Bluff body uses deep-reinforcement-learning trained active flow control to achieve hydrodynamic stealth

Feng Ren,[1,2] Chenglei Wang,[1] and Hui Tang[1, a)]

[1)] *Research Center for Fluid Structure Interactions, Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong, China*

[2)] *School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, China*

(Dated: 5 August 2021)

We propose a novel active-flow-control strategy for bluff bodies to hide their hydrodynamic traces, i.e., strong shears and periodically shed vortices, from predators. A group of windward-suction-leeward-blowing (WSLB) actuators are adopted to control the wake of a circular cylinder submerged in a uniform flow. An array of velocity sensors is deployed in the near wake to provide feedback signals. Through the data-driven deep reinforcement learning, effective control strategies are trained for the WSLB actuation to mitigate the cylinder's hydrodynamic signatures. Only a 0.29% deficit in streamwise velocity is detected, which is a 99.5% reduction from the uncontrolled value. The same control strategy is found to be also effective when the cylinder undergoes transverse vortex-induced vibration. The findings from this study can shed some lights on the design and operation of underwater structures and robotics to achieve hydrodynamic stealth.

[a)] Electronic mail: h.tang@polyu.edu.hk

I. INTRODUCTION

Predator-prey interactions influence ecosystem structures. In aquatic communities, these interactions are strongly affected by disturbances in the surrounding liquid medium, especially when vision detection does not work well. One example is that harbor seals (Phoca vitulina) use whiskers to sense hydrodynamic perturbations produced by their preys, predators or conspecifics, thus helping to forage in dark or turbid water[1–3]. Another example is that fish use their lateral lines to perceive weak hydrodynamic perturbations induced by surrounding stationary or moving bodies[4]. Inspired by these aquatic animals, scientists and engineers have strived to develop sensible sensors for underwater detection[5].

From the preys' perspective, if achieving hydrodynamic stealth of a certain level by mitigating or even eliminating their hydrodynamic traces in the form of shears and/or shed vortices, they can significantly reduce the chance of being captured by their predators. Even though the idea of stealth has been extensively explored and even implemented in engineering applications in various disciplines such as optics[6], acoustics[7], etc, it has occasionally been tackled in hydrodynamics under a different name "hydrodynamic cloaking". In this emerging field, existing studies are only focused on either creeping flows[8–10] or water waves[11,12], both governed by linear equations. The control methods include the use of well-designed metamaterials[8,9], applying mass/momentum sources[10], or varying water depth[11] around an obstacle.

In this study, we propose to apply active flow control (AFC) to mitigate or even eliminate the hydrodynamic traces, so that preys can achieve hydrodynamic stealth on demand and escape from their predators. Due to the unsteadiness, nonlinearity and high dimensionality of this flow control problem, it is challenging to find effective control laws in an explicit form. Instead, we adopt deep reinforcement learning (DRL), a machine learning technique, to search feasible control strategies through the interactions between an intelligent agent and environment. DRL uses a data-driven approach, i.e., artificial neural networks (ANNs), to approximate strongly nonlinear dynamics by learning through trial-and-error[13]. Compared to other popular machine learning techniques such as the genetic programming used in our previous AFC work[14], it is more suitable for multi-input multi-output (MIMO) control problems like in the present study. Besides the striking, high-profile victories in the game of Go against the best human players[15,16], DRL has recently been successively applied in automatic

control[17] and complex flow-related problems, e.g., fish schooling[18], bird flying in turbulent environment[19], flow navigation of smart particles[20], etc. Applications of DRL trained AFC for drag reduction at various flow conditions have also been actively explored[21–24]. Successful DRL applications in fluid mechanics can be found in recent review papers by Brunton et al.[25], Rabault et al.[26], and Ren et al.[27]. All these successful applications of DRL in flow problems especially AFC further make us feel confident to realize the above idea of applying AFC to achieve the hydrodynamic stealth for bluff bodies.

II. PROBLEM DESCRIPTION AND METHODOLOGY

A. Problem description

We adopt a generic bluff body, i.e., a circular cylinder, for the control. As shown in FIG 1(a), a circular cylinder with mass m and diameter D is immersed in a uniform flow of velocity U. The corresponding diameter-based Reynolds number is fixed at $Re = 100$, such that asymmetric vortices will periodically shed from the cylinder, forming a Kármán vortex street. The cylinder's streamwise motion is refrained all the time while it is transversely connected to a spring of stiffness K. The asymmetric vortex shedding results in a periodic hydrodynamic lift (F_L) exerted on the cylinder, which thus undergoes vortex-induced vibration (VIV) governed by

$$m\ddot{y} = -Ky + F_L \tag{1}$$

Two key dimensionless parameters, i.e., the mass ratio m^* and the reduced velocity U_R, can be used to characterize the cylinder's dynamics

$$m^* = \frac{m}{\rho D^2}, \quad U_R = \frac{U}{f_N D} \tag{2}$$

where ρ is the fluid density and $f_N = \sqrt{K/m}/2\pi$ is the natural frequency of this mass-spring system in vacuum. By varying m^* and U_R, the cylinder may vibrate in a "locked-in" regime where the vibration amplitude is significantly large due to resonance, making the control very challenging[28]. FIG 1(b) to (d) show the uncontrolled wake patterns behind the stationary, locked-in vibrating, and unlocked-in vibrating cylinders, respectively, all revealing strong hydrodynamic signatures. Due to the large-amplitude vibrations, the wake of the locked-in VIV cylinder seems quite different from the wakes in the other two cases in terms of the wake width and vortex orientation.

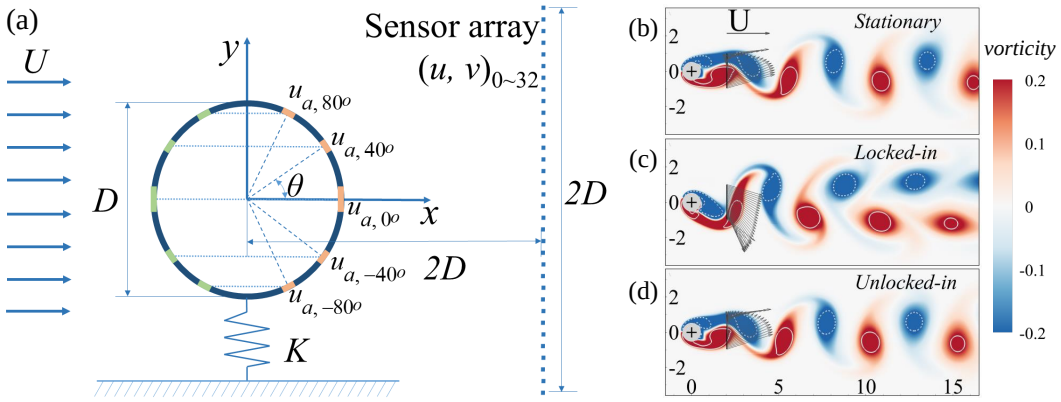

FIG. 1. (a) Schematic of the AFC problem. Five pairs of WSLB actuators are adopted to control the cylinder wake, arranged symmetrically ($\theta = \pm 80°$, $\pm 40°$, and $0°$) about the cylinder's centerline. Each pair of actuators consists of a suction slot (green arc) at the windward side of the cylinder and a blowing slot (red arc) at the leeward side. (b) to (d) Wake patterns of stationary cylinder, locked-in VIV cylinder ($m^* = 2$, $U_R = 5$), and unlocked-in VIV cylinder ($m^* = 1.5$, $U_R = 3$), respectively, at the same instant when the cylinder reaches its equilibrium position in the upstroke. The wakes are colored by vorticity. The vortices are identified using the λ_{ci} criterion[29] and enclosed by grey lines, where solid and dashed lines indicate positive and negative vortices, respectively. Velocity vectors measured by the sensor array are also shown to demonstrate strong hydrodynamic signatures in the cylinder wakes.

