

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0303905

Deep Reinforcement Learning in Action: Real-Time Control of Vortex-Induced Vibrations

Hussam Sababha,¹ Bernat Font,² and Mohammed Daqaq^{1,3}

¹*Division of Engineering, New York University Abu Dhabi*

²*Mechanical Engineering, Delft University of Technology*

³*Department of Mechanical Engineering, Tandon School of Engineering, NYU*

(*Electronic mail: haa385@nyu.edu)

(Dated: December 5, 2025)

This study showcases an experimental deployment of deep reinforcement learning (DRL) for active flow control (AFC) of vortex-induced vibrations (VIV) in a circular cylinder at a high Reynolds number ($Re = 3000$) using rotary actuation. Departing from prior work that relied on low-Reynolds-number numerical simulations, this research demonstrates real-time control in a challenging experimental setting, successfully addressing practical constraints such as actuator delay. When the learning algorithm is provided with state feedback alone (displacement and velocity of the oscillating cylinder), the DRL agent learns a low-frequency rotary control strategy that achieves up to 80% vibration suppression which leverages the traditional *lock-on* phenomenon. While this level of suppression is significant, it remains below the performance achieved using high-frequency rotary actuation. The reduction in performance is attributed to actuation delays and can be mitigated by augmenting the learning algorithm with past control actions. This enables the agent to learn a high-frequency rotary control strategy that effectively modifies vortex shedding and achieves over 95% vibration attenuation. These results demonstrate the adaptability of DRL for AFC in real-world experiments and its ability to overcome instrumental limitations such as actuation lag.

I. INTRODUCTION

Vortex-induced vibration (VIV) of elastic bodies in cross-flow is a widely encountered and practically important fluid-structure interaction phenomenon. It arises when vortices are shed from an elastic body at a frequency close to its natural frequency, enabling effective energy transfer from the flow and into the body. This resonance phenomenon, also known as *lock-in*, can incite large-amplitude oscillations, potentially compromising the performance and structural integrity of many engineering systems^{1,2}. Notable examples of such adverse effects include, but are not limited to, vibration of heat exchanger tubes³, fatigue of marine risers in offshore operations⁴, and wind-induced oscillations in tall buildings and long-span bridges⁵.

The practical significance of VIV has resulted in an extensive body of research aimed at both elucidating its underlying physics and developing strategies to mitigate it^{1,6,7}. This includes ongoing efforts to devise effective passive and active control techniques^{8–11}. Passive control strategies seek to influence flow separation and wake development by modifying the geometry or surface characteristics of the structure. Common approaches include the addition of helical strakes¹¹, fairings¹², splitter plates⁹, and tripping wires¹⁰. The main advantage of passive strategies is their simplicity and independence from external energy input, which makes them cost-effective and relatively easy to implement. However, such methods rarely eliminate VIV entirely and often introduce undesirable trade-offs, such as increased drag, which may negatively impact the overall system performance¹³.

Given the inherent limitations of passive control methods, considerable research has also been directed toward active flow control techniques, which utilize external energy input to directly influence the vortex shedding dynamics.

Such approaches aim to either weaken the strength of vortex shedding¹⁴, or to shift the vortex shedding frequency away from the natural frequency of the structure. The latter strategy, which concerns shifting the vortex shedding frequency, was first demonstrated on a rigidly mounted cylinder, and involved forcing the cylinder to rotate at an angular speed, Ω , which varies sinusoidally such that¹⁵,

$$\Omega(t) = \Omega_0 \sin(2\pi f_r t), \quad (1)$$

where Ω_0 is the magnitude of the angular velocity and f_r is its frequency. It was shown that vortex shedding can synchronize or *lock-on*¹⁶ to the imposed sinusoidal rotary motion^{17,18}, and that the *lock-on* occurs across a wide range of forcing parameters leading to significant reduction in drag and transverse fluid forces¹⁹. More recent studies have extended the implementation of *lock-on* control to the suppression of VIV in elastically-mounted cylinders^{20,21}. It was observed that, within the *lock-on* regime, significant reduction in the amplitude of transverse VIV can be achieved, although notable exceptions arise when the frequency, f_r , of the imposed rotary motion is close to the natural frequency of the elastically-mounted cylinder, in which case large VIV amplitudes may still persist²⁰. It was also observed that the *lock-on* regime widens as the magnitude of the prescribed angular velocity, Ω_0 , increases.

