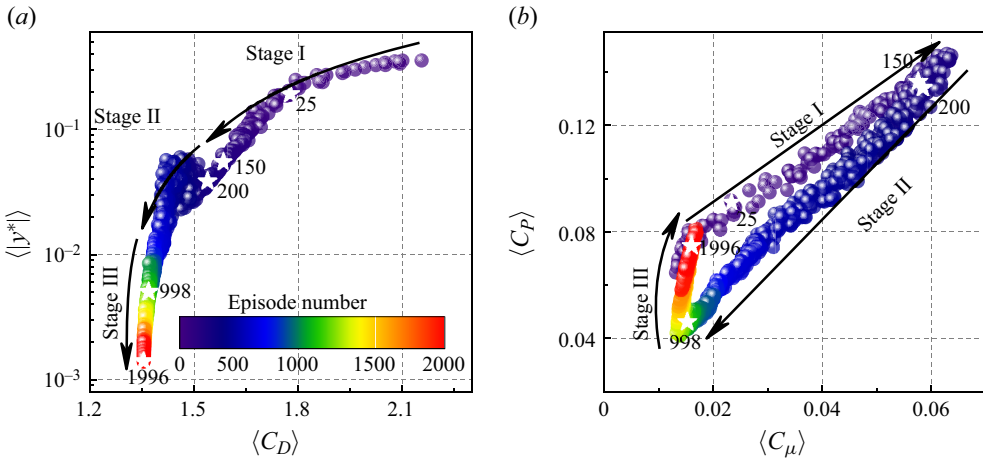
Figure 6. The trajectories of (*a*) the mean drag versus the absolute value of the transverse displacement and (*b*) the momentum coefficient versus power coefficient. Training with the full state space $\{\ddot{y}^*, \dot{y}^*, y^*\}$ is used in this case.

To better understand the physics behind the search path, we draw the $\langle C_D \rangle \sim \langle |y^*| \rangle$ curve and the $\langle C_\mu \rangle \sim \langle C_P \rangle$ curve in figure 6. The nearly monotonic variation of $\langle C_D \rangle$ against $\langle |y^*| \rangle$ implies that the reduction in the mean drag of the controlled cylinder is mainly accompanied by the reduction in the distance swept by the VIV cylinder. This close connection also suggests that C_D could potentially be a good indicator of VIV control. In other words, in future applications, the drag and lift forces exerted on the cylinder could be good alternatives to provide feedback control signals. Figure 6(*b*) shows the connection between the kinetic energy of the rotary cylinder and the power supplied to the actuator. It is quite surprising to see that the three stages identified in figure 6 are the three edges of a triangle-like closed loop. That is, the initial $\langle C_\mu \rangle \sim \langle C_P \rangle$ pair roughly overlaps that in the final stage. To unveil more insights, a comparison will be made in § 3.3 between the initial control strategy (taking the 25th episode as a representative) and the final control strategy (the 1996th episode), which consume similar amounts of energy. It is also noted that Stage II proceeds along an energy-saving direction. However, throughout Stage III, C_μ is almost fixed while the energy supplied to the actuator monotonically increases. Eventually, the performance is further improved at the cost of an almost doubled energy input.

During the exploration process, since the actuation is sampled from a probabilistic distribution determined by the DRL agent, the recorded actuation and state both involve a certain degree of noise. To remove this impact, we utilize the control strategies learned from the 5th to the 1995th episode (*ep* is the episode number), and perform deterministic runs every five episodes. In the second half of theses deterministic runs, the mean value, the fluctuation amplitude and the frequency of actuation are obtained, as depicted in figure 7(*a–c*). In figure 7(*c*), because the fast Fourier transformation is performed using the second half of each episode that involves $100T_0$, i.e. 10 000 actuation samples, the finest frequency resolution is 0.01. This is why the action frequency appears discontinuous. From the case validations illustrated in Appendix A, we show that the actuation frequency follows exactly the same frequency as the kinematic variables, which is consistent with the overlapping frequency of ω^* and y^*. We also calculate the phase lag between the
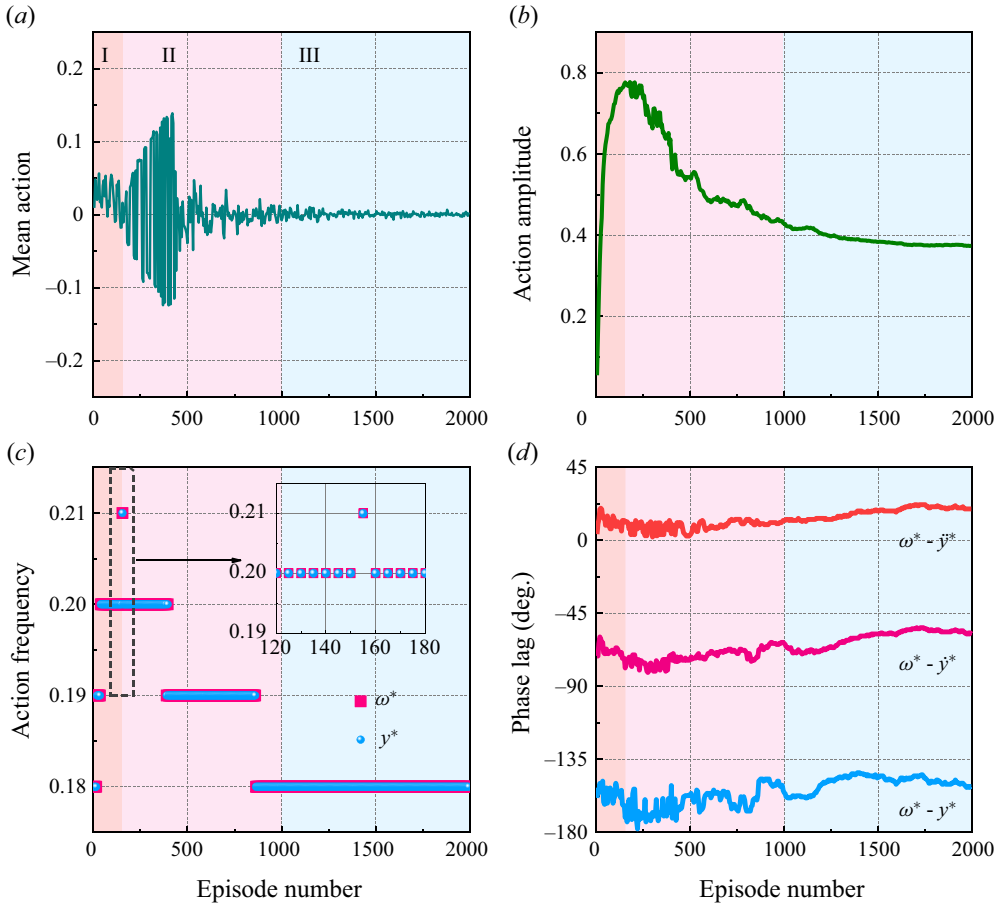
990 A7-13

Figure 7. The variations of the control quantities during a deterministic run from the fifth episode to the 2000th episode with an interval of five episodes. The (*a*) mean actuation, (*b*) fluctuation amplitude of actuation, (*c*) frequency of actuation and (*d*) phase lag between the actuation and kinematic variables. The three stages are identified again and classified using different background colours.

actuation and the kinematic variables based on correlation analysis and present the results in figure 7(*d*).

From figure 7, one can deduce the following features of the three stages. In Stage I ($ep < 150$), the DRL agent adjusts both the amplitude and frequency of the AFC forcing, both being increased. In Stage II ($150 < ep < 1000$), the DRL agent also adjusts both the amplitude and frequency of the AFC forcing, both being gradually reduced though. In Stage III ($ep > 1000$), the DRL agent gradually adjusts its phase lag against the kinematic variables. A phase lag close to $20°$ between ω^* and $\ddot{y}^*$ is finally obtained. In both Stages I and II, action biases, i.e. non-zero mean actions, are observed. It takes a long trial process for the bias to disappear at the end of Stage II. This problem almost vanishes in Stage III.

In the above divisions into three stages, the transition boundary from Stage I and Stage II is quite clear based on either the peak of the action amplitude or the location of the largest action frequency. However, because the change from Stage II to Stage III is relatively vague, in figure 7, the boundary between Stage II and Stage III is only a rough approximation; from the aspect of energy utilization shown in figure 6(*b*), where
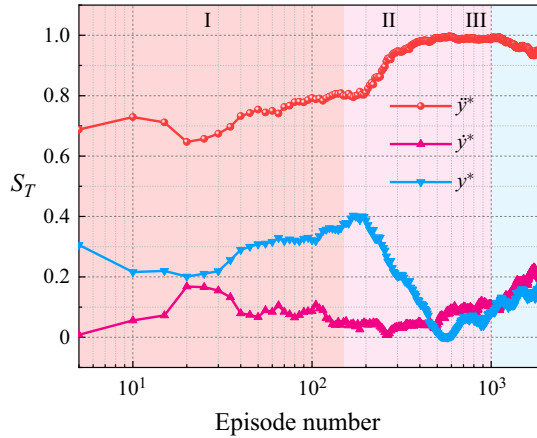
Figure 8. The total sensitivity index for all sensory-motor cues during one training. The sensitivity analysis is conducted from the fifth to the 2000th episode with an interval of five episodes.

C_μ reaches its minimum and C_P starts to increases again. It also remains a question why Stage I transits to Stage II, which is not simply a consequence of saturated action. This question will be discussed in § 3.3.

It is observed from figure 7(*d*) that the phase between ω^* and $\ddot{y}^*$ is the closest, suggesting that $\ddot{y}^*$ always reacts with changes in ω^* faster than the other two sensors. We guess this is a key reason why the $\ddot{y}^*$ sensor is selected from the very beginning, and plays a key role during the whole exploration process.

Note that the above interpretations of the learned control strategy are essentially different from the cluster method (Fernex *et al.* 2021; Li *et al.* 2021), where the state information is projected onto a low-dimensional phase space and different strategies are then clustered. By contrast, the way we interpret the search path is less general but more straightforward from the physics perspective, and benefits from a small number of sensors. In addition, the effect of hyperparameters used in the DRL is examined and the data are shown in Appendix B, where it is indicated that the above findings are less sensitive to the hyperparameters. The fact that different hyperparameters lead to the very similar division of three stages confirm that the above findings are reasonable and solid.

3.2. *Sensitivity analysis*

For control with three sensor signals, it is worthwhile to know how sensitive each sensor is in the control strategy. Therefore, we perform a sensitivity analysis via the Sobol method (Saltelli 2002; Sobol 2014), based on SALib – a sensitivity analysis library implemented in Python. Here the 'actor' network that establishes the mapping between sensor signals and the control action is utilized, together with randomly selected values within the actuation range. Total sensitivity indices S_T are used to quantify the sensitivity of each control input to the control output. According to the definition in Sobol (2014), S_T is an invariance-based sensitivity index that measures the contribution of the variance of each input argument to the variance of the output, which including all variance caused by its interactions, of any order, with any other input variables. In a whole training, S_T of each sensory-motor cue is shown in figure 8, where, the same as in the deterministic runs, we perform the sensitivity analysis from the 5th to 2000th episode with an interval of five episodes.

First, it is seen that the $\ddot{y}^*$ sensor maintains a much higher sensitivity level than the other two sensors. This qualitative conclusion is consistent with results in figure 4, where the $\ddot{y}^*$ sensor is the most effective one. Second, the evolution of S_T also exhibits similar features to the three stages identified in § 3.2. As revealed in figure 7, the DRL agent adjusts the actuation amplitude and frequency during Stages I and II, where theoretically one signal among the three kinematic variables is enough. This is why S_T of $\ddot{y}^*$ shows an increasing trend during these two stages. However, because Stage III is a phase-adjusting stage and the $\ddot{y}^*$ sensor alone cannot achieve this, the contributions of the other two sensors start to grow during this stage. This explains why the S_T of both y^* and $\dot{y}^*$ show a dramatically increasing trend in Stage III. From the sensitivity analysis, one can also see that adjustments of the actuation amplitude and frequency are easier than the long-march phase adjustment process.

3.3. *Deterministic control using strategies learned at different stages*

After the DRL training reaches convergence, we obtain the control law established by the 'actor' network. Among the 23 training processes, we select a few cases with a full state space $\{\ddot{y}^*, \dot{y}^*, y^*\}$ as representatives, occurring at the 25th, the 150th, the 200th, the 998th and the 1996th episode. We conduct a group of deterministic controls using these strategies learned at different episodes to elucidate the difference of control performance.