To mitigate the cylinder's hydrodynamic traces, five pairs of windward-suction-leeward-blowing (WSLB) actuators are adopted to provide control authority[30,31], as depicted in FIG 1(a). In line with the streamwise direction, each pair of actuators consists of a suction slot at the windward side of the cylinder and a blowing slot at the leeward side. The five leeward actuators are located azimuthally at $\theta = \pm 80°$, $\pm 40°$, and $0°$, measured from the centerline. The actuator pair in the centerline has slots of $12°$ arc, while the other eight off-centerline actuators have slots of $6°$ arc, altogether occupying one fifth of cylinder's surface. For each pair of actuators, fluid is horizontally sucked into the windward slot and injected out from the leeward slot with an identical velocity, hence realizing zero-net-mass-flux actuation. Since the targeted wake flow is symmetric about the centerline, to simplify the control, the actuator pairs located at $\theta = \pm 80°$ are always operated in phase, and so are the pairs located at $\theta = \pm 40°$. Hence only three independent controls are employed in the current study.

To provide feedback signals to the controller, an array of velocity sensors is placed $2D$ downstream of the cylinder, sensing the cylinder's hydrodynamic signatures through measuring the streamwise and transverse velocities at 33 equal-spacing locations along a vertical line of $2D$ length, as depicted in FIG 1(a). This sensor array is chosen according to an experiment setup for drag measurement[32]. In practice, it can be a rake of hot-wire probes or Pitot tubes.

With the velocity information collected from the sensor array, a cost function J is defined for the purpose of flow control

$$J = \sum_{i=1}^{N_S} \left[(u_i - 1)^2 + v_i^2 \right] + C \sum_{i=1}^{N_A} u_{a,i}^2 \tag{3}$$

where $N_S = 33$ and $N_A = 5$ are the numbers of sensors and WSLB actuators, respectively. u_i and v_i are the instantaneous streamwise and transverse velocities read by the i-th sensor, and u_a is the actuation velocity. This cost function consists of two components. The first component, i.e., the first term of Eq. 3, quantifies the velocity deficit, namely, the discrepancy in velocities between the wake and the uniform incoming flow (denoted as J_1 hereafter). $J_1 = 0$ when the wake is identical to the incoming flow, indicating a zero hydrodynamic trace. The second component reflects the energy consumed by the WSLB actuators for the control (the second term of Eq. 3), in which a small weighting coefficient $C = 10^{-4}$ is applied to prevent over fluctuation of the actuation. To avoid overloading, the upper limit of the actuation velocity (u_a) is set as five times the incoming flow velocity.

B. CFD simulation environment

The above flow control problem is studied in a computational-fluid-dynamics (CFD) simulation environment. The details of the solver can be found in Appendix A. When performing DRL-based AFC, hundreds or even thousands of episodes (each one representing an independent case) are required to train the control strategy. This poses great challenges to the computational efficiency of the flow solver. To address this issue, our flow solver is accelerated using the GPU parallel algorithm[33] and run in an NVIDIA Tesla K40c GPU server. With this platform, the averaged computational time for each control case can be managed within 6 minutes.

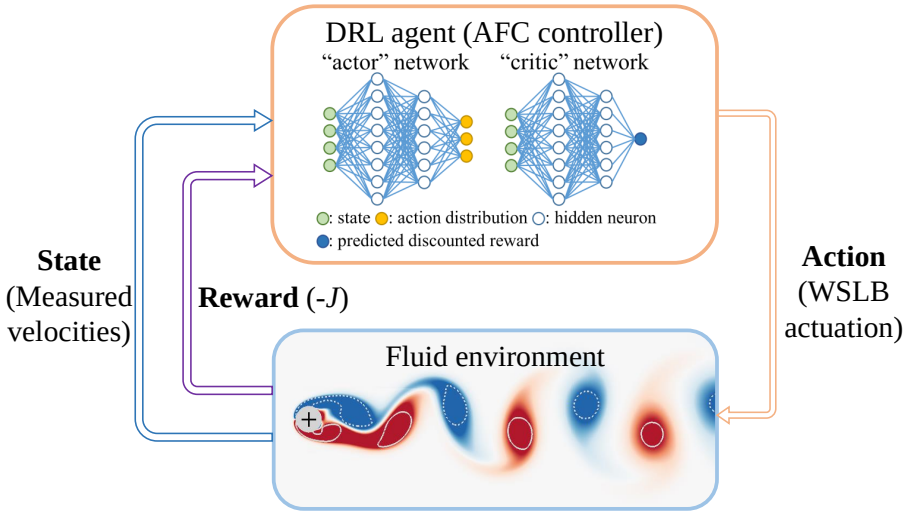

FIG. 2. Schematic of the DRL loop used in the present study

C. DRL trained AFC

In the present DRL framework, effective control strategies are learnt through interactions between the controller, i.e., a smart "agent", and the cylinder wake, i.e., the "fluid environment". As illustrated in FIG 2, the agent feels the "state" of the environment using the sensor array, and exerts changes to the environment through "action", i.e., the WSLB actuations. To encourage the agent to keep improving its performance during the learning process, a "reward" is defined as $r = -J$, such that a better action gets a higher reward.

The proximal policy optimization (PPO) model[34], currently often regarded as the state-of-the-art algorithm for performing continuous control, is adopted to train the DRL agent through stochastic trial-and-error. In this model, two sets of ANNs are used in the agent, known as the "critic" and "actor" networks. As depicted in FIG 2, both networks use the state as their input. On the output side, the "critic" network estimates the reward, whereas the "actor" network models the agent's policy π_Θ, i.e., the control strategy. By using this PPO-based DRL framework, it is expected that the agent can learn effective control strategies in a similar way like humans. More details about this framework are provided in Appendix B.

The control strategies will be learnt in two scenarios. In the first scenario, the spring stiffness K is set infinitely large, so that the cylinder is stationary. The simulation in each episode lasts for $27T$, where $T = D/U$, approximately 5 vortex shedding periods in the

uncontrolled case. Here an episode is defined as a complete run with reinitialized flow environment and a control strategy updated according to the states, actions and rewards obtained in previous runs. In the second scenario, K is set finite, so that the cylinder experiences VIV. The simulation lasts for $54T$, doubled because the vibrating cylinder and its wake require extra time to reach the steady state. In each episode, states are read from the flow field 50 times in a duration of T, and accordingly, the actions are also enforced in the same pace.

In general, appropriate choices of network width (i.e., the number of neurons in each hidden layer) can improve the training performance, while bad choices will lead to over-fitting or under-fitting problems. To determine suitable hyper-parameters for the current DRL training, a number of trial learnings have been conducted on the stationary cylinder. The tested hyper-parameters in five selected trial learnings are listed in TABLE I, and the corresponding learning curves are presented in FIG 3. Here a learning curve describes the variation of the averaged cost function value against episodes. It is seen that, as a successful learning, the learning curve of Trial I not only shows a generally decreasing trend followed by a level-off in the late stage, but also gradually becomes smoother. For other trials, however, the learning processes have to be terminated earlier due to either large fluctuations or non-decreasing trends. In addition, two more trial learnings with ANNs of width 1024×128 and 512×256, respectively, blow up in just the first 50 episodes, and therefore are not presented in FIG 3.

In the DRL framework, the weights of ANNs are randomly initialized. This poses challenges in the consistency of the learning, not mentioning the discrepancies caused by difference in ANN structures. Thus, Trial V is chosen to compare with Trial I so that the effect of different initial ANN weights can be demonstrated. It is seen from FIG 3 that, although Trial V shows a much faster decreasing trend in the early stage, its performance is worse off at approximately the 450th episode, reflected by the unexpected rising and violently fluctuating learning curve. Nevertheless, its performance is still better than the other trials using different ANN or learning rate settings. Hence we adopt the hyper-parameters used in Trials I and V in this study.