Some researchers have also investigated closed-loop control strategies, in which feedback signals acquired using sensors are used to dynamically adapt the actuation of the elastic structure in a way that suppresses VIV. Compared to the more extensively studied open-loop approaches, closed-loop implementations remain relatively rare and are often limited to variants of the classical proportional-integral-derivative (PID) framework^{22,23}.

In parallel with these developments, machine learning, par-

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0303905

ticularly deep reinforcement learning (DRL), has recently emerged as a promising alternative to discover effective active flow control (AFC) strategies. Unlike traditional control methods, DRL enables environment-driven learning and adaptation without requiring explicit modeling of the complex governing equations. Along this line, Rabault *et al.*²⁴ reported the first application of an artificial neural network trained with a DRL agent for AFC. In their numerical study, which involved a 2D simulation of a circular cylinder at low Reynolds number ($Re = 100$), the DRL effectively learned a control policy, which manipulates the mass flow rates of two jets positioned on either side of the cylinder in such a way that stabilizes the vortex street and achieves an 8% reduction in drag with minimal actuation effort. Since then, DRL has been successfully applied to a number of AFC problems including drag reduction^{25–30}, collective swimmers and gliding control^{31,32}, separation control^{33,34}, and heat transfer³⁵, among others. The reader is referred to^{36,37} for comprehensive reviews on DRL for AFC.

Recent progress has also demonstrated the strong potential of DRL for mitigating VIV. In one numerical study, Zheng *et al.*³⁸ compared DRL and Gaussian-process-based active learning for controlling VIV in a circular cylinder using jet actuation at a low Reynolds number ($Re = 100$). While active learning achieved moderate vibration reduction, the DRL approach, employing a soft actor-critic algorithm, reduced vibration amplitudes by over 80%. In another also numerical study, Chen *et al.*³⁹ applied DRL to suppress the VIV of a square cylinder using synthetic jets positioned on different sides of the cylinder. They achieved a vibration reduction of up to 96% and demonstrated that DRL agents can adapt actuation strategies to actuator placement and reduced velocity conditions.

Of particular relevance to the present study is the numerical investigation of Ren *et al.*⁴⁰, who applied a DRL algorithm to uncover a new control strategy that imposes rotation on a circular cylinder to suppress VIV under *lock-in* conditions. Using only partial observations from the cylinder kinematics (displacement, velocity, and acceleration), as state inputs, the DRL agent was able to successfully reduce the amplitude of VIV by up to 99.6%. Interestingly, the mechanism learned by the DRL differs from the well-established *lock-on* effect, highlighting that DRL is capable of discovering novel strategies for attenuation. More recently, Ren *et al.*⁴¹ extended DRL-based AFC to wake-induced vibrations in tandem and staggered cylinder configurations. Using rotary actuation, the control mechanism based on the DRL agent achieved vibration reductions exceeding 95% across various flow conditions.

While previous numerical studies have demonstrated the effectiveness of DRL for VIV suppression, these efforts have been limited to simulated environments and low Reynolds numbers, e.g. $Re \leq 300$. On one hand, simulations provide absolute control on the environment and system's state, but the high computational cost of DRL for scale-resolving simulations is currently a computational bottleneck preventing high-Reynolds-number investigations³³. As a result, the applicability of DRL-based control strategies to VIV suppression for moderate-to-high Reynolds-number flows remains unex-

plored. On the other hand, experimental implementation introduces practical challenges that are typically absent in simulations, including communication and actuation delays, sensor noise, and limited control bandwidth. These factors can significantly affect learning dynamics and may lead to control strategies that differ from those developed in idealized settings. Therefore, experimental validation is a critical step toward assessing the robustness and generalizability of DRL in VIV applications.

To address this gap, the present work experimentally investigates DRL-based flow control for mitigating VIV of a circular cylinder using rotary actuation. Conducting experiments at high Reynolds numbers, $Re = 3000$, enhances the relevance of the results to real-world fluid systems, which predominantly operate in turbulent regimes. Unlike earlier studies that primarily controlled the cylinder's rotational velocity, the present approach directly modulates the actuator voltage via pulse-width modulation (PWM), thereby eliminating an intermediate control step and allowing the DRL agent to learn the voltage–rotation mapping implicitly. Beyond evaluating vibration suppression performance, we highlight the critical influence of actuator delay by contrasting policies trained with and without delay consideration, underscoring its significance for reliable flow control. Collectively, these contributions advance learning-based flow control by demonstrating the feasibility of translating DRL policies to real-time experimental hardware in high Reynolds number VIV applications.