Figure 9 demonstrates temporal variations of the rotary forcing recorded in episodes of different stages. From figure 9(*e*), one can see that the actuation reaches saturation in the early few control cycles, and then gradually becomes quasisteady. To maintain the control performance, a rotational velocity with an amplitude of approximately 0.4 is required, corresponding to the linear velocity at the edge of the cylinder being less than $U_0/5$. Note that the vortex shedding frequency does not deviate from the natural resonance frequency, as confirmed in Appendix A, suggesting that the main physical mechanism of VIV suppression lies in that the DRL-guided control successfully alters the flow field to an attenuated vortical flow, and that the lift generated by the vortices are balanced by the Magnus effect induced by the rotary forcing. Intriguingly, from all selected cases, as well as the information in figure 7(*c*), the DRL agent does not proceed with seeking a proper frequency being the vortex shedding frequency in the uncontrolled situation, or its super harmonics or subharmonics, so as to utilize the 'lock-on' effect, which is what the AFC community usually adopted.

One can now observe another significant fact that at Stage II, the DRL agent gradually learns the control strategy that a large forcing shall be exerted at the first few periods so as to rapidly mitigate the vibration to a new and low-amplitude state. After that, the periodic or quasisteady forcing is applied to fine tune this state. In this sense, the initial stage of each episode that involves a few vibration cycles with higher forcing strength actually plays a vital role in achieving the final VIV suppression objective. Without this initial process, the VIV control can hardly be realized even with a higher-level of forcing, just like cases in Stage I. This finding suggests that the DRL agent is really smart. This type of control strategy can hardly be done in the open-loop manner.

Temporal variations of quantities related to energy consumption, i.e. C_P and C_μ, are depicted in figure 10. Under the converged control strategy at the 1996th episode, after the initial stage, in which the DRL-guided control strategy requires a relatively large energy input to rapidly reduce the vibration, only a small amount of energy input is needed to maintain the control performance in the quasisteady state. In contrast, the 25th episode does not show this kind of adaptation. Further quantitative comparisons are based on the
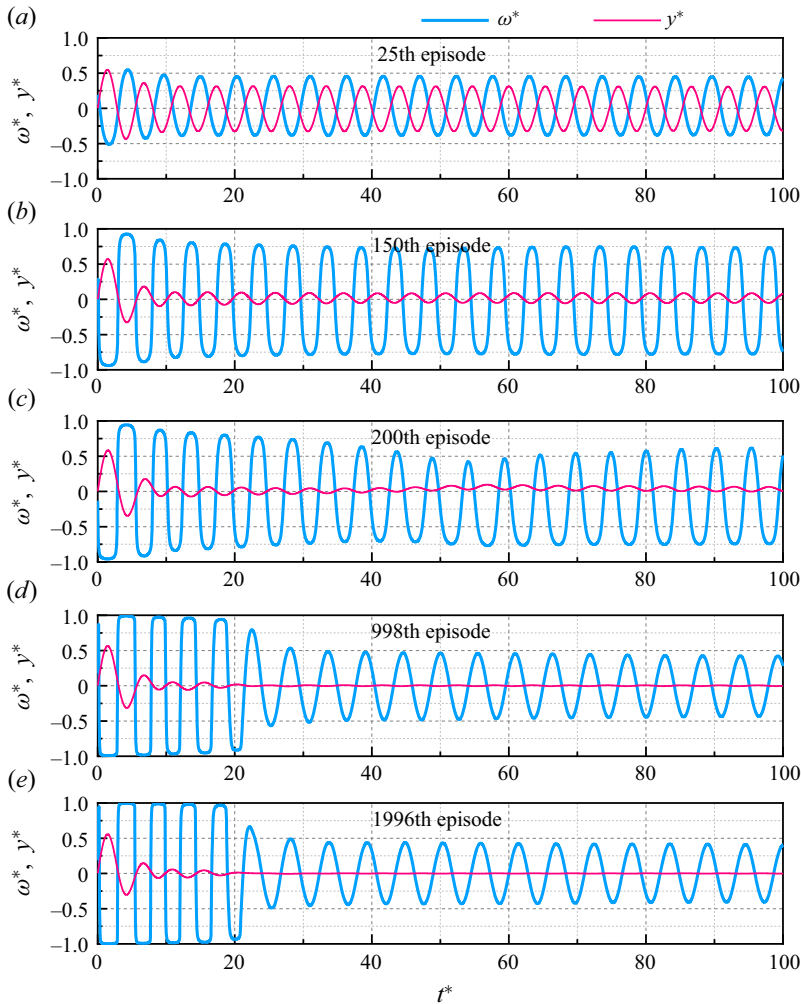
Figure 9. (*a*) Temporal variations of the rotational velocity after the AFC is turned on. (*b*) The frequency spectrum of ω^* in the quasisteady state. Here f^* denotes the frequency normalized by T_0^{-1}. The 25th, 200th and 1996th episode are representatives of Stage I, Stage II and Stage III, respectively. The 150th episode is the boundary between Stage I and Stage II and the 998th episode the boundary between Stage II and Stage III.

power-saving ratio (*PSR*), defined as

$$PSR = \frac{\langle \Delta P_D \rangle_T}{\langle P_C \rangle_T}, \tag{3.1}$$

where $\Delta P_D = -\Delta F_D U_0$ denotes the saved power used to drive the streamwise motion in the situation where the cylinder is cruising with a speed of U_0. This quantity is averaged within a time horizon of one vortex shedding period. Here P_C is the power consumed by the AFC. After normalization, we have $PSR = -\langle \Delta C_D \rangle_T / \langle C_P \rangle_T$, where ΔC_D is the reduced drag coefficient. Therefore, as estimated in the quasiperiodic state, the power saving ratio is 78.7 if compared with the uncontrolled VIV case, and 1.85 if compared with the uncontrolled stationary case. In some prior studies (Protas & Wesfreid 2002; Bergmann, Cordier & Brancher 2005), where rotary control was used for drag reduction
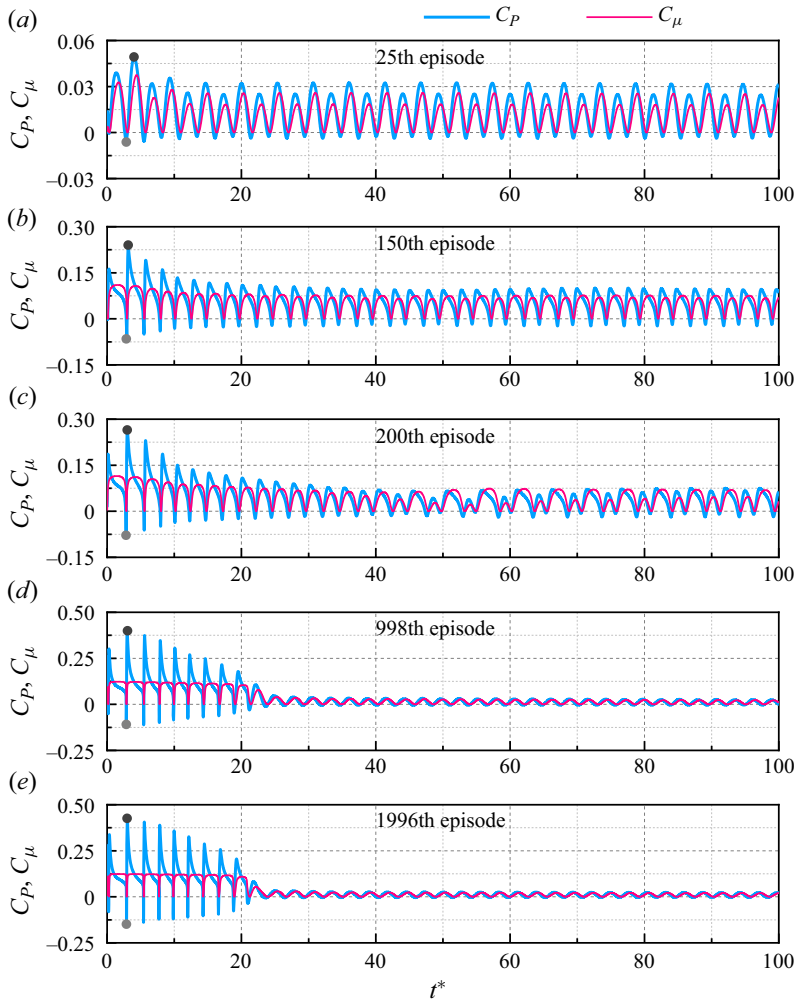
Figure 10. Temporal variations of the power coefficient and the momentum coefficient in the deterministic control using controllers learned in selected episodes.

under laminar conditions, *PSR* did not exceed 1.0. Therefore, the DRL-guided control presented in this study is energetically efficient.

Recalling that in figure 5(*a* i–iii), similar levels of rotary amplitude were found in the initial and final training stages, it is interesting to see how apparently similar actuations lead to distinct control results. To find the reason, we illustrate the temporal variations of ω^*, y^*, $\dot{y}^*$ and $\ddot{y}^*$ during one quasisteady cycle of the 25th and the 1996th episodes in figure 11. Four instantaneous snapshots are selected, when $\ddot{y}^*$, ω^*, $\dot{y}^*$ and y^* are, in sequence, at their lowest positions. Although a slightly biased ω^* appears in the 25th episode, the amplitudes of ω^* in the two cases are very close, i.e. approximately 0.4. It can be found that the forcing frequency of the 25th episode is slightly higher than that of the well learned strategy in the 1996th episode. Furthermore, the $\omega^* - \ddot{y}^*$ phase lag in the 25th episode is also slightly smaller, coinciding with the data in figure 7.

In figure 5, although the fluctuation amplitude of the rotary actuation in the two episodes do not show big differences, the two strategies do lead to distinct FSI dynamics.

990 A7-18
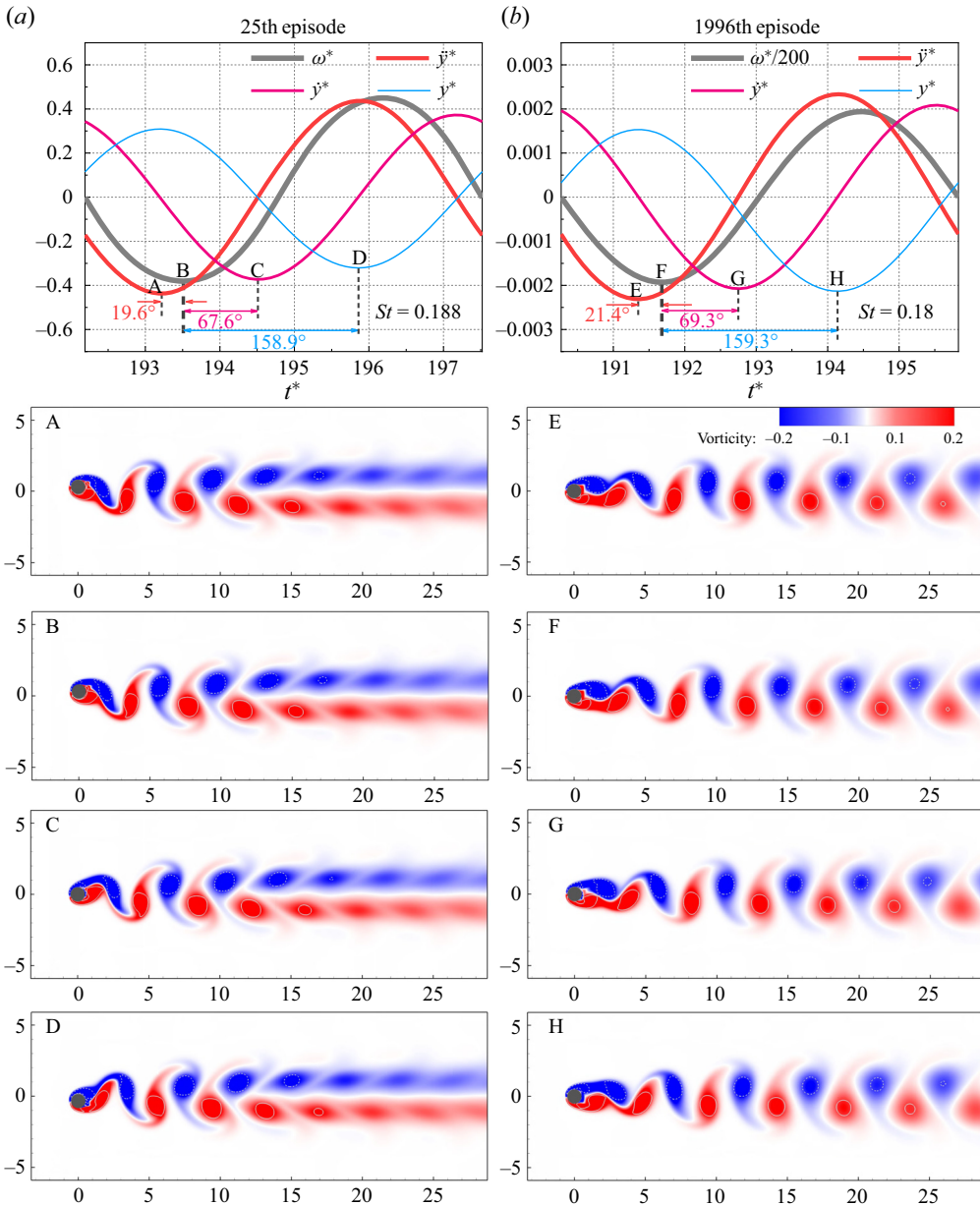
Figure 11. The temporal variation of the rotational forcing as well as the kinematic variables in the quasisteady state. Flow fields at four representative instants are demonstrated. The unsuccessful 25th episode is shown here for comparison purposes.