TABLE I. Hyper-parameters used in the DRL

Learning	Neural Network	Learning Rate		Discount Factor
		Critic	Actor	
I	512 × 128	0.0002	0.0003	0.97
II	256 × 128	0.0002	0.0003	0.97
III	512 × 128	0.0002	0.0003	0.99
IV	512 × 128	0.0002	0.0003	0.93
V	512 × 128	0.0002	0.0003	0.97

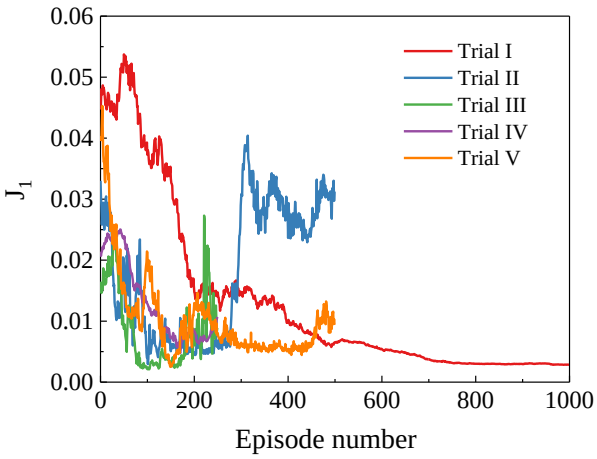

FIG. 3. Comparison of learning curves of DRL control over a stationary cylinder with different hyper-parameter setups listed in TABLE I

III. RESULTS AND DISCUSSIONS

A. Stationary cylinder

FIG 4 presents more details of the successful learning introduced in FIG 3 (i.e., Trial I). It is seen that the control strategy improves in general with a fast rate in the first 500 episodes, and then gradually converges. Four representative episodes, i.e., Episode 1, 256, 493 and 1000, are chosen to show details of the WSLB actuation and the control effects at different stages of the learning process. In Episode 1, without any prior knowledge, the DRL agent manages to break down the asymmetric vortex shedding from the cylinder by operating the 0° actuator with the strongest actuation and the ±80° actuators with the

weakest. As a result, the originally unsteady wake becomes nearly steady, consisting of two strong shear layers mirrored about the cylinder's centerline, as shown in subplot (A) on the right. However, eminent velocity deficit in the wake can still be detected by the sensor array. As the learning proceeds and more experience is obtained, the DRL agent gradually reduces the 0° actuator's strength while increasing the $\pm 40^\circ$ actuators'. The two shear layers become less evident and the resulting velocity profile gradually approaches to a uniform one.

Once a mature control strategy is learnt, the DRL agent decides to operate the $\pm 40^\circ$ actuators with the greatest strength, i.e., $u_{a,+40^\circ} = u_{a,-40^\circ} = 3.1$, and operate the 0° actuator with the smallest strength, i.e., $u_{a,0^\circ} = 2.3$. This strategy is somewhat consistent with what has been revealed in previous wake/VIV control studies[31,35–37]: to make the control effective and efficient at $Re = 100$, the jet-like actuation should be applied close to the mean separation points, i.e., $\theta = \pm 58^\circ$. With this control strategy, the shear layers are no longer tangible and the measured velocity profile almost coincides with the uniform one as in the incoming flow (the mean discrepancy is only 0.29%, reduced by 99.5% if compared with the uncontrolled case), as shown in subplot (D) in FIG 4. A video showing the evolution of the flow field with this well-trained control strategy can be seen in FIG 5 (multimedia view). Through this control the cylinder's hydrodynamic traces are almost eliminated, and any predator just $2D$ downstream can hardly feel its existence. Furthermore, as revealed in the four insets, the fluctuations in the actuation are significantly mitigated when the learning process converges, indicating great confidence of the learnt control strategy.

Time histories of the cost function, hydrodynamic forces and velocity fields obtained in Episode 1000 are presented in FIG 6. It is seen that the control efficiently makes J_1, the velocity deficit in the wake, reach and maintain its lowest values right after about $t = 5$, slightly more than one oscillation cycle of the uncontrolled cylinder. The control also quickly reduces the hydrodynamic forces, in both streamwise and transverse directions, to values very close to zero, as shown in FIG 6(b). The nearly zero transverse force, i.e., lift, reflects that the asymmetric vortex shedding has been successfully suppressed by the WSLB actuation. The slight oscillation around zero in the streamwise force indicates that the flow induced drag is almost balanced by the WSLB generated trust, which, from the momentum perspective, confirms the recovery of the uniform flow in the wake.

FIG 6(c) to (d) compare the spatio-temporal distributions of the net streamwise velocity, $u - 1$, and transverse velocity, v, measured by the sensor array before and after the WSLB

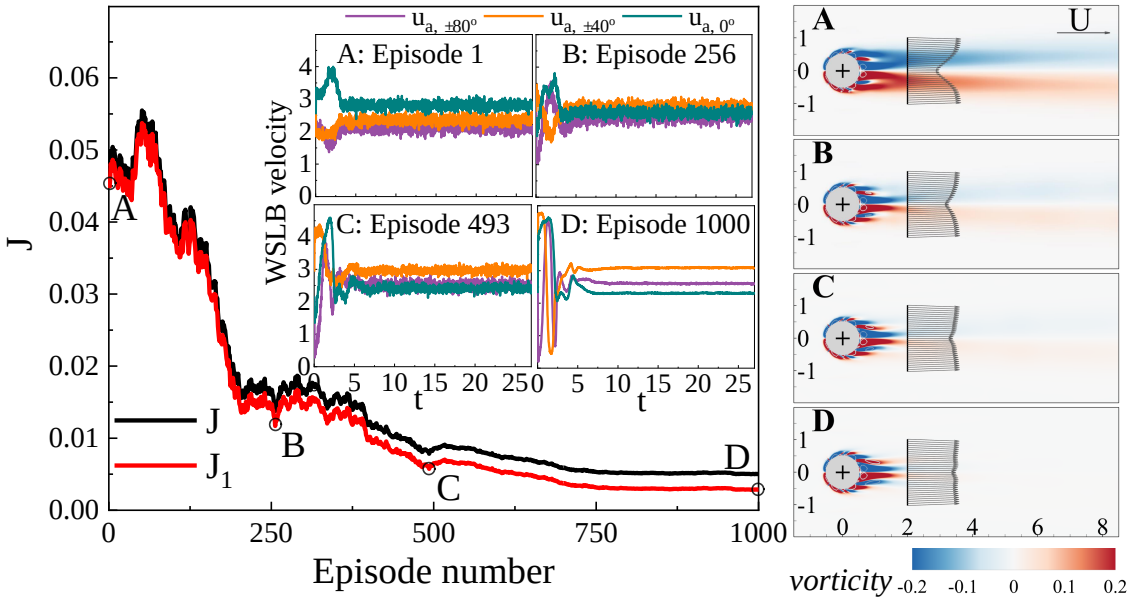

FIG. 4. Left: Learning process represented by the variation of cost function values against episode number. The four insets show WSLB actuations generated by the DRL agent at different stages of the learning. Right: Instantaneous wake patterns and measured velocity profiles at the four selected stages. The progressive disappearance of velocity deficit well demonstrates the effectiveness of DRL trained wake control.

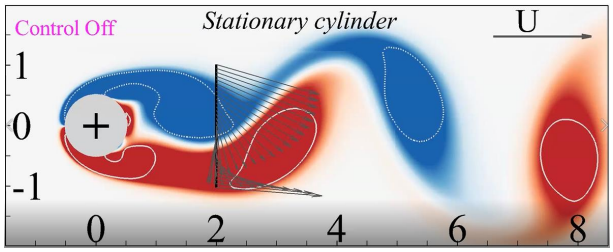

FIG. 5. Evolution of hydrodynamic traces of the stationary cylinder subject to the DRL-trained WSLB control (Multimedia view).

control, both normalized using the freestream velocity U. It is seen that, before the control, both the net streamwise and transverse velocities are spatio-temporally symmetric. Near the cylinder's centerline, the net streamwise velocity is always negative, reflecting the drag generation characteristics. After the control is switched on, the spatio-temporal patterns are quickly broken down, both evolving into nearly zero velocity fields.