The remainder of this paper is organized as follows. Section II describes the problem statement. Section III illustrates the experimental setup, including the fluid–structure apparatus, sensor, actuation system, and validation against published results. Section IV details the deep reinforcement learning framework, including the state and action representations, reward formulation, and training procedure. Section V presents the experimental findings of the implementation of the DRL. Finally, Section VI summarizes the key outcomes and outlines directions for future research in experimental DRL-based flow control.

II. PROBLEM STATEMENT

We consider the system illustrated in Fig. 1, which consists of a cylinder of diameter, D , and mass, M , elastically mounted on a spring with stiffness, K . The cylinder is free to oscillate transversely, i.e., in a direction perpendicular to a uniform incoming flow with density, ρ , and steady velocity, V . For certain values of V , vortices are shed from the cylinder at a frequency near the natural frequency of the system, $f_n = (1/2\pi)\sqrt{K/M}$. This frequency matching leads to resonant pressure forces that drive the cylinder into large-amplitude oscillations, denoted by $Y(t)$, a phenomenon commonly referred to in the literature as vortex-induced vibration.

Our goal here is to use DRL to discover a control policy that imposes a rotation $\Omega(t)$ (recall Equation (1)) on the cylinder such that the amplitude of the VIV is suppressed. This will be realized in an experimental setting using only sensor measurements of $Y(t)$ and its derivatives, and the actuation force

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0303905

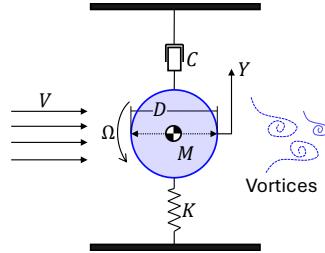


Figure 1. A schematic representation of a circular cylinder of diameter, D , subjected to a uniform flow of velocity, V . The cylinder is free to oscillate in the transverse, Y , direction and can actively rotate about its axis with angular velocity, Ω . The dynamics of the cylinder is lumped into a single-degree-of-freedom model involving a lumped mass, M , stiffness, K , and damping coefficient, C .

itself when needed. The DRL has no knowledge of the flow parameters, or the lift forces which are a complex function of time, t , $\Omega(t)$, $Y(t)$ and its derivatives.

III. EXPERIMENTAL SYSTEM

To achieve the goal of this study, an experimental system was created to mimic the conditions stated in the problem statement. The setup is shown in Fig. 2 and consists of four main components: (*i*) the test cylinder and the mounting mechanism, (*ii*) the actuation mechanism, (*iii*) the measurement system, and the (*iv*) testing environment.

A. Test cylinder and its mounting mechanism

The test cylinder denoted by 1 in Fig. 2 is machined from Aluminum and has a diameter $D = 17.5 \pm 0.01$ mm and an immersed length of $L = 160$ mm, yielding an aspect ratio $L/D \approx 9.14$. The cylinder is free on one end (the immersed end) and mounted on the other end onto a custom housing using two precision rotary bearings “2” as shown in the figure. The rotary bearings are rigidly mounted to a custom housing using alignment rods, which carry the weight and fix the cylinder along a defined rotational axis. The mounted end of the test cylinder is connected to a motor “4” using a flexible coupler “3”, which mitigates any residual misalignment between the motor shaft and the rotating cylinder. The whole system including the test cylinder, the rotary motor, the rotary bearings and the housing are mounted on a linear track using a linear air bearing “7”. The air-bearing ensures low structural damping while constraining the motion solely in the cross-flow direction. The air-bearing is connected on either side to two linear springs “10” that provided the elasticity necessary to incite VIV.