With the control strategy learned in the 25th episode, the wake exhibits a C(2S) vortex shedding pattern (Williamson & Roshko 1988) like in the uncontrolled case, where two single vortices shed within one vibration cycle, and the like-signed vortices coalesce in the downstream region, which is approximately 15 diameters from the cylinder. On the contrary, a transition to the 2S pattern is observed in figure 11(b). As seen in the VIV phase diagram provided by Williamson & Govardhan (2004), the 2S pattern is usually associated
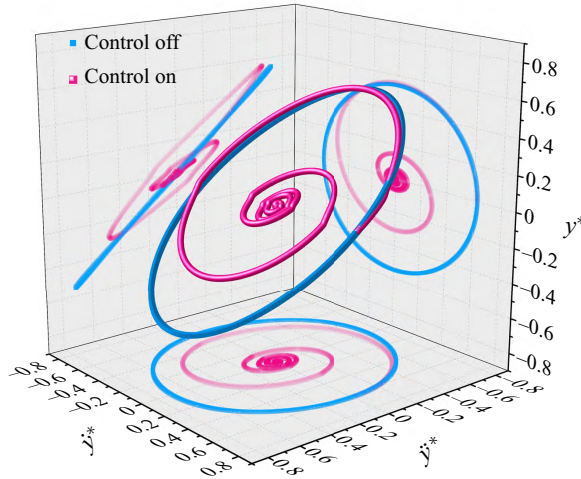
Figure 12. The 3-D scatters of $\ddot{y}^* - \dot{y}^* - y^*$, with projections on the 2-D planes. The scatters are denoted by blue square cubes and pink spheres, respectively, for instants before and after the control is turned on.

with a larger wavelength, or smaller vortex shedding frequencies, and this is confirmed by the above results. Intuitively, the near-wall shear layer in figure 11(*a*) forms an inclined angle using the 25th episode strategy, which is measured between the centreline of the wake and the boundary of the positive and negative shear layers. This phenomenon appears to be almost symmetrical in figure 11(*b*). This flow feature also reveals well-controlled structure vibrations.

From instant E to instant F, the negative rotational velocity (clockwise direction) transports momentum from the upper right-hand side of the wake to the lower right-hand side, mitigating the asymmetrical force induced by the near-wall flow structures. In this sense, maintaining the relatively good stability of the upper and lower shear layers is vital for the success of the present DRL-guided VIV control. Now, one can also imagine that if the rotary actuation is performed somewhat earlier (corresponding to a smaller $\omega^* - \ddot{y}^*$ phase lag) or later (corresponding to a larger $\omega^* - \ddot{y}^*$ phase lag), the influence of shear layers on the lift could become quite different. Evidence in the later § 3.4 would confirm that the DRL agent does remarkably enhance the stability of this FSI system.

3.4. *Deterministic control using the converged control strategy*

With the control strategy learned at the 1996th episode, we illustrate in figure 12 the three-dimensional (3-D) trajectory of $\ddot{y}^*–\dot{y}^* - y^*$, with and without the control. In this 3-D space, the uncontrolled trajectory forms a periodic orbit. With a phase lag of $180°$ between $\ddot{y}^*$ and y^*, the projection on the two-dimensional (2-D) $\ddot{y}^*–y^*$ plane appears to be roughly a straight line. After the control is turned on, this orbit gradually shrinks to its centre and finally reaches the quasiequilibrium state, a tiny closed loop in 3-D space.

Temporal variations of the hydrodynamic force and torque coefficients, as well as the vibration response, are depicted in figure 13. Less than five VIV periods, after the DRL agent starts to mediate the originally periodic FSI system, the FSI system falls into a quasisteady state. In this controlled state, the drag exerted on the cylinder is reduced by 37 % compared with the uncontrolled VIV case, accompanied by a reduced drag fluctuation of 98 %. Furthermore, the quasisteady drag resembles that in the uncontrolled
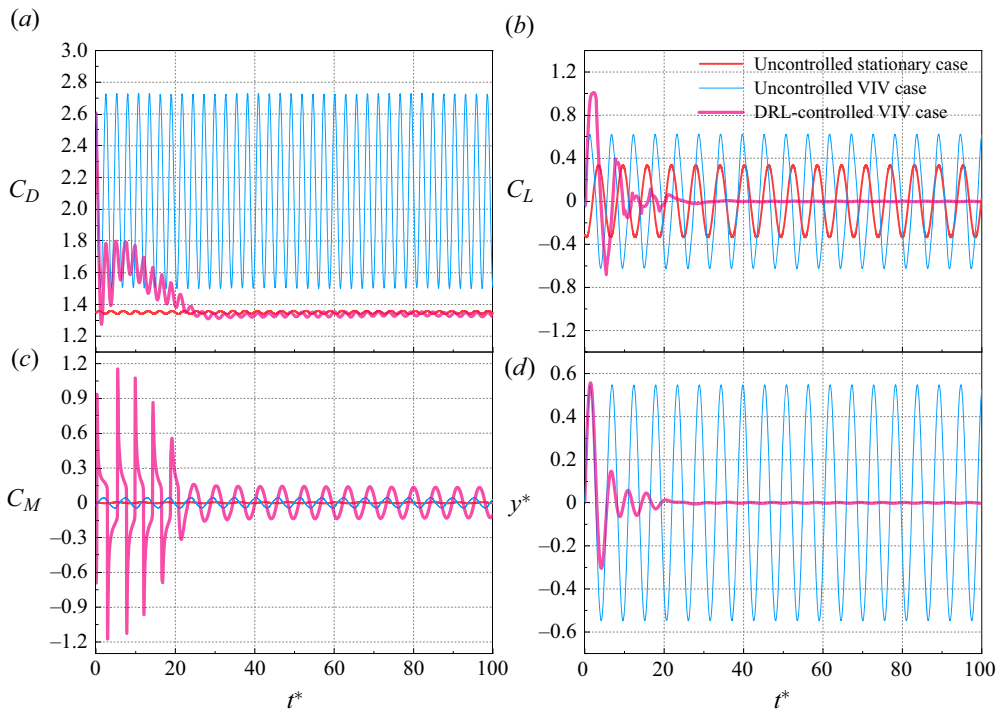
Figure 13. Time histories of (*a*) the drag coefficient C_D, (*b*) the lift coefficient C_L, (*c*) the moment coefficient C_M and (*d*) the transverse displacement y^*. The uncontrolled stationary case, the uncontrolled VIV case, and the DRL-controlled case are organized together for comparison purposes.

stationary case. Therefore, the drag reduction mainly stems from the narrowed distance swept by the cylinder, as shown in figure 13(*d*). In figure 13(*b*), the lift experiences a sharp change at the beginning of the control, even exceeding that of the uncontrolled VIV case in the first vortex shedding period, then decays quickly as the control continues and is eventually suppressed to almost zero. Meanwhile, a large torque with sharp variations in the beginning is required to drive the rotary motion, as can be seen in figure 13(*c*). Figure 13(*d*) demonstrates the successfully suppressed VIV, where the final vibration amplitude is only approximately 0.2 % of the cylinder diameter (an enlarged view can be found in figure 11), corresponding to a VIV amplitude suppression rate of 99.6 %. Furthermore, for the VIV case, only a 1.3 % frequency shift is observed after control. This almost unaffected VIV frequency indicates that the system is still in the locked-in condition.

Comparisons of the three lift force components are depicted in figure 14, where C_E is reduced by a factor of five for the sake of clarity. From the definitions in § 2.1, one can see that the vortex force is the hydrodynamic source that drives the VIV. The non-harmonic variation of C_V stems from the vortex shedding mode. After control, the fluctuation of C_V is reduced by a factor of over 700. This can be viewed as further evidence of well-suppressed VIV. It is noted that the mean C_V is not zero, which is a consequence of the non-zero mean actuation, as evidenced by the fast Fourier transformation results in figure 9(*b*).

In order to figure out why the VIV can be suppressed in the 1996th episode, the lift variation in one cycle is shown in figure 15 as well as three instants corresponding to
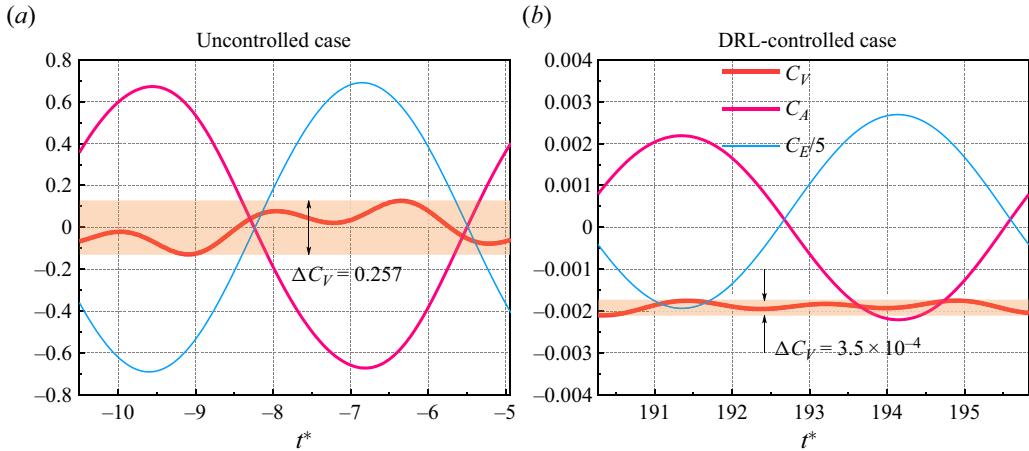
Figure 14. Temporal variations of the three lift components of (*a*) the uncontrolled VIV case and (*b*) the DRL-controlled VIV case (the vortex force, the added-mass force and the elastic force).

minimum, moderate and maximum lift force, respectively. The pressure field near the cylinder and the vorticity field are also depicted. For the three cases, the vortices that form and shed from the cylinder surface play key roles in lift generation and the resulting VIV.

In the uncontrolled stationary case, when C_L is at its minimum at instant A, the anticlockwise vortex (labelled as 'acw-vortex') at its lower surface is forming and the clockwise vortex (labelled as 'cw-vortex') is about to shed from the upper surface. Correspondingly, the pressure on its lower surface is much lower than that on the upper surface, resulting in a negative lift force. At instant B, both the acw-vortex and cw-vortex have moved to the downstream. The positive and negative shear layers lead to a balanced pressure distribution on the cylinder.

In the uncontrolled VIV case, at instant D, the cylinder is moving downwards and is near its lowest position. The cw-vortex is about to shed and has a lesser effect on the cylinder. The negative shear layer on the upper surface provides larger pressure than the positive shear layer on the lower surface, resulting in the maximum negative lift. At instant E, the cylinder is near its equilibrium position and is moving upwards with almost its largest transverse velocity. The upward motion creates a high-pressure region on the upper left-hand surface, meanwhile the cw-vortex provides a low-pressure region on the upper right-hand surface, leading to a balanced transverse force.

In the DRL-controlled VIV case, the well-controlled cylinder is nearly stationary. At instant G, since the cylinder is rotating in the anticlockwise direction (i.e. positive ω^*), the shedding of an acw-vortex is promoted earlier (compared with the uncontrolled stationary case), thus exerting little effect on the pressure distribution over the cylinder surface. The lift force is thus suppressed to a value around zero. Similarly, at instant I, the early shedding of a cw-vortex is induced by the clockwise rotation of the cylinder.

The above results reveal that the DRL knows when to exert the right amount of rotary forcing, so that effective control can be conducted by actively adjusting the phase lag between the rotational forcing and the lift (or the transverse displacement).