Through the learnt WSLB control, the vortex evolution and resulting wake pattern are

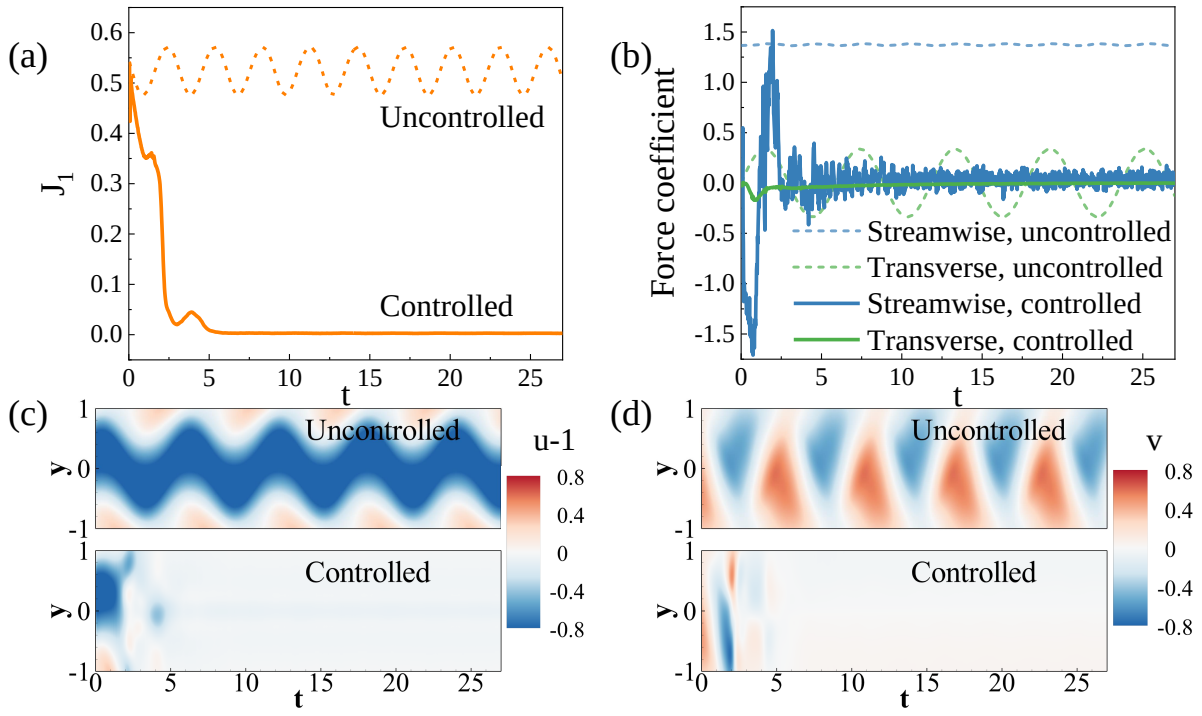

FIG. 6. Evolution of (a) J_1 value and (b) force coefficients of uncontrolled and well-controlled cases for the stationary cylinder. The control strategy learnt in Episode 1000 (see FIG 4) is adopted in the controlled case. The force coefficients were calculated by nondimensionalizing the streamwise and transverse components of hydrodynamic forces using $\rho U^2 D/2$. (c) and (d): Evolution of the corresponding net streamwise velocity, u–1, and transverse velocity, v, measured by the sensor array, respectively. The control starts at $t = 0$.

significantly changed, indicating the change of associated wake instability. To quantify the latter change, we perform a linear stability analysis by following similar procedures adopted for cylinder wake flows in previous works[35,38–40] (detailed in Appendix C). FIG 7 presents the distribution of complex critical frequency of disturbance, ω_0, along the downstream direction for both the uncontrolled and controlled cases. Here the real and imaginary parts, i.e., $\mathrm{Re}(\omega_0)$ and $\mathrm{Im}(\omega_0)$, represent the frequency and growth rate of disturbance, respectively. It is seen from FIG 7(a) that the learnt control reduces the absolute instability region (defined by the x range of positive $\mathrm{Im}(\omega_0)$) by about 73%. The control also reduces the disturbance growth rate in the convective instability region (defined by the x range of negative $\mathrm{Im}(\omega_0)$) by about one order of magnitude. From FIG 7(b), it is seen that the control significantly enhance the dominant frequency of instability in the near wake, indicating that, with the control, the

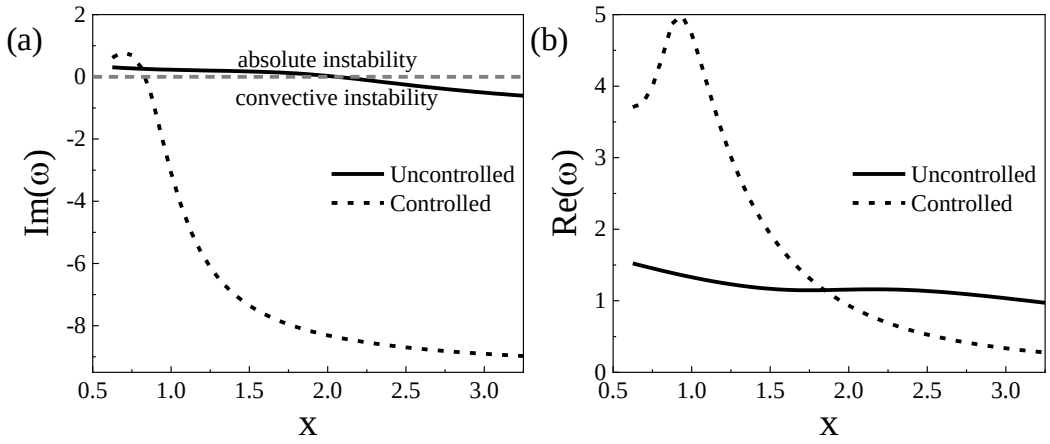

FIG. 7. Comparisons of (a) imaginary and (b) real parts of critical ω_0 along the streamwise direction between uncontrolled and controlled cases.

near wake flow becomes less sensitive to disturbance close to the original vortex shedding frequency (i.e., $\mathrm{Re}(\omega) = 1.06$). All these changes in the wake instability confirm that the learnt control is able to stabilize the cylinder wake and greatly reduce the hydrodynamic traces.

B. VIV cylinder

The success of DRL trained control strategies for the stationary cylinder has motivated us to further explore the control in a more complicated scenario where the cylinder undergoes VIV in the transverse direction. In this scenario, the fixed sensor array has a relative motion to the cylinder, generating different readings as revealed in the velocity vectors shown in FIG 1(b) to (d). This change may cause the deterioration of the WSLB control and pose challenges to the DRL training. Intriguingly, since the control learnt from the stationary cylinder makes the hydrodynamic lift very close to zero (as revealed in FIG 6(b)), the same control strategy may be effective in suppressing the VIV first and then mitigating the hydrodynamic traces. Hence, the control strategy learnt in Section III A, i.e., the well trained ANNs, is directly applied as the starting point in the new learning. This is a machine learning technique called transfer learning where the knowledge acquired from one task is utilized in another different but related task.

To show the necessity and effectiveness of using transfer learning for the VIV cylinder, we compare the first 250 episodes of the learning for the selected two VIV cases, i.e., a locked-in