To minimize end effects and encourage uniform vortex shedding, the gap between the cylinder's end and the platform is set to approximately 1 mm. This distance is chosen

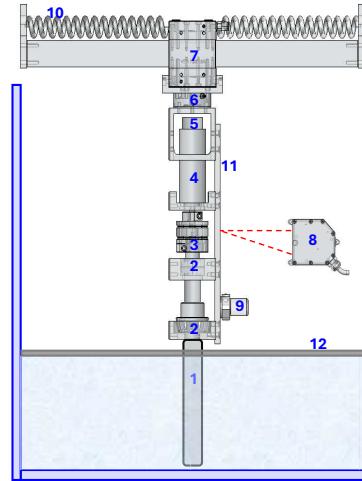


Figure 2. Schematic of the experimental setup with numbered components: 1 Test cylinder, 2 Precision rotary bearings, 3 Flexible coupler, 4 DC motor, 5 Rotary encoder, 6 Force sensor controller, 7 Linear air bearing, 8 Laser displacement sensor, 9 Accelerometer, 10 Linear springs, 11 Precision connecting rod, 12, Plexiglass sheet.

based on prior studies to minimize downwash effects and end wall interference, consistent with the setup used by Zhao et al.⁴². The total mass of the oscillating system, encompassing the cylinder, motor assembly, and movable parts of the air-bearing system, is $M = 1095$ g. The displaced fluid mass is $M_d = 360.07$ g, resulting in a mass ratio $m = M/M_d = 30.35$. Free-decay tests conducted in both air and still water were used to determine the natural frequencies and structural damping. The natural frequency in quiescent water was measured at $f_n = (1/2\pi)\sqrt{K/M} = 1.96$ Hz, and the structural damping ratio in air was found to be $\zeta_{\text{air}} = C/(2\sqrt{KM}) \approx 1.02 \times 10^{-2}$.

B. Actuation mechanism

The cylinder is actuated to rotate around its minor axis using a brushed DC motor with a rated torque of 0.24 kg·cm. The DC motor is actuated using a DCC1002 motor controller (Phidget). The controller drives the motor using a pulse-width modulation (PWM) signal, where the duty cycle determines the applied voltage and thereby regulates the motor speed and direction. For system integration, the controller is interfaced through the Phidget library in Python, which provides a dedicated class straightforward instantiation of motor objects, parameter configuration, and execution of control commands.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0303905

D. Testing environment

All tests are performed in the recirculating free-surface water channel at the Laboratory for Applied Nonlinear Dynamics (LAND), see Fig. 3. The test section of the water channel measures 400 mm in width, 500 mm in depth, and 6000 mm in length. The free-stream velocity can be continuously varied within the range $0.056 \leq V \leq 0.45$ m/s, and the turbulence intensity in the free stream is maintained below 1%.

C. Measurement system

The measurement system is designed to capture the dynamic response of the rotating cylinder with high accuracy. Transverse body displacement was recorded using a laser range sensor (Micro-Epsilon) denoted by “8” in Fig. 2 with a measurement range of ± 0.1 mm. The digital output of the sensor is converted into an analog voltage signal via a C-Box signal converter, making it compatible with the acquisition system. Local acceleration of the body is measured using a piezoelectric accelerometer, denoted by “9” in Fig. 2, which is connected to a Dytran signal conditioning unit for filtering, thereby producing an analog voltage output. Rotational motion is monitored using a digital rotary encoder (E5-1000, US Digital, USA), whose quadrature pulses are converted into an analog signal before acquisition (denoted by “5” in Fig. 2). All measurement signals are thus available in analog form and are routed to NI6281 data acquisition board (National instruments). Finally, the NI DAQ board is interfaced with Python through a dedicated class, allowing motor actuation and sensing to be integrated and executed within a single environment. This platform serves as the agent in the DRL framework, directly interacting with the experimental system by executing actions and receiving feedback for learning.

The EduPIV system (Dantec Dynamics) is employed to capture the flow structures in the near wake of the rotating cylinder. The flow is seeded with polyamic tracer particles of nominal diameter 50 μm , which are illuminated by a high-intensity LED light source. Imaging is performed using a high-speed camera with a resolution of 1960×600 pixels. To mitigate disturbances from the free surface and ensure image clarity, a plexiglass sheet is placed above the measurement section to stabilize the flow. See Fig. 3 for details of the PIV system. Images are acquired at a frame rate of 310 frames per second. For each measurement, 1000 images are recorded. The velocity field is computed using adaptive PIV in DynamicStudio (Dantec Dynamics). The algorithm is an iterative, multi-pass approach, where the interrogation area (IA) is adaptively reshaped according to local velocity gradients corresponding to image deformation. The interrogation area is set with a minimum size of 64x64 pixels and a maximum size of 128x128 pixels, and the grid step size is 32x32 pixels, resulting in a 50% overlap between adjacent measurement points.