To reveal control in a transient process, we present the time histories of the cross-flow displacement y^*, the rotational velocity ω^* and the rotational acceleration $\mathrm{d}\omega^*/\mathrm{d}t^{**}$ in figure 16, as well as the pressure and vorticity fields at the selected instants figure 16(*a*)
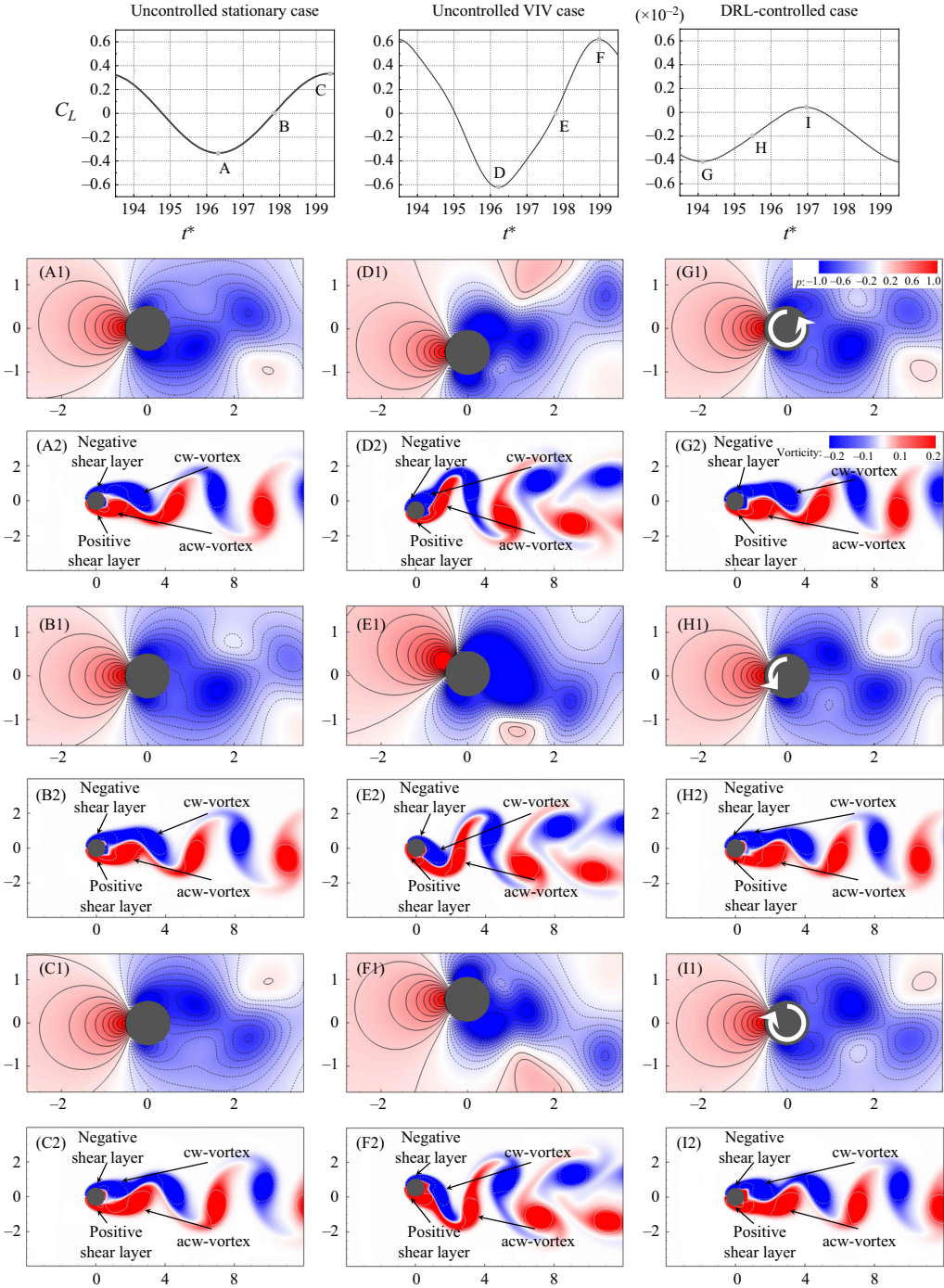
990 A7-22

Figure 15. Temporal variation of lift coefficient. Pressure field and vorticity field at three representative instants are demonstrated. For the DRL-controlled VIV case, arrows are shown to represent the corresponding rotary direction and amplitude, wherein a full circle means $|\omega^*| = 0.4$.
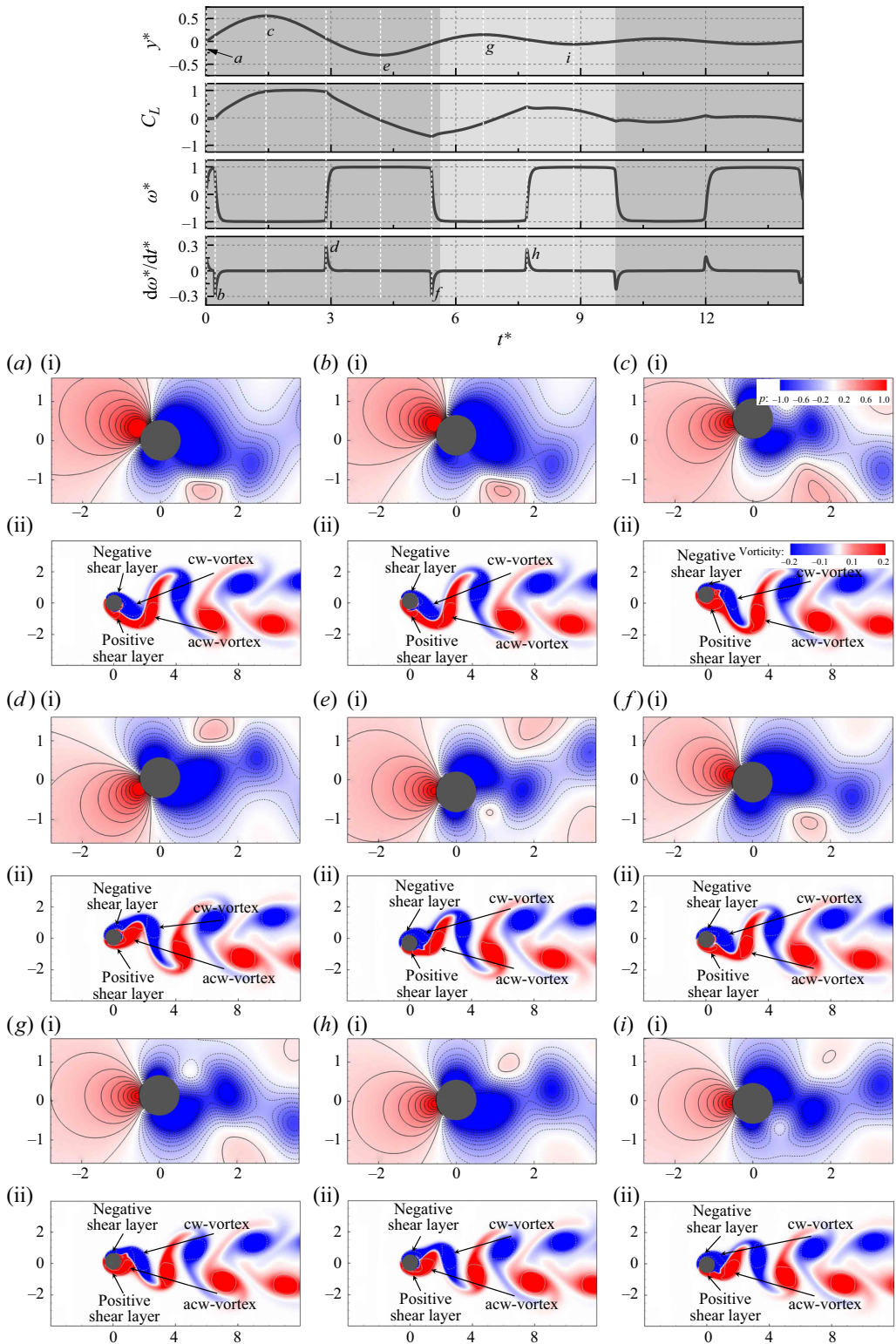
Figure 16. Temporal variation of the cross-flow displacement y^*, the rotational velocity ω^* and its temporal derivative $\mathrm{d}\omega^*/\mathrm{d}t^*$. Pressure field and vorticity field at representative instants are demonstrated.

to figure 16(*i*). The dark and light grey background marks three cycles, all starting from a zero y^* with upwards motion. The instant in figure 16(*a*) is the starting point. The control starts at figure 16(*a*). At instants in figure 16(*a,d,f*) and figure 16(*h*), sharp changes of the control forcing occur, whereas at instants in figure 16(*c,e,g*) and figure 16(*i*), the oscillation displacement reaches its local maximum or minimum. In these cycles, the control forcing maintains mostly at its maximum allowed value, either positive or negative.

It is observed that the sign of ω^* is always opposite to that for y^*, suggesting that the DRL agent adopts a simple yet effective opposition control in this initial stage, by utilizing the Magnus effect, i.e. the anticlockwise rotary motion induces a downward force, and *vice versa*. Therefore, the force induced by rotary motion would counteract the recovery elastic force. As revealed in figure 14, the elastic force is much larger than the lift components. As such, the rotary induced lift force can mitigate the very large recovery elastic force, gradually slowing down the cylinder's transverse motion.

Note that the Magnus effect is only one of the lift components. Another lift component induced by the vortex formation and shedding is nonlinear and complex. The phase difference between the vortex shedding and the cylinder rotation can affect this lift component, which is reflected by the clear phase lags between y^* and ω^*, at instants in figure 16(*b,f*). As having been discussed in § 3.1, this type of phase difference also appears in the quasisteady stage, suggesting that the DRL agent has learned a more complicated strategy than the opposition control.

From instant in figure 16(*a*) to figure 16(*b*), ω^* rapidly increases, which retards the growth of the cw-vortex. At the instant in figure 16(*b*), ω^* reaches its local maximum and then sharply decreases, promoting for the shedding of the cw-vortex and the formation of the following positive shear layer as evidenced at the instant in figure 16(*c*). Similar phenomena can also be found at instants in figure 16(*d,f,h*). Comparing instants in figure 16(*b,d,f,h*), the large low-pressure region gradually becomes smaller and the pressure difference between the upper and lower surfaces is gradually attenuated. This indicates that the DRL-guided control successfully modulates the vortex dynamics. At the instant in figure 16(*c*), the elastic force is at its negative maximum, the clockwise rotation stimulates the growth of the negative shear layer while retarding the growth of the positive shear layer, leading to an upwards lift force so as to counteract downward elastic force. Similar phenomena can be found at instants in figure 16(*e,g,i*).

The time-averaged flow fields displayed in figure 17 further reveal the impact of DRL-guided control. Through comparisons, one can note that the low-velocity far downstream region in the uncontrolled VIV case vanishes after control, making the DRL-controlled flow resemble the uncontrolled stationary case, while the length of the recirculation bubble represented by the zero streamwise velocity contour is increased by 16 %. Prior studies (Rabault *et al.* 2019; Ren *et al.* 2021*a*) have shown that the change in the drag force is strongly related to the length of the recirculation bubble, where an elongated recirculation bubble is usually accompanied by attenuated and delayed vortex shedding. Furthermore, the attenuated transverse flow after control indicates a much weaker source of cross-flow vibration.

3.5. *Dynamic mode decomposition analysis*

To demonstrate changes in the coherent flow structures and flow stability characteristics before and after control, we perform a series of DMD analyses over three flow types, i.e. the uncontrolled stationary case, the uncontrolled VIV case and the DRL-controlled case. Compared with the more popular method of proper orthogonal decomposition,
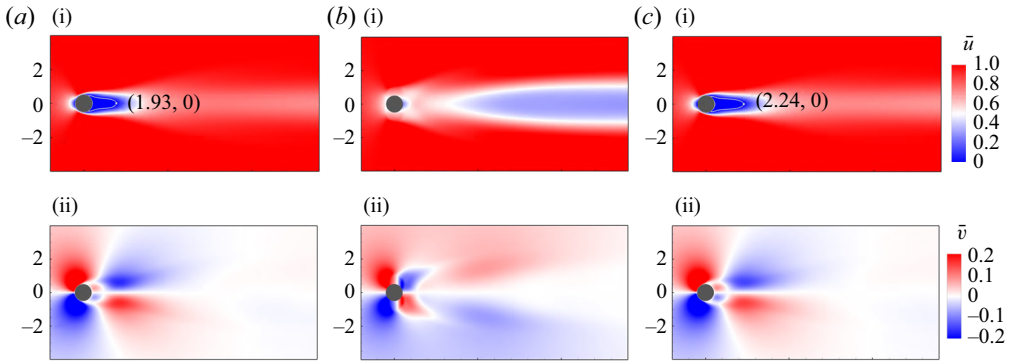
Figure 17. Comparisons of the time-averaged streamwise velocity, transverse velocity and pressure field between (*a*) the uncontrolled stationary cylinder, (*b*) the uncontrolled VIV cylinder and (*c*) the DRL-controlled VIV cylinder. In subpanels (*a* i) and (*c* i), the recirculation bubble is represented by the $\bar{u} = 0$ contour line.

which extracts flow structures based on their energy ranking, DMD provides more clear information related to time or frequencies. With DMD, the unsteady flow field can be reconstructed using limited DMD modes, i.e.

$$\boldsymbol{v}(\boldsymbol{x}, t) = \sum_{i=0}^{L \leqslant N} \mathrm{e}^{\mu_i t} \boldsymbol{\Phi}_i(\boldsymbol{x}), \qquad (3.2)$$

where N is the total number of snapshots. The real and imaginary parts of μ_i represent the growth rate and angular frequency of the ith DMD mode $\boldsymbol{\Phi}_i(\boldsymbol{x})$, respectively. Note that we have included in $\boldsymbol{\Phi}_i(\boldsymbol{x})$ the corresponding amplitude, which reflects the influence of the initial data on the DMD mode. Detailed information about the DMD algorithm can be found in Schmid (2010) and Wang *et al.* (2017*a*).