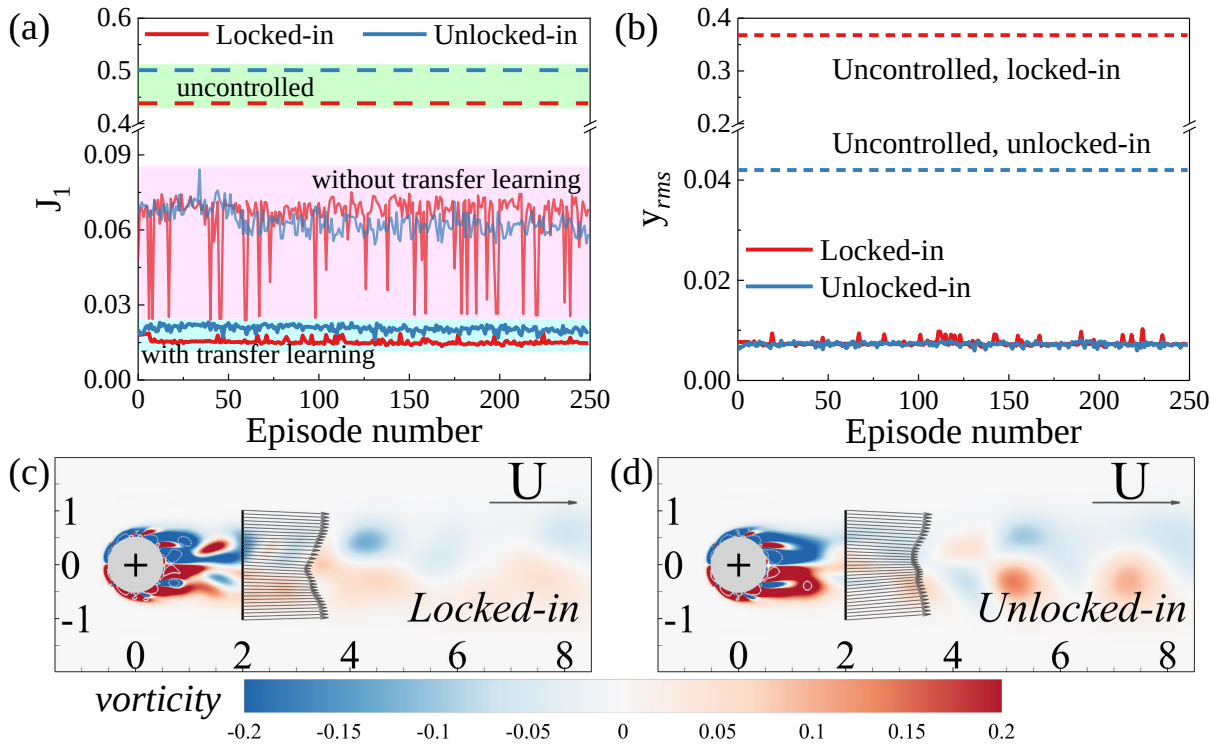

FIG. 8. (a) Learning processes of DRL trained AFC on the locked-in and unlocked-in VIV cylinder. (b) Variation of standard deviations of the cylinder's transverse displacement throughout the learning process, which are evaluated using the data in the second half of each episode. (c) and (d): Instantaneous wake patterns and measured velocity profiles for the well-controlled locked-in and unlocked-in VIV cylinder, respectively. In both cases, the control strategies learnt in Episode 250 are adopted.

case and an unlocked-in case, as shown in FIG 8(a). It is seen that with the transfer learning, J_1 in both cases remains low with slight fluctuations. Its mean is almost constant starting from the very beginning of the learning. Without using the transfer learning, however, the learning curve in the locked-in case fluctuates violently and does not show a decreasing trend. Although fluctuating mildly, J_1 in the unlocked-in case also remains at a high level and does not show an obvious decreasing trend either.

With the control strategies obtained from transfer learning, the cylinder's VIV can be suppressed to a large extent in both the locked-in and unlocked-in situations, as evidenced in FIG 8(b) by the very small standard deviations in transverse displacement. As expected, these control strategies works well as long as the VIV is suppressed. They effectively mitigate the cylinder's hydrodynamic traces, reducing the mean velocity deficit by 96.5% in

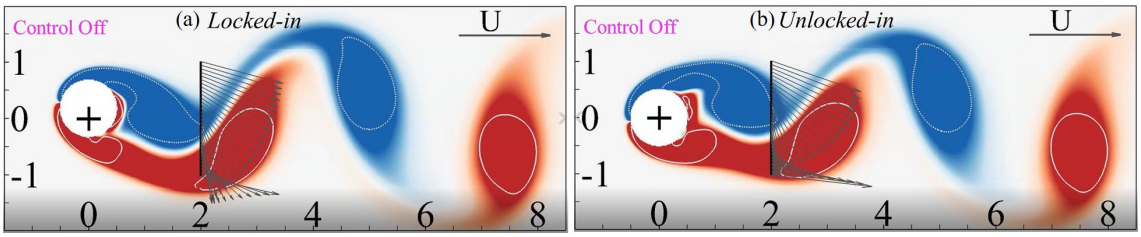

FIG. 9. Evolution of hydrodynamic traces of the VIV cylinder subject to the DRL-trained WSLB control: (a) locked-in and (b) unlocked-in (Multimedia view).

the locked-in case and 95.9% in the unlocked-in case, respectively, compared with the corresponding uncontrolled cases. The controlled wakes are shown in FIG 8(c) and (d). Unlike in the controlled case for the stationary cylinder, vortex shedding can still be observed in these two cases, mainly due to the imperfect VIV control. However, with the learnt control, this strong hydrodynamic signature is to a great extent constrained in the near wake, i.e., $x < 2D$, beyond which only very weak vorticities are observed. The remarkably weakened vorticities in the mid-wake will soon disappear from the background flow. Two videos showing the control effects for the VIV cylinder in the locked-in and unlocked-in situations can be seen in FIG 9 (multimedia view).

Time histories of the WSLB velocities in the selected episodes for the two VIV cases are presented in FIG 10, and other quantities reflecting the control effect are presented in FIG 11. Unlike in the stationary cylinder case, the WSLB actuations still oscillate significantly even when the learnt control achieves its steady state. The $\pm 40^\circ$ actuators are still strongest, with their means slightly higher than that in the stationary cylinder case by 12% in the locked-in case and by 5.7% in the unlocked-in case, respectively. The wake controls are not as good as in the stationary cylinder case, as shown in FIG 8(c) and (d) as well as in FIG 11(c) and (d). All these discrepancies are probably caused by the slight vibration of the controlled cylinder, as revealed in FIG 11(b). If comparing these two VIV cases, it is seen that the locked-in case requires much longer time to achieve its steady state, as indicated by the transient processes shown in FIG 10 and 11.

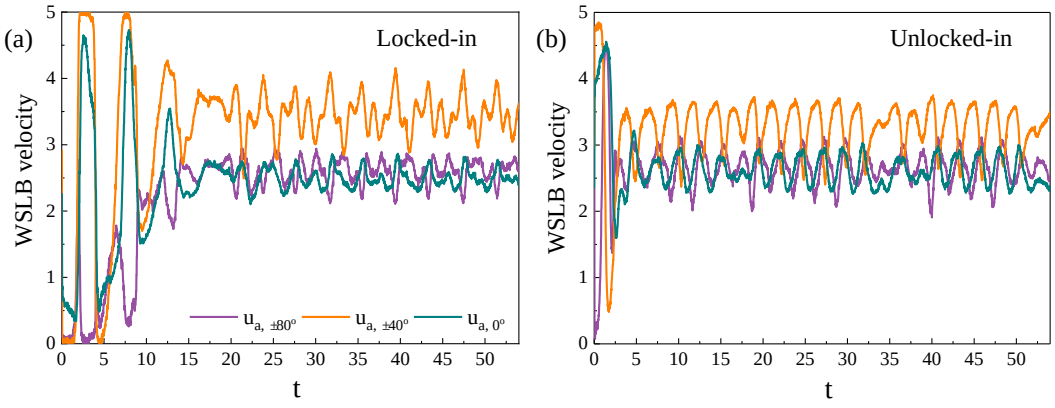

FIG. 10. Transient WSLB actuations generated by the DRL in (a) the locked-in and (b) unlocked-in cases