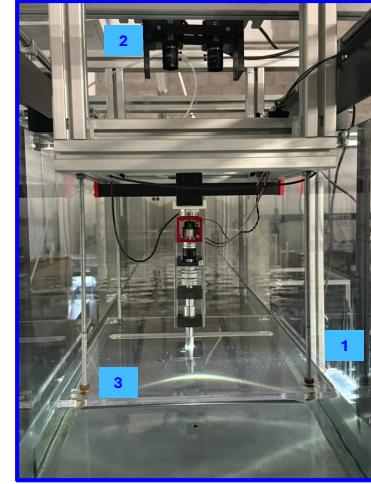


Figure 3. Experimental setup of the rotating cylinder within the flume. Key components of the EduPIV system, including 1 the LED illumination, 2 high-speed camera, and 3 plexiglass flow stabilizer are also shown.

E. Experimental verification

Before we delve into the design of the DRL controller, we verify that the experimental system described in section III A does indeed undergo VIV in the absence of a prescribed rotary actuation. To this end, we first study variation of the dimensionless steady-state amplitude, A/D , of the cylinder as the reduced flow speed, $U = V/(f_n D)$ is varied. It can be clearly seen in Fig. 4(a) that the steady-state response exhibits the typical *lock-in* phenomenon, where large-amplitude responses occur over a range of reduced wind speeds that correspond to vortex shedding frequencies close to the natural frequency of the oscillator. The response clearly exhibits the typical characteristics of the *lock-in* phenomenon with two distinct branches of solution: the initial branch, which exists for $4.5 \leq U \leq 5.1$, the lower branch , which covers the range $5.1 \leq U \leq 9.0$. The peak amplitude response is observed to reach $A/D \approx 0.6$ in the present experiments, compared with $A/D \approx 0.95$ in Khalak and Williamson⁴³ and $A/D \approx 1.0$ in Zhao et al.⁴⁴. The reduction in the peak steady-state amplitude is expected considering the influence of the mass ratio, which is inherently higher in this setup due to the additional weight from the motor assembly.

As an additional confirmation of our findings that the cylinder oscillations are indeed due to VIV, we plot the peak amplitude A^*/D of the steady-state response against the Skop-Griffin number, SG , of the system defined as^{45,46}

$$SG = 2\pi^3 St^2 (1 + m) \zeta, \quad (2)$$

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0303905

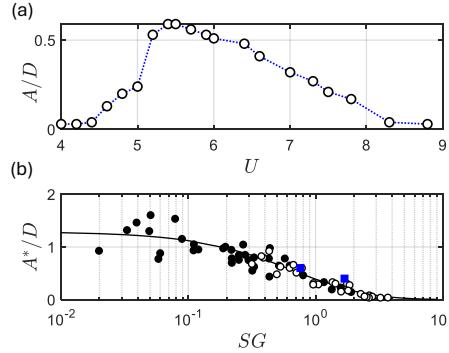


Figure 4. (a) Variation of the normalized steady-state VIV amplitude, A/D , with the reduced velocity, U , highlighting the typical *lock-in* curve. (b) Effect of the Skop-Griffin number, SG , on the peak response amplitude, A^*/D , within the lock-in region. Square (blue) markers denote the present experimental results. The solid line corresponds to the model of Griffin and Ramberg⁴⁷, while the solid and hollow circles represent data obtained from Paidoussis et al.⁴⁸ in water and air, respectively.

where $St \approx 0.21$ is the Strouhal number of the cylinder at large Reynolds numbers, $m = 30.3$ is the mass ratio of the cylinder and $\zeta = 0.012$ is the damping ratio of the system. As shown in Fig. 4(b), we can clearly see that the coordinate of the resulting point falls nicely on the Skop-Griffin universal curve further confirming that the oscillations are indeed due to VIV. The process is repeated for a different cylinder with a mass ratio $m = 40.1$ and a damping ratio $\zeta = 0.018$. Again, the coordinate of the resulting point falls on the Skop-Griffin universal curve.