In this study, we extract a series of flow fields with a time interval of $0.02T_0$ within a duration of one vortex shedding period, and the numbers of snapshots used in the three flow systems are 307, 275 and 279, respectively. In our tests, these numbers of snapshots are sufficient to obtain converged DMD results. When conducting the DMD analysis, we select part of the whole domain ranging from $x^* = 0.50$ to $x^* = 14$ and from $y^* = -4$ to $y^* = 4$. It should be noted that the zone that contains the moving cylinder is removed, so as to avoid the effect of moving structure on the DMD analysis. As can be seen from the DMD modes of figure 19, this subdomain is large enough to cover most flow features both around the cylinder and in the wake.

Sorting the DMD modes by their frequencies, we obtain the frequencies (non-dimensionalized as Strouhal numbers) and growth rates of all DMD modes. For the sake of brevity, we list the data for the first four DMD modes in table 3. Here, the Strouhal number of Mode 1 is consistent with the lift frequency extracted in figure 13, and is the dominating fluctuation frequency, i.e. the vortex shedding frequency. The frequencies of Mode 3 and Mode 5 correspond to two and three times the vortex shedding frequency, respectively. The first four growth rates for all three cases are of an order of 10^{-4} or smaller, implying that the corresponding modes are neutrally stable.

In addition to table 3, we plot the growth rate against the Strouhal number for all DMD modes in figure 18. The negative-frequency points are the adjoint modes, making the data symmetrical along the vertical axis. For the uncontrolled stationary case, only a few modes including Mode 1 have positive growth rates. For the uncontrolled VIV case, approximately one-third of the DMD modes have positive growth rates. However, after

990 A7-26

Configuration	DMD mode	Strouhal number	Growth rate ($\times 10^{-5}$)
Uncontrolled stationary case	0	0	-2.491
	1	0.163	0.290
	3	0.327	-2.258
	5	0.490	-4.025
Uncontrolled VIV case	0	0	-1.774
	1	0.182	0.373
	3	0.364	-1.024
	5	0.546	8.453
DRL-controlled VIV case	0	0	-12.1278
	1	0.180	-21.415
	3	0.359	-10.916
	5	0.539	-8.982

Table 3. Comparisons of the amplitude, Strouhal number and growth rate of the first four DMD modes.
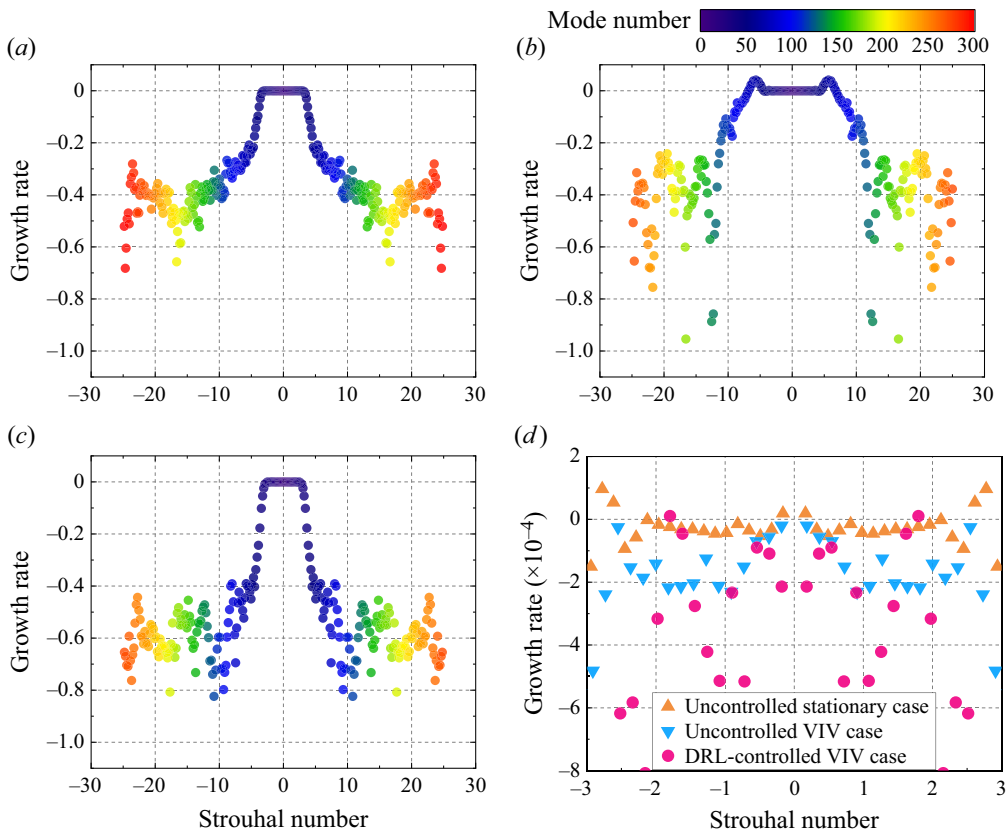
Figure 18. Scatter plots of the Strouhal number versus growth rate. (*a–c*) Results for the uncontrolled stationary case, the uncontrolled VIV case and the DRL-controlled VIV case, respectively. (*d*) A comparison between the three cases involving only the low-frequency modes. The colours determine the mode number. Negative Strouhal numbers represent the corresponding adjoint modes.
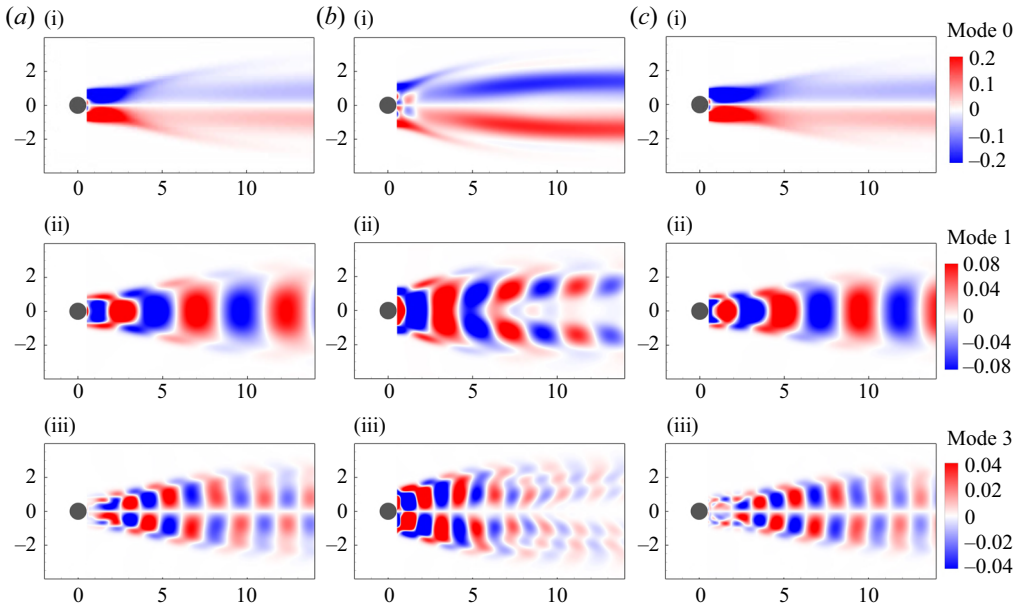
Figure 19. The three dominant DMD modes calculated from the uncontrolled stationary case, the uncontrolled VIV case and the DRL-controlled VIV case.

control, no positive growth rate is found, proving that the DRL-guided control does enhance the flow stability. Moreover, the two peaks in the uncontrolled case suggest that the flow instability can be triggered by either low-frequency or high-frequency perturbations. This potential source of instability is eliminated by the DRL-guided control.

Figure 19 shows the dominant DMD modes of the uncontrolled stationary case, the uncontrolled VIV case and the DRL-controlled VIV case. Among these modes, Mode 0 is the mean mode that corresponds to the time-averaged vorticity field, where one can see a pair of symmetrical shear layers. After control, Mode 1 shows a much narrower shear layer compared with figure 19(*b* i), as well as an elongated shear layer compared with figure 19(*a* i), consistent with the time-averaged variables illustrated in figure 17. Mode 1 is connected to the lift fluctuation and is thus strongly associated with VIV suppression. In Mode 1, one can see in-line alternating vortices behind the cylinder in figure 19(*a* ii) and figure 19(*c* ii), which are much narrower and weaker than those in the uncontrolled VIV case. Vortices whose centre are off the centrelines appear in the uncontrolled VIV case, suggesting that their momentum along the cross-flow direction is larger, gained when interacting with the cylinder. This phenomenon is evidence of larger lift fluctuations of the cylinder exerted by the surrounding fluid. These flow features explain the changes after the VIV is suppressed. In Mode 3, one can see pairs of counter-rotating vortices whose centres are symmetrical along the centreline of the wake flow, which is the main cause of drag fluctuation. Comparing figures 19(*b* iii) and 19(*c* iii), the attenuated near-wall flow structures after control explain the remarkably mitigated drag fluctuation.

Therefore, the DMD analysis explains the remarkable enhancement of flow stability achieved with DRL-guided control. Furthermore, changes in the coherent structures show that DRL-guided control mediates the flow to a state resembling that of the stationary case while generating an elongated recirculation bubble as well as weak small-scale flow structures.
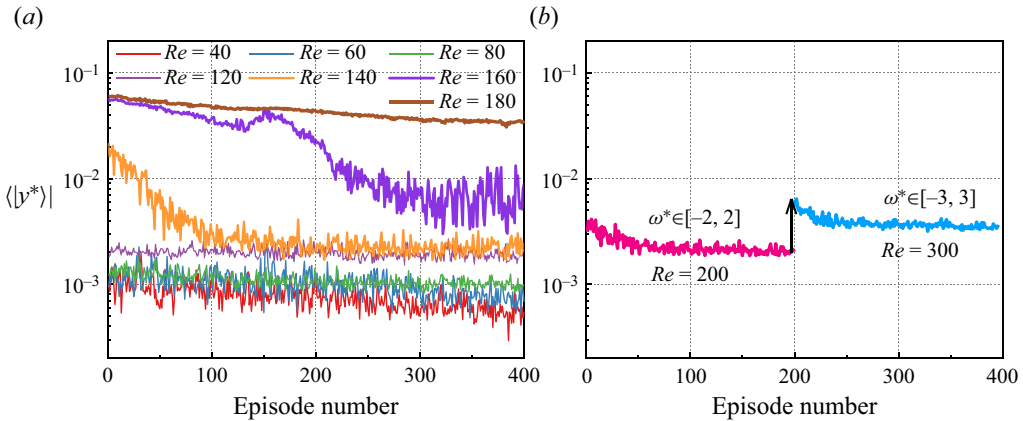
990 A7-28

Figure 20. Transfer learning for cases at Reynolds numbers ranging from 40 to 300. In panels (*a*,*b*), the first episode starts from the learned strategy at $Re = 100$. For the $Re = 300$ case in (*b*), the DRL agent inherits the converged strategy from the $Re = 200$ condition.

3.6. *Robustness at different Reynolds numbers*

In this section, we test the robustness of the DRL-guided control strategy within a Reynolds number range of [40, 300]. Instead of directly applying the learned strategy to the off-design conditions, we introduce the idea of transfer learning to further improve performance. We retain almost all of the same set-ups as in the $Re = 100$ case but reinitialize the 'actor' and 'critic' neural networks with well-trained model parameters as in the control strategy of § 3.2. The training then continues until reaching convergence. The converging trends for various Reynolds numbers are shown in figure 20. Note that in figure 20(*a*), the actuation range of $\omega^* \in [-1, 1]$ is retained. However, due to the saturated forcing occurring around $Re = 180$, we enlarge the actuation range and choose an interval of $[-2, 2]$ for the $Re = 200$ case. For the $Re = 300$ case, we choose a larger actuation interval of $[-3, 3]$ and start the transfer learning via inheriting the strategy learned at $Re = 200$. Instantaneous vorticity fields, as well as the vibration responses and AFC forcing at selected Reynolds numbers, are organized into figure 21.

At $Re = 40$, the VIV is under the subcritical Re condition (Mittal & Singh 2005). When the DRL-guided control is turned on at $t^* = 0$, the flow becomes steady and only slight rotary forcing is needed to sustain the suppressed VIV. For $Re \leqslant 120$, the starting point of the transfer learning shows that the learned strategy at $Re = 100$ is already good, capable of suppressing the VIV to a level around 0.001 times the cylinder diameter. In the following transfer learning process, the curve only shows a very slow decreasing trend, indicating that there is almost no more space for the DRL agent to improve its performance. As Re is further increased to 140, a larger VIV amplitude is noted in the beginning, suggesting that the learned strategy cannot be directly applied under this condition. However, with transfer learning, the control performance can be further improved to a level of roughly $\langle |y^*| \rangle = 0.003$. This case vigorously verifies the advantage of transfer learning.