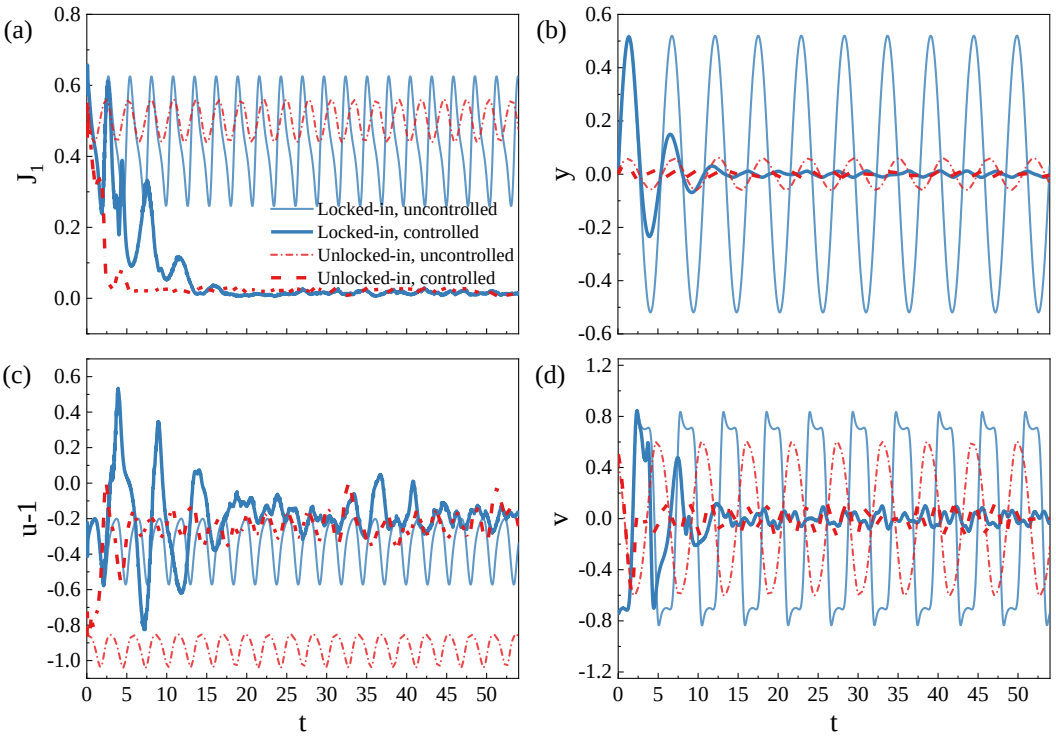

FIG. 11. Evolution of selected quantities in the locked-in and unlocked-in cases: (a) J_1 value; (b) Transverse displacement of the cylinder; (c) Net streamwise velocity at point $(2, 0)$; (d) Transverse velocity at point $(2, 0)$

IV. SUMMARY

We demonstrated that using DRL trained AFC a bluff body can achieve hydrodynamic stealth from its predators. The well-trained WSLB control not only balances the body's

hydrodynamic drag during cruising, but also greatly reduces the velocity deficit in the very near wake, hence mitigating or even eliminating the body's hydrodynamic traces. The linear stability analysis revealed that the control can reduce by about 73% the length of absolutely unstable region in the cylinder wake. It was also shown that the trained WSLB control can effectively suppress VIV, and hence can mitigate the hydrodynamic traces of VIV bodies. Although many parameters are unexplored, e.g., the effects of Reynolds number, WSLB actuator settings and sensor settings, the findings from this interesting study can shed some lights on the design and operation of underwater robotics for achieving low detectability.

In the near future, we will extend the study to more realistic conditions, where the flow can be turbulent and three-dimensional, the DRL trained AFC may be realized in experimental environment, and uncertainties in measurements and actuations may be involved. To perform efficient learning in these challenging studies, making the use of system invariances and accelerating the DRL through a multi-environment approach[41] can play vital roles.

ACKNOWLEDGMENTS

This study was financially supported by Research Grants Council of Hong Kong under General Research Fund (Project No. 15249316 & 15214418), PolyU Faculty of Engineering Dean's Reserve (Project No. ZVVT) and Departmental General Research Fund (Project No. YBXQ).

DATA AVAILABILITY

The data that supports the findings of this study are available within the article.

APPENDIX

A. CFD solver

High-fidelity CFD simulations are conducted to provide training data for DRL based machine learning. In the simulations, the fluid is assumed incompressible. We adopt the lattice Boltzmann method (LBM) to numerically solve the Navier-Stokes equations, using an evolution procedure with separated collision and streaming steps. In this method, we use

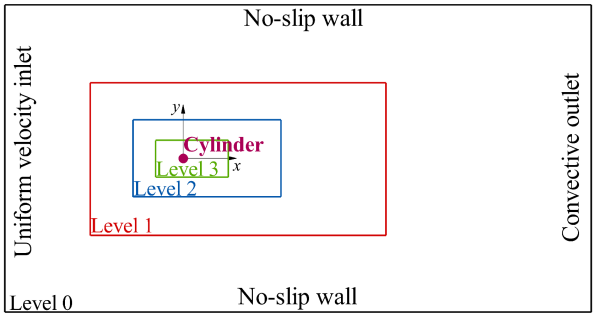

FIG. 12. Schematic of the computational domain, multiblock meshing arrangement, and boundary conditions

the multi-relaxation time algorithm[42] to enhance numerical stability, and use the He-Luo model[43] to better satisfy the fluid incompressibility condition.

FIG 12 shows a schematic of the computational domain, multiblock meshing arrangement and boundary conditions. The size of the computational domain is $64D\times20D$. The circular cylinder is initially located at the centerline, $20D$ downstream from the inlet. The multi-block grid partition method[44] 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, each having 512×160 uniform mesh nodes. Around the cylinder, the finest block (level 3) with mesh spacing $\delta x = D/64$ is used. The inlet velocity is set as $U = 0.02c$, where c is the lattice speed, corresponding to a time step $\delta t = T/3200$, where $T = D/U$. Unless otherwise stated, in the following discussions all velocities, lengths and time are nondimensionalized using U, D and T, respectively.

We apply the velocity boundary condition at the inlet and top/bottom walls, which is achieved via a modified bounce-back scheme with momentum exchange[45]. The convective flow condition, i.e., $\partial_t u+U\partial_x u = 0$, is utilized at the outlet to allow the vortices to smoothly cross the boundary with the least reflection[46]. The WSLB actuations are realized via the Dirichlet boundary condition at the designated slots. Regarding the vibration of the cylinder, we employ the double-linear interpolations for the treatment of curved boundary[47], the third-order non-equilibrium extrapolation scheme for the mesh refilling process[48], and the corrected momentum exchange method[49] to calculate the hydrodynamic forces. This combination of algorithms has been proven to be sufficiently accurate for simulating moving boundary problems, via comparisons with groups of existing methods[50]. The grid partition

and boundary setups have also been verified in our previous studies[14,31,35–37].

Note that the simulations in the current study are conducted at a relatively low Reynolds number where the flow is laminar. For simulations at high Reynolds numbers where the flow could be turbulent and carries high-frequency information, directly solving the Navier-Stokes equations using other algorithms may be necessary to ensure the simulation accuracy.

B. PPO-based DRL

In each episode, a DRL agent runs the policy for N times and collects a trajectory, i.e., a sequence of state-action-reward combinations

$$\tau = (s_1, a_1, r_1), (s_2, a_2, r_2), (s_t, a_t, r_t), ..., (s_N, a_N, r_N) \tag{A1}$$

To consider a long-term effect of the actions, a discounted reward is used

$$R_t = \sum_{t'>t} \gamma^{t'-t} r_{t'} \tag{A2}$$

where γ is the discount factor usually close to 1. This definition gives later rewards more weights in calculating R_t.

In order to update the policy π_Θ, properly defining the objective function for each set of ANNs is vital. First, the objective of the "critic" network is to minimize the discrepancy between the predicted and actual discounted rewards, using the objective function

$$J_{critic} = \hat{E}_t(-\hat{A}_t^2) \tag{A3}$$

where $\hat{E}_t$ denotes the empirical expectation over time, and $\hat{A}_t$ is called "advantage" that evaluates the difference between the predicted and actual discounted rewards

$$\hat{A}_t = R_t - V_\Theta(s_t) \tag{A4}$$

However, as learning proceeds, the agent would always wish to achieve larger reward than its prediction, i.e., to obtain larger advantage. To achieve this, the policy-based DRL algorithms usually utilize the second set of network, i.e., the "actor" network. By using the PPO model, we follow the work in Schulman et al.[34], where a clipped surrogate objective function is used, i.e.,

$$J_{actor} = \hat{E}_t[\min(q_t(\Theta)\hat{A}_t, \text{clip}(q_t(\Theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)] \tag{A5}$$

where $q_t(\Theta) = \pi_\Theta(a_t|s_t)/\pi_{old}(a_t|s_t)$ is the ratio of the probability of current policy π_Θ in adopting action at according to state s_t to the probability of previous policy π_{old}. The clip term inside Eq. A5 indicates that $R_t(\Theta)$ is constrained to an interval $[1-\epsilon, 1+\epsilon]$, where ϵ is a hyper-parameter set as 0.2 as recommended by Schulman et al.[34]. The use of the minimum operation is to avoid an excessively large policy update.