IV. DEEP REINFORCEMENT LEARNING

Unlike traditional active control methods in fluid mechanics, which typically rely on reduced-order models or prior knowledge of the governing equations, deep reinforcement learning (DRL) enables the design of controllers without requiring an explicit model of the flow physics. In this study, the agent, implemented as a computer program, interacts iteratively with the experimental setup and adapts its control strategy to minimize vibrations. The control agent is trained using the Proximal Policy Optimization algorithm (PPO) implemented in Stable-Baseline3 (SB3 – version 2.6.0) built on PyTorch (version 2.7.0). The training environment is custom-developed in Python (version 3.10.0), to interface directly with the experimental hardware.⁴⁹.

The PPO agent samples a sequence of states, s , actions, a and rewards, r , represented by η defined as

$$\eta = (s_1, a_1, r_1), (s_2, a_2, r_2), \dots, (s_N, a_N, r_N), \quad (3)$$

where N is the number of times the DRL agent applies a policy $\pi_\theta(a_t, s_t)$ in each episode. Then, the agent pauses learning and updates the policy after each trajectory using the accumulated experience. The state space consists of the kinematic variables of the oscillating cylinder, represented by Y/D and $\dot{Y}/2\pi f_n D$. The velocity Y is computed from the finite difference of consecutive displacement measurements and subsequently filtered using an exponential moving average (EMA) with a smoothing factor of $\beta = 0.2$ to mitigate sensor noise. The action space consists of one action corresponding to the voltage duty cycle applied to the motor controller which can take a value between $[-0.4, 0.4]$. The reward aims to mitigate the cylinder displacement (vibrations) and is defined as

$$r = -|Y/D|. \quad (4)$$

Figure 5 presents the network architectures of the standard PPO method. The algorithm comprises two neural networks: the *actor* and the *critic*. Both networks receive the same state vector as input and share a similar architecture, consisting of an input layer, two fully connected hidden layers with 64 units each, and an output layer. The critic network outputs an estimate of the state value function $V^\pi(s)$, while the actor network outputs the agent's policy π_θ distribution, i.e. the control strategy.

In the present study, the duration of each training episode is set to $T = 25T_0$, where T_0 is the natural period of the cylinder in the uncontrolled case. With $T_0 \approx 0.5$ s, the total episode duration is $T \approx 12.5$ s. Within each episode, the agent interacts with the environment 128 times, producing a sequence of flow states and control actions. The policy is updated at the end of each episode. Each new episode begins from the terminal physical state of the preceding episode. In other words, control (the first action of the new episode) resumes immediately without a recovery period. After training converges, we switch to a deterministic control phase in which the neural network weights remain fixed. During this stage, the actor network alone is used to map the observed state to the corresponding control action. The hyperparameters used in this study are listed in Table I.

Deploying DRL in real-time experiments introduces challenges that are absent in simulations. First, the motor controller hardware operates with a minimum command update interval of 100 ms, which imposes an upper limit on the control frequency. To accommodate this limitation, we deliberately select a system operating frequency around 1.96 Hz, corresponding to approximately five control actions per oscillation period. Second, the motor exhibits an actuation delay: preliminary tests showed a lag of roughly 200 ms between the application of a voltage input and the attainment of its maximum rotational speed. Such delays can complicate reinforcement learning because the executed command is not immediately reflected in the observed state, thereby violating the Markov property assumed in classical Markov Decision Processes⁵⁰. In practice, due to these delays and sensor limitations, the agent perceives only a partial view of the system state, and the problem is more accurately described as a partially observable Markov decision process (POMDP)⁵¹. In-

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0303905

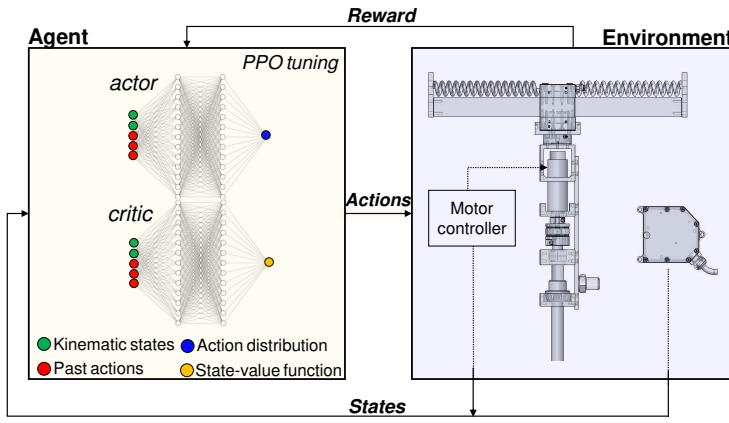


Figure 5. Illustration of the deep reinforcement learning framework. The environment is coupled with the learning agent

Table I. Hyperparameters used in the PPO implementation.