As Re is further increased to 160, a similar trend is observed as in the case of $Re = 140$, while showing a higher starting point and less effective control performance. This change suggests that the increased Re starts to make the VIV control more difficult. The DRL-guided control experiences a failure when Re is increased to 180, and the performance is only slightly improved as the learning proceeds. The finally converged strategy shows unsuccessful VIV suppression, as depicted in figure 21(*c* iv). Meanwhile,
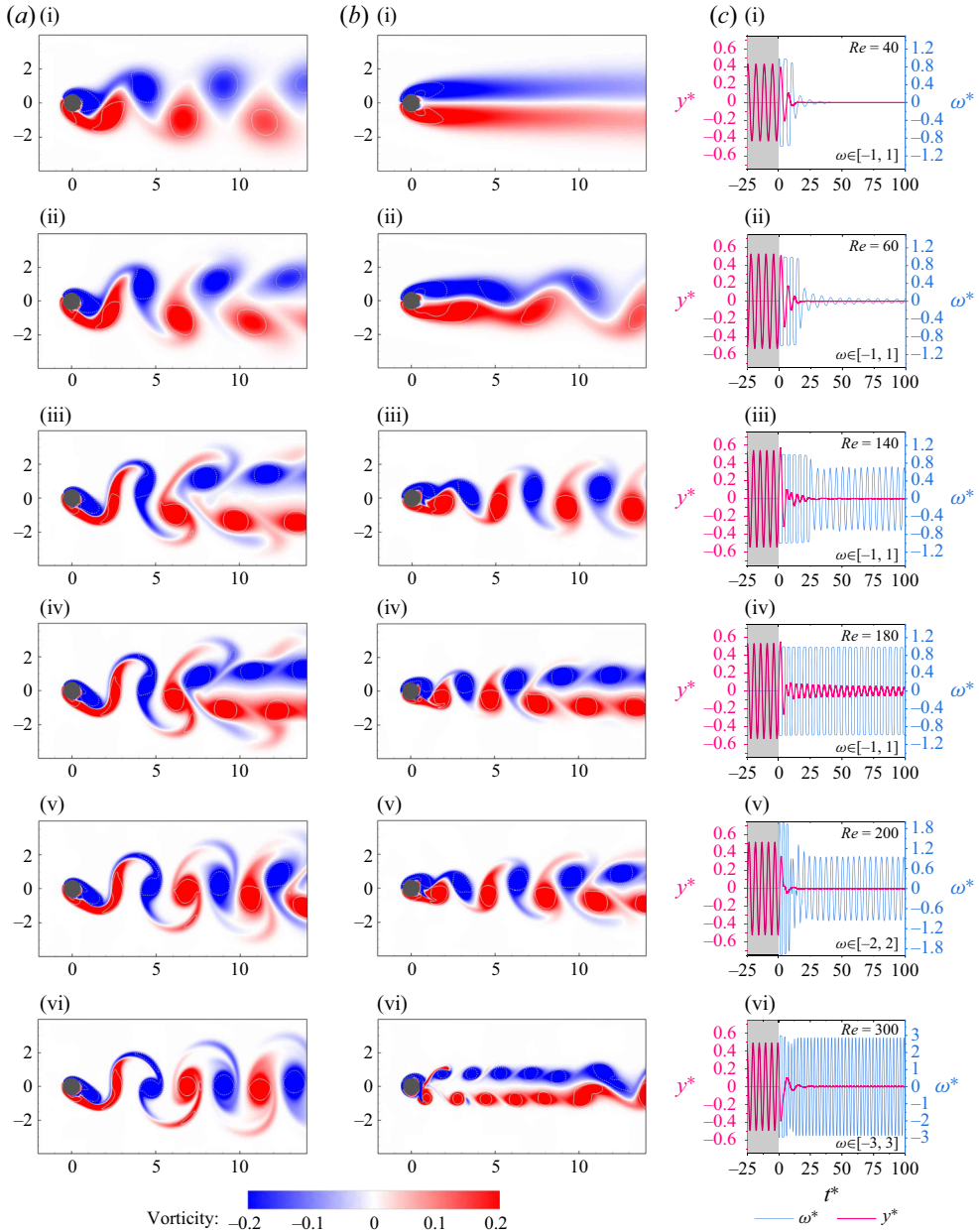
Figure 21. The flow fields, vibration response and rotational forcing from DRL-guided control at selected Reynolds numbers. Instantaneous snapshots of (*a*) the uncontrolled VIV case and (*b*) the DRL-controlled VIV case. (*c*) Temporal variations of the vibration amplitude before and after the control is turned on, as well as the AFC forcing. In panels (*a,b*), instants are selected when the cylinder is crossing its equilibrium position with upward velocity.

we also note that the actuators have already reached saturation at this point, explaining why the DRL can hardly improve the performance during transfer learning. Therefore, we expand the actuation range in cases when a higher Reynolds number is involved.

990 A7-30

From the perspective of the flow field, one can also observe that in figure 21(*a* ii,iii,*b* ii,iii, *c* ii,iii), a transition from the C(2S) vortex shedding pattern to the 2S pattern occurs. The phenomenon that the DRL-controlled flow field shows a larger centre-to-centre distance between two adjacent and like-signed vortices suggests that the vortex shedding period is enlarged by the rotary control. The less successful case, the $Re = 180$ case, does not show a pattern transition but the distance between the parallel-moving vortices is narrowed. At $Re = 200$ and $Re = 300$, a dramatic transition of the alternating 2S vortex shedding pattern to the C(2S) pattern is observed, which can be explained by the high-frequency rotary forcing in figures 21(*c* v) and 21(*c* vi). This frequency shift in the rotary forcing reveals a distinct mechanism of VIV control, which, as concluded by Du & Sun (2015), is due to the fact that the vortex shedding frequency is attracted to the forcing frequency. In other words, for the cases when $Re = 200$ and $Re = 300$, the control strategy discovered by the DRL actually utilizes the 'lock-on' effect. A similar vortex shedding pattern of the $Re = 300$ case can also be found in the work of Protas & Wesfreid (2002), where harmonic rotary control with a frequency of two times the vortex shedding frequency was utilized. In fact, this frequency ratio matches the forcing frequency over the vibration frequency in figure 21(*c* vi) well.

It should be noted that in this section, while performing the transfer learning, we do not follow the prior practice in § 3.1, in which a group of three or five trials was conducted so as to avoid the possible effects of training randomness, and thus we do not intend to imply that the strategies showcased in figure 21 are the only optimal ones. Our main objective is to demonstrate the robustness of DRL-guided control in terms of different Re, and these cases indeed prove that the DRL-guided control is qualified. In addition, we understand that the VIV at $Re = 300$ involves the 3-D effect. Here we still choose the 2-D simulations so as to keep consistent with the other cases. In the future, 3-D and even the fully turbulent conditions would be thoroughly studied. Note that in higher-Reynolds-number conditions, our choice of state space, action and reward for VIV control would not be affected, thus we anticipate that the model-free and data-driven DRL could possibly be still effective.

3.7. *Robustness with a perturbed upstream flow*

In the Introduction, we mentioned that current studies related to DRL-guided AFC lacks sufficient tests and confirmations with perturbed environments, which is the more realistic situation. In this section, we further verify the robustness of the DRL-guided control strategy in a perturbed upstream flow. We superpose a harmonic perturbation to the uniform velocity U_0 at the inlet (Konstantinidis & Bouris 2009; Zhao *et al.* 2013; Ma *et al.* 2021), and thus the inlet velocity becomes

$$U_0' = U_0(1 + \epsilon_p \sin(2\pi f_p t)), \tag{3.3}$$

where ϵ_p is the perturbation amplitude and f_p the perturbation frequency. In the study of Konstantinidis & Bouris (2009), both the harmonic (as (3.3)) and the non-harmonic upstream perturbations induced complicated flow patterns and hydrodynamic force responses. In a recent study that focused on the effect of a sinusoidal streamwise gust on the vortex-induced force on an oscillating rectangular cylinder using both wind tunnel experiments and a CFD approach (Ma *et al.* 2021), it was found that the flow structure was sensitive to the upstream flow condition. For demonstration purposes, we fix the perturbation amplitude to quite a challenging level, $\epsilon_p = 0.5$ (a value of 0.325 is used in Konstantinidis & Bouris (2009) and Zhao *et al.* (2013)). In this sense, the robustness is essentially different from that of § 3.4, due to the fact that the Reynolds number of the

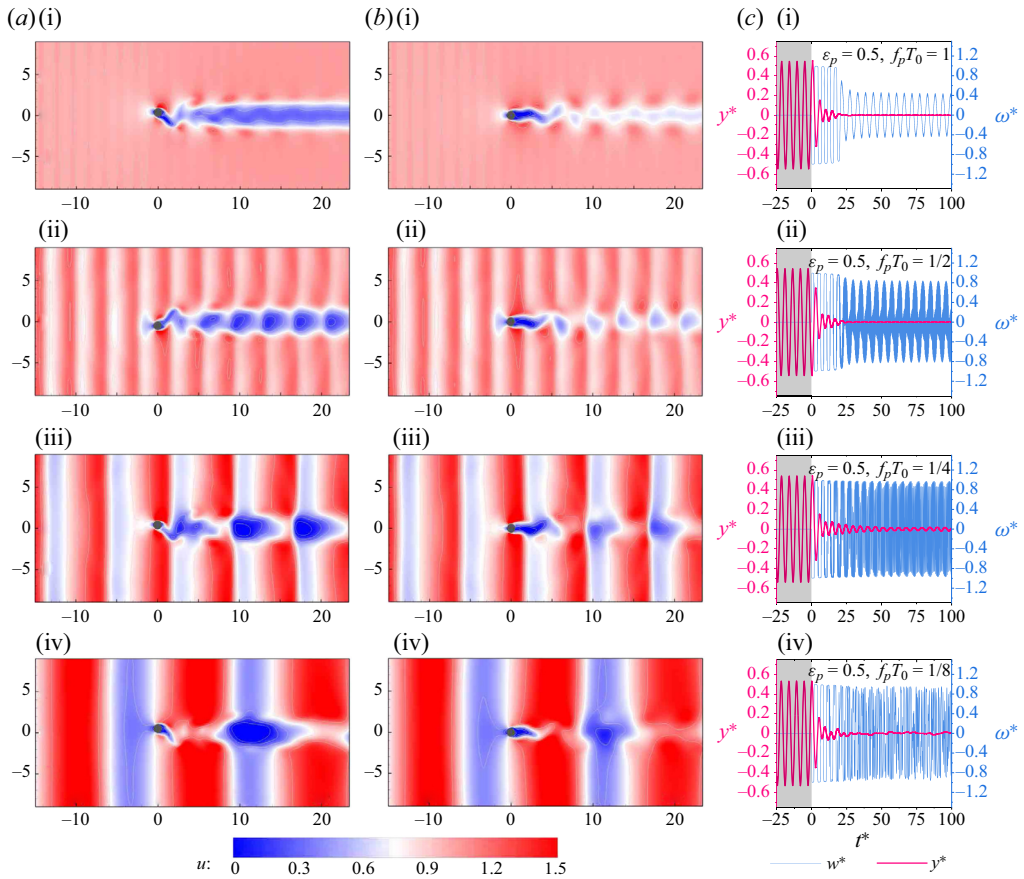almost reaches saturation. When the perturbation frequency approaches the VIV resonance frequency (see table 1), we see the most challenging case with $f_p T_0$ being 1/4, as shown in figure 22(*c* iii). In this case, the asynchronous perturbation and vibration cause difficulties for the DRL in making decisions and induce high-frequency and saturated actuations.

The onset of complicated frequency features reflects the significant impact of the perturbed upstream flow on the flow and vibration response, whereas the DRL agent can still successfully realize the vibration suppression objective. Therefore, the above robustness tests prove that the DRL-guided control is not only efficient under uniform upstream flow conditions, but is also successful when the environment is experiencing a certain level of perturbation.

In our prior study (Ren *et al.* 2021*b*) targeted at eliminating the wake signatures of a circular cylinder, it was observed that the attenuated wake signature is closely related to the hydrodynamic forces exerted on the cylinder, as well as the vibration characteristics. The findings here, on the opposite side, show that the suppression of vibration also facilitates the attenuation of wake signatures, even with strong upstream perturbations.

4. Summary and conclusion

In the present work, we have demonstrated the application of DRL-guided AFC to the classical VIV problem. Through self-rotary control guided by the DRL agent in a real-time and self-adaptive manner, the VIV was eventually suppressed, with a reduction in the VIV amplitude of 99.6 %. Through case studies and detailed analysis, we have drawn the following conclusions.