When updating the policy, we choose the Adam optimizer, which performs better in fast convergence than conventional stochastic gradient descent optimizers[51]. In this optimizer, two decay rates are usually fixed and only the learning rate needs to be set. In our study, the learning rates are set as very small values, i.e., 0.0002 and 0.0003 for the actor and critic networks, respectively, which we found can help produce stable learning curves.

When dealing with continuous control, the actor network does not directly generate actions. Instead, it generates a combination of parameters for a certain probability distribution for actions. In this study, we choose the Gaussian distribution, from which the actions are sampled in a predefined range. After 500 episodes, we narrow the sampling range via reducing the standard deviation of the Gaussian distribution. By this means, randomness in the sampling process is reduced and hence the fluctuations of sampled actions are suppressed. Meanwhile, as confirmed from the learning curves, both the trend and smoothness of the learning process are unaffected, implying the appropriateness of this adjustment. Note that a narrow Gaussian distribution at the early stage is not suggested, because it will lead to a small searching space.

C. Linear stability analysis

In the linear stability analysis for the flow around a circular cylinder, the mean flows of both uncontrolled and controlled cases are chosen, as shown in FIG 13(a), which are assumed to be inviscid and locally parallel downstream of the cylinder, the same as in Wang et al.[35], Thiria and Wesfreid[40]. Introduce a disturbance stream function

$$\Psi(x, y, t) = \psi(y)e^{i(kx-\omega t)} \tag{A6}$$

where ψ is the disturbance amplitude, k is the complex wave number, and ω is the complex frequency. Substituting this stream function into the inviscid Orr-Sommerfeld or Rayleigh equation gives

$$(k\bar{u} - \omega)(\psi'' - k^2\psi) - k\bar{u}''\psi = 0 \tag{A7}$$

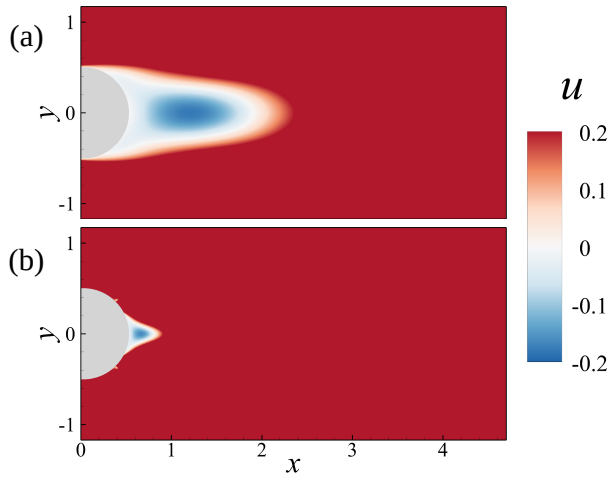

FIG. 13. Time-averaged streamwise velocity in the (a) uncontrolled and (b) controlled cases.

with boundary conditions $\psi(\pm\infty) = 0$. $\bar{u}$ is the time-averaged streamwise velocity at a given downstream location x, and the double prime denotes the second derivative with respect to y.

Equation A7 defines an eigenvalue problem and yields a dispersion relation $\omega = \omega(k)$ at a selected downstream location x. By mapping a number of constant $\mathrm{Im}(k)$ (i.e., imaginary part of k) lines in the complex ω plane, a "critical point", ω_0, can be identified at the edge of a cusp-like trajectory[35,38–40], as exemplified by FIG 14 at $x = 1$ in the uncontrolled case. Its real and imaginary parts, $\mathrm{Re}(\omega_0)$ and $\mathrm{Im}(\omega_0)$, represent the frequency and growth rate of disturbance, respectively. If $\mathrm{Im}(\omega_0)$ is positive, the flow is absolutely unstable, meaning that the disturbance will grow with time and eventually affect the entire flow field. Otherwise, the flow is convectively unstable, and the disturbance decays with time.

Usually, ω_0 can also be determined when the group velocity of disturbance vanishes for a given complex absolute wave number (k_0), i.e.

$$\left.\frac{\partial \omega}{\partial k}\right|_{k_0} = 0 \tag{A8}$$

Using the iterative method introduced by Deissler[52], one can quickly obtain the k_0-ω_0 pair at the critical point without repeating the above cusp map procedure many times. For example, the k_0-ω_0 pair calculated at $x = 1$ in the uncontrolled case is $k_0 = 1.578 - 0.799i$ and $\omega_0 = 1.329 + 0.228i$, consistent with the result shown in FIG 14.

Using the method described above, the critical frequency ω_0 is determined at various downstream locations from $x = 0.625$ to 3.25, as shown in FIG 7. It is seen that the $\mathrm{Im}(\omega)$

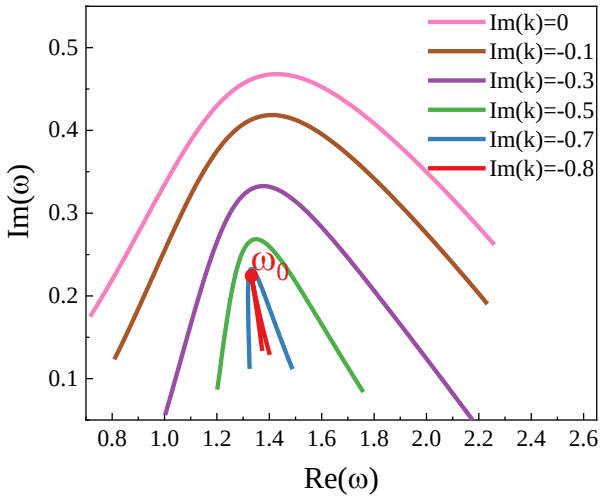

FIG. 14. Constant Im(k) lines in the ω plane at $x = 1$ for the uncontrolled case. The critical ω_0 is identified at the edge of the cusp-like trajectory.

transition from positive to negative occurs at about $x = 2.06$ in the uncontrolled case, which coincides with the length of recirculation area as can be read in FIG 13. This indicates that the flow in the recirculation area is absolutely unstable, and the flow becomes convectively unstable further downstream.

REFERENCES

[1]G. Dehnhardt, B. Mauck, W. Hanke, and H. Bleckmann, "Hydrodynamic trail-following in harbor seals (phoca vitulina)," Science **293**, 102 (2001).

[2]H. R. Beem and M. S. Triantafyllou, "Wake-induced 'slaloming' response explains exquisite sensitivity of seal whisker-like sensors," Journal of Fluid Mechanics **783**, 306–322 (2015).

[3]C. T. Murphy, C. Reichmuth, W. C. Eberhardt, B. H. Calhoun, and D. A. Mann, "Seal whiskers vibrate over broad frequencies during hydrodynamic tracking," Scientific Reports **7**, 8350 (2017).

[4]Y., Yang, J., Chen, J., Engel, S., Pandya, N., and Chen, "Distant touch hydrodynamic imaging with an artificial lateral line," Proceedings of the National Academy of Sciences (2006).

[5]Y. Jiang, Z. Ma, and D. Zhang, "Flow field perception based on the fish lateral line system," Bioinspiration & Biomimetics **14**, 041001 (2019).

[6]C. Qian, B. Zheng, Y. Shen, L. Jing, E. Li, L. Shen, and H. Chen, "Deep-learning-enabled self-adaptive microwave cloak without human intervention," Nature Photonics **14**, 383–390 (2020).

[7]S. Zhang, C. Xia, and N. Fang, "Broadband acoustic cloak for ultrasound waves," Phys. Rev. Lett. **106**, 024301 (2011).

[8]Y. A. Urzhumov and D. R. Smith, "Fluid flow control with transformation media," Physical Review Letters **107**, 074501 (2011).

[9]J. Park, J. R. Youn, and Y. S. Song, "Hydrodynamic metamaterial cloak for drag-free flow," Physical Review Letters **123**, 074502 (2019).

[10]E. Boyko, V. Bacheva, M. Eigenbrod, F. Paratore, A. D. Gat, S. Hardt, and M. Bercovici, "Microscale hydrodynamic cloaking and shielding via electro-osmosis," Physical Review Letters **126**, 184502 (2021).