Parameter	Value / Setting
Policy network architecture	64×64
Value network architecture	64×64
Length of each episode	$25T_0$
Steps per episode (n_{steps})	128
Batch size	64
Number of epochs	10
Total training steps	5×10^4
Discount factor (γ)	0.99
GAE parameter (λ)	0.95
PPO clipping factor (ϵ)	0.2
action bounds	$[-0.4, 0.4]$
Optimizer	Adam
Learning rate	3×10^{-4}
Initializer	Orthogonal
Entropy coefficient	0.01

cluding past actions and past state measurements can mitigate the effects of partial observability by effectively forming a sufficient statistic that reduces the POMDP to a Markovian formulation⁵². This approach has demonstrated success across various domains, including robotics⁵³ and, more recently, active flow control^{29,54}. To compensate for actuator lag, we adopt a similar strategy by augmenting the state vector with past control inputs. Although incorporating past states measurements could further enhance the DRL training, preliminary tests indicated that augmenting past actions alone was sufficient to compensate for the actuation delay. Given that the motor requires approximately 200 ms (i.e., two sampling intervals) to reach full speed, the two most recent control actions are included in the state vector. This design enables the policy to associate the current kinematic states with the cumu-

lative influence of previous actions, thereby embedding both the delayed actuation dynamics and the effects of partial observability into the training framework.

V. RESULTS AND DISCUSSION

This section presents the experimental findings of the DRL-based rotary control framework. Here, we assess the performance of the DRL strategy when implemented using only state feedback and compare it to the well-known open-loop *lock-on* sinusoidal control strategy. We examine the effect of actuator delays within the control loop on the overall performance of the DRL control policy and discuss an approach to account for them.

A. DRL control policy

We first implement the DRL strategy on the experimental system assuming that the control action can be applied instantly without any actuator lag. In such a scenario, the state, action, and reward are directly coupled within each control cycle. Figure 6(a) shows how the training reward develops over successive episodes. Specifically, each episode consists of 128 environment interactions (steps), and the episode reward is computed by summing the instantaneous rewards at each step and dividing by the number of steps to obtain the mean reward. The reward initially takes on negative values, then rapidly increases, and eventually stabilizes, indicating that the DRL agent has converged and established a stable control policy. Convergence is observed after approximately 300 episodes, which corresponds to around 60 minutes of training. The solid line in the figure represents the mean reward across five independent training runs, while the shaded

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0303905

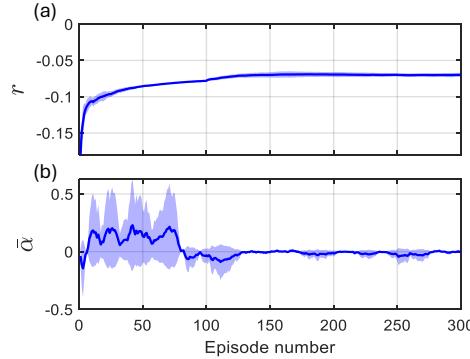


Figure 6. (a) Variation of the average reward, r , with the number of episodes. (b) Variation of the mean rotational speed, $\bar{\alpha}$, with the number of episodes

region denotes the corresponding standard deviation.

Evolution of the cylinder's normalized mean rotational speed defined as $\bar{\alpha} = \bar{\Omega}_0 D / (2U)$ ($\bar{\Omega}_0$ is the dimensional mean rotational speed), throughout the training process is shown in Fig. 6(b). The solid line represents the mean across five independent training runs, and the shaded area denotes the corresponding standard deviation. At the onset of training, $\bar{\alpha}$ deviates from zero, reflecting a directional bias in the initial control input. As training progresses however, the mean rotational speed gradually converges toward zero, indicating that the learned policy adopts a symmetric back-and-forth rotation of the cylinder. This oscillatory motion is crucial for maintaining the cylinder displacement near $Y = 0$, thereby stabilizing the structure and mitigating the VIV.