   (i) Under the lock-in condition, the DRL-guided AFC was found to be capable of suppressing the VIV with relatively small actuations and energy input, indicating that the control method is effective and efficient.

  (ii) With all possible combinations of sensory-motor cues being tested, the combination of displacement, velocity and acceleration sensors showed the best performance, while the acceleration sensor alone played the most significant individual role. This finding was confirmed by sensitivity analysis. Three distinct search paths during training were observed, intuitively illustrating the logical approach of the AI and showing that the DRL exploration process is interpretable.

 (iii) The DMD analysis showed that the DRL control remarkably enhanced the flow stability and transformed the flow field into a state resembling an uncontrolled stationary one, while leading to an elongated recirculation bubble and mitigating small-scale flow structures. The DMD analysis offered solid evidence that the DRL exploitation was physically interpretable.

 (iv) With various Reynolds numbers and a perturbed upstream flow, DRL-guided control could still successfully suppress flow-induced vibration, indicating that the DRL-guided control is robust. Moreover, the idea of transfer learning has also been proven to be a powerful tool with which to extend the capability of DRL in complicated tasks.

Therefore, we can conclude that the present DRL-guided VIV control method is effective, efficient, interpretable and robust. Although the physical model of VIV is relatively simple, the case we have demonstrated in this study can provide insight for future studies concerning more complicated flow-related systems, especially the control in fully turbulent regime. The rich physics and practical significance of turbulence control make us firmly believe that the DRL-guided turbulence control is a very valuable field to explore. Moreover, the DRL-guided control method with interpretations from the physical aspect

can help us understand how AI identifies an optimal controller and the physical rationale behind it.

Declaration of interests. The authors report no conflict of interest.

Author ORCIDs.

Feng Ren https://orcid.org/0000-0001-7523-0239;

Chenglei Wang https://orcid.org/0000-0003-0771-8667;

Jian Song https://orcid.org/0009-0008-6533-6679;

Hui Tang https://orcid.org/0000-0002-6774-507X.

Appendix A. Response to a specified sinusoidal signal through the controller

Because the learned controller is established as the form of a neural network, it is less straightforward when describing its features. In this supplementary section, we investigate how the controller learned at the 1996th episode reacts when simple sinusoidal signals are specified as the input. We consider a non-normalized vibration amplitude formulated as

$$y = y_A D_0 \sin\left(\frac{2\pi St}{T_0} t\right), \tag{A1}$$

where y_A is the specified dimensionless vibration amplitude, and St represents the dimensionless vibration frequency. The dimensionless state variables y^*, $\dot{y}^*$, $\ddot{y}^*$ can then be deduced and function as the input of the DRL controller, i.e.

$$\left.\begin{array}{l} y^* = y_A \sin\left(\dfrac{2\pi St}{T_0} t\right), \\[2mm] \dot{y}^* = 2\pi St y_A \cos\left(\dfrac{2\pi St}{T_0} t\right), \\[2mm] \ddot{y}^* = -4\pi^2 St^2 y_A \sin\left(\dfrac{2\pi St}{T_0} t\right). \end{array}\right\} \tag{A2}$$

A systematic parametric study concerning values of y_A ranging from 0.001 to 1 and St ranging from 0.1 to 0.3 provides an overall view of the response of the controller. These ranges can sufficiently cover all possible $y_A - St$ pairs during the training and deterministic control process. The results are organized into figure 23.

From the phase diagrams, one can draw the following conclusions.

(i) A marginal actuation bias exists, especially when the vibration amplitude is small. However, compared with the actuation amplitude, this bias is still very small. In the aforementioned discussions in § 3.3, we did not find clear evidence that this bias plays a significant role in the control strategy. Therefore, we believe it is just a consequence of an imperfectly trained controller.

(ii) The action amplitude reaches the saturation state easily. This fact also coincides with the saturated actuation in the initial stage of control, as shown in figure 9. As having been discussed in § 3.3, this control ensures sufficiently large forcing strength in
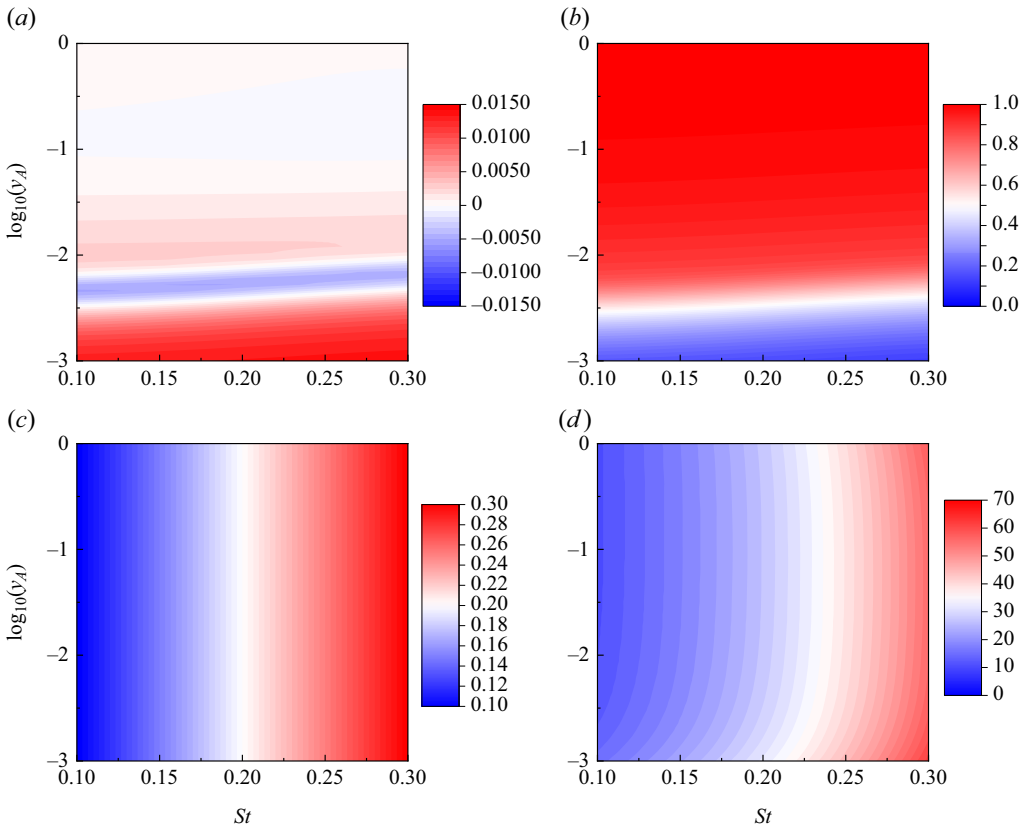
Figure 23. The AFC response with specified sinusoidally varying sensor signals as the control input. (*a*) The mean rotational forcing, (*b*) the fluctuation amplitude of the rotational forcing, (*c*) the frequency of the rotational forcing, and (*d*) the phase lag between the rotational forcing and the transverse displacement. (*a*) Average ω^*. (*b*) Amplitude of ω^*. (*c*) Frequency of ω^*. (*d*) $\omega^* - \ddot{y}^*$ phase lag (deg.).

Case	Network architecture (actor, critic)	Learning rate (actor, critic)	Discount factor
I	$64 \times 64, 64 \times 64$	$3 \times 10^{-4}, 4 \times 10^{-4}$	0.97
II	$32 \times 32, 32 \times 32$	$3 \times 10^{-4}, 4 \times 10^{-4}$	0.97
III	$50 \times 50, 50 \times 50$	$1 \times 10^{-4}, 1 \times 10^{-4}$	0.97
IV	$50 \times 50, 50 \times 50$	$1 \times 10^{-4}, 1 \times 10^{-4}$	0.97
V	$50 \times 50, 50 \times 50$	$3 \times 10^{-4}, 4 \times 10^{-4}$	0.96
VI	$50 \times 50, 50 \times 50$	$3 \times 10^{-4}, 4 \times 10^{-4}$	0.98

Table 4. Extra hyperparameters used to examine the effect of hyperparameters on the DRL-guided control.

the initial stage, which is significant in rapidly controlling the vibration to a new quasisteady state.

(iii) The actuation frequency is always identical to the vibration frequency. This once again confirms the finding in § 3.2 that the DRL agent does not focus on finding optimal frequencies that utilize the 'lock-on' effect. The DRL agent is actually adapting to the system frequency.
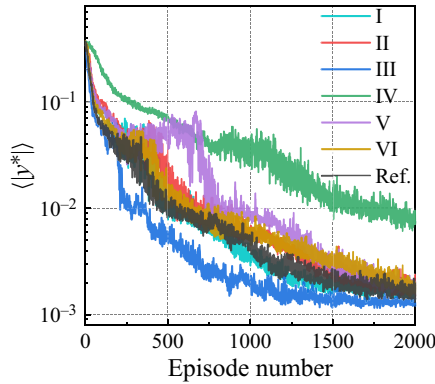
Figure 24. Learning process with six sets of hyperparameters, labelled from I to VI. Here the curve with hyperparameters used in figure 4(*c*) functions as the reference.
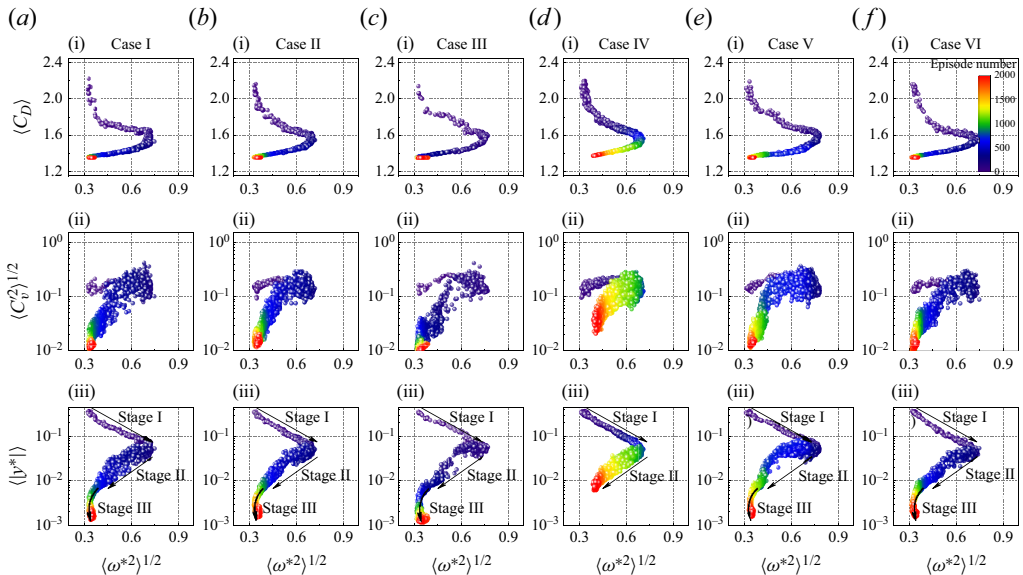

Figure 25. Trajectories of the (i) mean drag (ii) r.m.s. of the vortex force and (iii) absolute value of transverse displacement against the r.m.s. of the rotational velocity.

(iv) The $\omega^* - \ddot{y}^*$ phase difference reveals somewhat complicated features. It is quite sensitive to the vibration frequency, while only being slightly affected by y_A, except when y_A is small. As shown in figure 4(*c*), in the training, especially in Stage III, the vibration amplitude lies right in this range. Moreover, in Stage III, the DRL agent is mainly seeking the optimal phase between the actuation and kinematic variables. The fact that the phase lag is affected by both the vibration frequency and vibration amplitude makes the phase-adjusting task quite challenging. This explains why the DRL agent requires much effort in Stage III.

Appendix B. Effect of hyperparameters

As a general view in the field of machine learning, hyperparameters could affect the final output. To examine this effect, we perform six extra learnings that utilize different combinations of hyperparameters. The set-ups and results are shown in table 4, figure 24 and figure 25. We choose some significant hyperparameters that could affect the training process, i.e. the neural network structure, the learning rate and the discount factor.

REFERENCES

BERGMANN, M., CORDIER, L. & BRANCHER, J.-P. 2005 Optimal rotary control of the cylinder wake using proper orthogonal decomposition reduced-order model. *Phys. Fluids* **17** (9), 097101.

BOURGUET, R. & LO JACONO, D. 2014 Flow-induced vibrations of a rotating cylinder. *J. Fluid Mech.* **740**, 342–380.

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.

BRUNTON, S.L., NOACK, B.R. & KOUMOUTSAKOS, P. 2020 Machine learning for fluid mechanics. *Annu. Rev. Fluid Mech.* **52**, 477–508.

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** (4), 047118.

CHOI, H., JEON, W.-P. & KIM, J. 2008 Control of flow over a bluff body. *Annu. Rev. Fluid Mech.* **40** (1), 113–139.