[11]S. Zou, Y. Xu, R. Zatianina, C. Li, X. Liang, L. Zhu, Y. Zhang, G. Liu, Q. H. Liu, H. Chen, and Z.Wang, "Broadband waveguide cloak for water waves," Physical Review Letters **123**, 074501 (2019).

[12]Z. Zhang, S. Liu, Z. Luan, Z. Wang, and G. He, "Invisibility concentrator for water waves," Physics of Fluids **32**, 081701 (2020).

[13]B. Lusch, J. N. Kutz, and S. L. Brunton, "Deep learning for universal linear embeddings of nonlinear dynamics," Nature Communications **9**, 4950 (2018).

[14]F. Ren, C. Wang, and H. Tang, "Active control of vortex-induced vibration of a circular cylinder using machine learning," Physics of Fluids **31**, 093601 (2019).

[15]D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. V. D. Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, and M. Lanctot, "Mastering the game of go with deep neural networks and tree search," Nature **529**, 484–489 (2016).

[16]D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, and D. Hassabis, "Mastering the game of go without human knowledge," Nature **550**, 354–359 (2017).

[17]M. Volodymyr, K. Koray, S. David, R. Andrei A, V. Joel, B. Marc G, G. Alex, R. Martin, F. Andreas K, and O. Georg, "Human-level control through deep reinforcement learning," Nature **518**, 529–533 (2019).

[18]S. Verma, G. Novati, and P. Koumoutsakos, "Efficient collective swimming by harnessing vortices through deep reinforcement learning," Proceedings of the National Academy of Sciences **115**, 5849–5854 (2018).

[19]G. Reddy, A. Celani, T. J. Sejnowski, and M. Vergassola, "Learning to soar in turbulent environments," Proceedings of the National Academy of Sciences of the United States of America , E4877 (2016).

[20]S. Colabrese, K. Gustavsson, A. Celani, and L. Biferale, "Flow navigation by smart microswimmers via reinforcement learning," Physical Review Letters **118**, 158004 (2017).

[21]J. Rabault, M. Kuchta, A. Jensen, U. Réglade, and N. Cerardi, "Artificial neural networks trained through deep reinforcement learning discover control strategies for active flow control," Journal of Fluid Mechanics **865**, 281–302 (2019).

[22]H. Tang, J. Rabault, A. Kuhnle, Y. Wang, and T. Wang, "Robust active flow control over a range of reynolds numbers using an artificial neural network trained through deep reinforcement learning," Physics of Fluids **32**, 053605 (2020).

[23]F. Ren, J. Rabault, and H. Tang, "Applying deep reinforcement learning to active flow control in weakly turbulent conditions," Physics of Fluids **33**, 037121 (2021).

[24]R. Paris, S. Beneddine, and J. Dandois, "Robust flow control and optimal sensor placement using deep reinforcement learning," Journal of Fluid Mechanics **913**, A25 (2021).

[25]S. L. Brunton, B. R. Noack, and P. Koumoutsakos, "Machine learning for fluid mechanics," Annual Review of Fluid Mechanics **52**, 477–508 (2020).

[26]J. Rabault, F. Ren, W. Zhang, H. Tang, and H. Xu, "Deep reinforcement learning in fluid mechanics: A promising method for both active flow control and shape optimization," Journal of Hydrodynamics **32**, 234–246 (2020).

[27]F. Ren, H.-B. Hu, and H. Tang, "Active flow control using machine learning: A brief review," Journal of Hydrodynamics **32**, 247–253 (2020).

[28]W. Zhang, X. Li, Z. Ye, and Y. Jiang, "Mechanism of frequency lock-in in vortex-induced vibrations at low reynoldsnumbers," Journal of Fluid Mechanics **783**, 72–102 (2015).

[29]J. Zhou, R. J. Adrian, S. Balachandar, and T. Kendall, "Mechanisms for generating coherent packets of hairpin vortices in channel flow," Journal of fluid mechanics **387**, 353–396 (1999).

[30]S. Dong, G. S. Triantafyllou, and G. E. Karniadakis, "Elimination of vortex streets in bluff-body flows," Physical Review Letters **100**, 204501 (2008).

[31]C. Wang, H. Tang, S. C. M. Yu, and F. Duan, "Active control of vortex-induced vibrations of a circular cylinder using windward-suction-leeward-blowing actuation," Physics of Fluids **28**, 053601 (2016).

[32]W. Chen, Y. Liu, F. Xu, H. Li, and H. Hu, "Suppression of vortex shedding from a circular cylinder by using a traveling wave wall," 51st AIAA Aerospace Sciences Meeting (2013).

[33]F. Ren, B. Song, Y. Zhang, and H. Hu, "A gpu-accelerated solver for turbulent flow and scalar transport based on the lattice boltzmann method," Computers & Fluids **173**, 29–36 (2018).

[34]J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv preprint arXiv:1707.06347 (2017).

[35]C. Wang, H. Tang, S. C. M. Yu, and F. Duan, "Lock-on of vortex shedding to a pair of synthetic jets with phase difference," Physical Review Fluids **2**, 104701 (2017).

[36]C. Wang, H. Tang, S. C. M. Yu, and F. Duan, "Control of vortex-induced vibration using a pair of synthetic jets: Influence of active lock-on," Physics of Fluids **29**, 083602 (2017).

[37]C. Wang, H. Tang, F. Duan, and S. C. M. Yu, "Control of wakes and vortex-induced vibrations of a single circular cylinder using synthetic jets," Journal of Fluids & Structures **60**, 160–179 (2016).

[38]G. S. Triantafyllou, M. S. Triantafyllou, and C. Chryssostomidis, "On the formation of vortex streets behind stationary cylinders," Journal of Fluid Mechanics **170**, 461–477 (1986).

[39]K. Kupfer, A. Bers, and A. K. Ram, "The cusp map in the complex-frequency plane for absolute instabilities," Physics of Fluids **30**, 3075–3082 (1987).

[40]B. Thiria and J. E. Wesfreid, "Stability properties of forced wakes," Journal of Fluid Mechanics **579**, 137–161 (2007).

[41]J. Rabault and A. Kuhnle, "Accelerating deep reinforcement learning strategies of flow control through a multi-environment approach," Physics of Fluids **31**, 094105 (2019).

[42]D. D'Humières, I. Ginzburg, M. Krafczyk, P. Lallemand, and L. S. Luo, "Multiple-relaxation-time lattice boltzmann models in three dimensions," Philosophical Transactions of the Royal Society A **360**, 437–451 (2002).

[43]X. He and L. S. Luo, "Lattice boltzmann model for the incompressible navier–stokes equation," Journal of Statistical Physics **88**, 927–944 (1997).

[44]D. Yu, R. Mei, and S. Wei, "A multi-block lattice boltzmann method for viscous fluid flows," International Journal for Numerical Methods in Fluids **39** (2002).

[45]Ladd, Anthony, J., and C., "Numerical simulations of particulate suspensions via a discretized boltzmann equation. part 1. theoretical foundation," Journal of Fluid Mechanics

271, 285–285 (1994).

[46]A. Fakhari and T. Lee, "Finite-difference lattice boltzmann method with a block-structured adaptive-mesh-refinement technique," Physical Review E **89**, 033310 (2014).

[47]D. Yu, R. Mei, L. S. Luo, and S. Wei, "Viscous flow computations with the method of lattice boltzmann equation," Progress in Aerospace Sciences **39**, 329–367 (2003).

[48]A. Caiazzo, "Analysis of lattice boltzmann nodes initialisation in moving boundary problems," Progress in Computational Fluid Dynamics **8**, 3–10 (2008).

[49]Y. Chen, Q. Cai, Z. Xia, M. Wang, and S. Chen, "Momentum-exchange method in lattice boltzmann simulations of particle-fluid interactions," Physical Review E **88**, 013303 (2013).

[50]S. Tao, J. Hu, and Z. Guo, "An investigation on momentum exchange methods and refilling algorithms for lattice boltzmann simulation of particulate flows," Computers & Fluids **133**, 1–14 (2016).

[51]D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Arxiv (2014).

[52]R. J. Deissler, "The convective nature of instability in plane poiseuille flow," Physics of Fluids **30**, 2303–2305 (1987).