CHOI, J.-II., OBEROI, R.C., EDWARDS, J.R. & ROSATI, J.A. 2007 An immersed boundary method for complex incompressible flows. *J. Comput. Phys.* **224** (2), 757–784.

CHOU, P.-W., MATURANA, D. & SCHERER, S. 2017 Improving stochastic policy gradients in continuous control with deep reinforcement learning using the beta distribution. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70* (ed. D. Precup & Y.W. Teh), 6–11 August, Sydney, NSW, Australia, pp. 834–843. JMLR.

COLLIS, S.S., JOSLIN, R.D., SEIFERT, A. & THEOFILIS, V. 2004 Issues in active flow control: theory, control, simulation, and experiment. *Prog. Aerosp. Sci.* **40** (4–5), 237–289.

D'HUMIÈRES, D., GINZBURG, I., KRAFCZYK, M., LALLEMAND, P. & LUO, L.S. 2002 Multiple-relaxation-time lattice Boltzmann models in three dimensions. *Phil. Trans. R. Soc. Lond.* A **360** (1792), 437–451.

DU, L. & SUN, X. 2015 Suppression of vortex-induced vibration using the rotary oscillation of a cylinder. *Phys. Fluids* **27** (2), 023603.

FAKHARI, A. & LEE, T. 2014 Finite-difference lattice Boltzmann method with a block-structured adaptive-mesh-refinement technique. *Phys. Rev.* E **89** (3), 033310.

FAN, D., YANG, L., WANG, Z., TRIANTAFYLLOU, M.S. & KARNIADAKIS, G.EM. 2020 Reinforcement learning for bluff body active flow control in experiments and simulations. *Proc. Natl Acad. Sci. USA* **117** (42), 26091–26098.

FENG, Z.-G. & MICHAELIDES, E.E. 2009 Robust treatment of no-slip boundary condition and velocity updating for the lattice-Boltzmann simulation of particulate flows. *Comput. Fluids* **38** (2), 370–381.

FERNEX, D., NOACK, B.R. & SEMAAN, R. 2021 Cluster-based network modeling-from snapshots to complex dynamical systems. *Sci. Adv.* **7** (25), eabf5006.

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.

GOTTLIEB, S. & SHU, C.W. 1996 Total variation diminishing Runge–Kutta schemes. *Maths Comput.* **67** (221), 73–85.

GOVARDHAN, R. & WILLIAMSON, C.H.K. 2000 Modes of vortex formation and frequency response of a freely vibrating cylinder. *J. Fluid Mech.* **420**, 85–130.

HASHEMINEJAD, S.M., RABIEE, A.H., JARRAHI, M. & MARKAZI, A.H.D. 2014 Active vortex-induced vibration control of a circular cylinder at low Reynolds numbers using an adaptive fuzzy sliding mode controller. *J. Fluids Struct.* **50**, 49–65.

HE, X. & LUO, L.S. 1997 Lattice Boltzmann model for the incompressible Navier–Stokes equation. *J. Stat. Phys.* **88** (3–4), 927–944.

HONG, K.-S. & SHAH, U.H. 2018 Vortex-induced vibrations and control of marine risers: a review. *Ocean Engng* **152**, 300–315.

KIM, J. & BEWLEY, T.R. 2007 A linear systems approach to flow control. *Annu. Rev. Fluid Mech.* **39** (1), 383–417.

KINGMA, D.P. & BA, J. 2014 Adam: a method for stochastic optimization. arXiv:1412.6980.

KONSTANTINIDIS, E. & BOURIS, D. 2009 Effect of nonharmonic forcing on bluff-body vortex dynamics. *Phys. Rev.* E **79**, 045303.

LADD, A.J.C. 1994 Numerical simulations of particulate suspensions via a discretized Boltzmann equation. Part 1. Theoretical foundation. *J. Fluid Mech.* **271** (25), 285–309.

LI, H., FERNEX, D., SEMAAN, R., TAN, J., MORZYŃSKI, M. & NOACK, B.R. 2021 Cluster-based network model. *J. Fluid Mech.* **906**, A21.

LI, J. & ZHANG, M. 2022 Reinforcement-learning-based control of confined cylinder wakes with stability analyses. *J. Fluid Mech.* **932**, A44.

LOU, M., WANG, Y., QIAN, G. & DONG, W. 2021 Investigation on the vortex-induced vibration active control of the riser in the "lock-in" region based on adaptive fuzzy sliding mode theory. *Ocean Engng* **238**, 109697.

MA, R., ZHOU, Q., WANG, P., YANG, Y., LI, M. & CAO, S. 2021 Effects of sinusoidal streamwise gust on the vortex-induced force on an oscillating 5:1 rectangular cylinder. *J. Wind Engng Ind. Aerodyn.* **213**, 104642.

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** (10), 103604.

MITTAL, S. & SINGH, S. 2005 Vortex-induced vibrations at subcritical *Re*. *J. Fluid Mech.* **534**, 185–194.

MNIH, V. *et al.* 2019 Human-level control through deep reinforcement learning. *Nature* **518**, 529–533.

MUDDADA, S. & PATNAIK, B.S.V. 2017 Active flow control of vortex induced vibrations of a circular cylinder subjected to non-harmonic forcing. *Ocean Engng* **142**, 62–77.

PARIS, R., BENEDDINE, S. & DANDOIS, J. 2021 Robust flow control and optimal sensor placement using deep reinforcement learning. *J. Fluid Mech.* **913**, A25.

PROTAS, B. & WESFREID, J.E. 2002 Drag force in the open-loop control of the cylinder wake in the laminar regime. *Phys. Fluids* **14** (2), 810–826.

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.

RABAULT, J., REN, F., ZHANG, W., TANG, H. & XU, H. 2020 Deep reinforcement learning in fluid mechanics: a promising method for both active flow control and shape optimization. *J. Hydrodyn.* **32** (2), 234–246.

RAIBAUDO, C., ZHONG, P., NOACK, B.R. & MARTINUZZI, R.J. 2020 Machine learning strategies applied to the control of a fluidic pinball. *Phys. Fluids* **32** (1), 015108.

RAMACHANDRAN, P., ZOPH, B. & LE, Q.V. 2017 Swish: a self-gated activation function. arXiv:1710.05941.

REDDY, G., CELANI, A., SEJNOWSKI, T.J. & VERGASSOLA, M. 2016 Learning to soar in turbulent environments. *Proc. Natl Acad. Sci. USA* **113** (33), E4877.

REN, F., HU, H.-B. & TANG, H. 2020 Active flow control using machine learning: a brief review. *J. Hydrodyn.* **32** (2), 247–253.

REN, F., RABAULT, J. & TANG, H. 2021*a* Applying deep reinforcement learning to active flow control in weakly turbulent conditions. *Phys. Fluids* **33** (3), 037121.

REN, F., SONG, B., ZHANG, Y. & HU, H. 2018 A GPU-accelerated solver for turbulent flow and scalar transport based on the lattice Boltzmann method. *Comput. Fluids* **173**, 29–36.

REN, F., WANG, C. & TANG, H. 2019 Active control of vortex-induced vibration of a circular cylinder using machine learning. *Phys. Fluids* **31** (9), 093601.

REN, F., WANG, C. & TANG, H. 2021*b* Bluff body uses deep reinforcement learning trained active flow control to achieve hydrodynamic stealth. *Phys. Fluids* **33** (9), 093602.

ROWLEY, C.W. & DAWSON, S.T.M. 2017 Model reduction for flow analysis and control. *Annu. Rev. Fluid Mech.* **49** (1), 387–417.

RUSSELL, D. & WANG, Z.J. 2003 A Cartesian grid method for modeling multiple moving objects in 2D incompressible viscous flow. *J. Comput. Phys.* **191** (1), 177–205.

SALTELLI, A. 2002 Making best use of model evaluations to compute sensitivity indices. *Comput. Phys. Commun.* **145**, 280–297.

SCHMID, P.J. 2010 Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.* **656**, 5–28.

SCHULMAN, J., WOLSKI, F., DHARIWAL, P., RADFORD, A. & KLIMOV, O. 2017 Proximal policy optimization algorithms. arXiv:1707.06347.

SILVER, D., HUANG, A., MADDISON, C.J., GUEZ, A., SIFRE, L., VAN DEN DRIESSCHE, G., SCHRITTWIESER, J., ANTONOGLOU, I., PANNEERSHELVAM, V. & LANCTOT, M. 2016 Mastering the game of go with deep neural networks and tree search. *Nature* **529**, 484–489.

SILVER, D., SCHRITTWIESER, J., SIMONYAN, K., ANTONOGLOU, I. & HASSABIS, D. 2017 Mastering the game of go without human knowledge. *Nature* **550** (7676), 354–359.

SOBOL, I.M. 2014 Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Maths Comput. Simul.* **55**, 271–280.

SONODA, T., LIU, Z., ITOH, T. & HASEGAWA, Y. 2023 Reinforcement learning of control strategies for reducing skin friction drag in a fully developed turbulent channel flow. *J. Fluid Mech.* **960**, A30.

SUZUKI, K. & INAMURO, T. 2011 Effect of internal mass in the simulation of a moving body by the immersed boundary method. *Comput. Fluids* **49** (1), 173–187.

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** (5), 053605.

VERMA, S., NOVATI, G. & KOUMOUTSAKOS, P. 2018 Efficient collective swimming by harnessing vortices through deep reinforcement learning. *Proc. Natl Acad. Sci. USA* **115** (23), 5849–5854.

VICENTE-LUDLAM, D., BARRERO-GIL, A. & VELAZQUEZ, A. 2018 Flow-induced vibration control of a circular cylinder using rotational oscillation feedback. *J. Fluid Mech.* **847**, 93–118.

WAN, H. & PATNAIK, S.S. 2016 Suppression of vortex-induced vibration of a circular cylinder using thermal effects. *Phys. Fluids* **28** (12), 123603.

WANG, C., TANG, H., DUAN, F. & YU, S.C.M. 2016*a* Control of wakes and vortex-induced vibrations of a single circular cylinder using synthetic jets. *J. Fluids Struct.* **60**, 160–179.

WANG, C., TANG, H., YU, S.C.M. & DUAN, F. 2016*b* Active control of vortex-induced vibrations of a circular cylinder using windward-suction-leeward-blowing actuation. *Phys. Fluids* **28** (5), 053601.

WANG, C., TANG, H., YU, S.C.M. & DUAN, F. 2017*a* Control of vortex-induced vibration using a pair of synthetic jets: influence of active lock-on. *Phys. Fluids* **29** (8), 083602.

WANG, C., TANG, H., YU, S.C.M. & DUAN, F. 2017*b* Lock-on of vortex shedding to a pair of synthetic jets with phase difference. *Phys. Rev. Fluids* **2** (10), 104701.

WANG, Z., FAN, J. & LUO, K. 2008 Combined multi-direct forcing and immersed boundary method for simulating flows with moving particles. *Intl J. Multiphase Flow* **34** (3), 283–302.

WILLIAMSON, C.H.K. & GOVARDHAN, R. 2004 Vortex-induced vibrations. *Annu. Rev. Fluid Mech.* **36** (1), 413–455.

WILLIAMSON, C.H.K. & ROSHKO, A. 1988 Vortex formation in the wake of an oscillating cylinder. *J. Fluids Struct.* **2** (4), 355–381.

WONG, K.W.L., ZHAO, J., LO JACONO, D., THOMPSON, M.C. & SHERIDAN, J. 2018 Experimental investigation of flow-induced vibration of a sinusoidally rotating circular cylinder. *J. Fluid Mech.* **848**, 430–466.

YAO, W. & JAIMAN, R.K. 2017 Feedback control of unstable flow and vortex-induced vibration using the eigensystem realization algorithm. *J. Fluid Mech.* **827**, 394–414.

YU, D., MEI, R. & WEI, S. 2002 A multi-block lattice Boltzmann method for viscous fluid flows. *Intl J. Numer. Meth. Fluids* **39**, 99–120.

ZHANG, M.M., CHENG, L. & ZHOU, Y. 2004 Closed-loop-controlled vortex shedding and vibration of a flexibly supported square cylinder under different schemes. *Phys. Fluids* **16** (5), 1439–1448.

ZHANG, W., LI, X., YE, Z. & JIANG, Y. 2015 Mechanism of frequency lock-in in vortex-induced vibrations at low Reynolds numbers. *J. Fluid Mech.* **783**, 72–102.

ZHAO, M., KAJA, K., XIANG, Y. & YAN, G. 2013 Vortex-induced vibration (VIV) of a circular cylinder in combined steady and oscillatory flow. *Ocean Engng* **73**, 83–95.

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** (6), 063607.

ZHOU, J., ADRIAN, R.J., BALACHANDAR, S. & KENDALL, T.M. 1999 Mechanisms for generating coherent packets of hairpin vortices in channel flow. *J. Fluid Mech.* **387**, 353–396.