Upon convergence of the DRL training, the control policy learned by the "actor" network is extracted. A deterministic control experiment is then performed for a duration of $t = 50$ s. Figure 7 illustrates the control actions learned by the agent along with the corresponding normalized rotational speed, $\alpha(t) = \Omega(t)D/(2U)$, of the cylinder. We observe that the motor actuates in discrete steps, yet the resulting rotational speed exhibits a smooth sinusoidal profile. This indicates that the DRL agent successfully approximated the non-linear mapping between the voltage duty cycle of the motor and its resulting velocity through interaction with the environment without the need for complex system identification and transfer function modeling; a step that is necessary in conventional control design.

A closer examination of Fig. 7(a) reveals that the learned control policy operates in two distinct phases characterized by different actuation frequencies. Initially, the agent applies high-frequency periodic inputs, which gradually transition to a lower frequency over time. This behavior suggests that the agent prioritizes a rapid suppression of oscillations at the onset, before settling into a more stable, sustained control regime. Figure 8 compares the time history of the cylinder

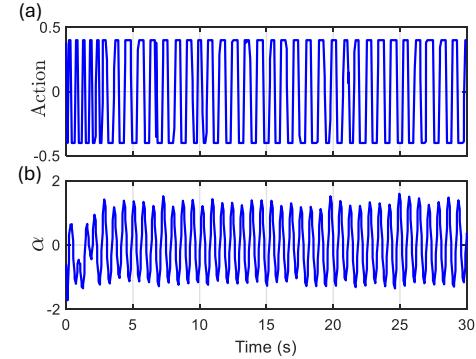


Figure 7. Results of the DRL-based deterministic control. (a) Control action prescribed by the agent, represented as the applied voltage duty cycle. (b) Resulting normalized motor rotational speed, α , as measured by the rotary encoder.

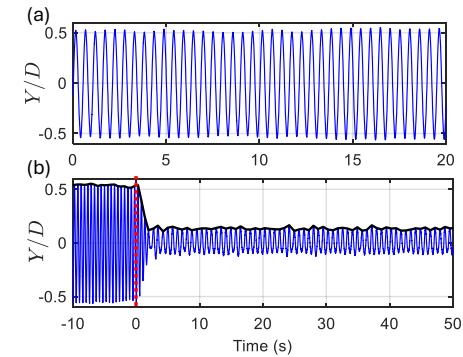


Figure 8. (a) Time history of the non-dimensional displacement, Y/D with no control. (b) Time history of the non-dimensional displacement, Y/D . Red-dashed line marks the onset of control actuation.

displacement, Y/D , for the uncontrolled and controlled cases. Figure 8(a) shows the baseline response without active flow control. In contrast, Fig. 8(b) shows the response under DRL-based control. It is clear that, once the control is activated (indicated by the dashed red line), the amplitude of VIV quickly drops to significantly lower levels, highlighting the impact of the initial high-frequency actuation. Over time, the response stabilizes at a lower frequency, below the natural frequency of the system. The post-control oscillation amplitude is approximately 80% of the uncontrolled case, highlighting the effectiveness of the learned control policy.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0303905

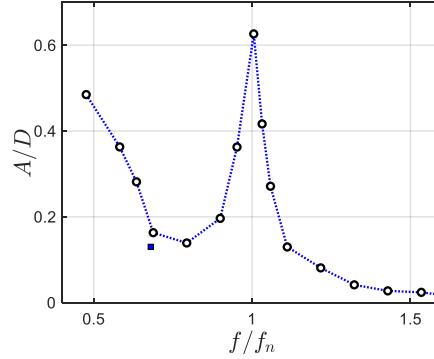


Figure 9. Amplitude response of the cylinder under sinusoidal actuation at varying frequency ratios. Square marker shows a comparison against the DRL-based control strategy.

B. Rotary sinusoidal lock-on control

Given that the learned DRL policy produces oscillatory control signals, it is natural to benchmark its performance against a traditional rotary sinusoidal *lock-on* control algorithm. To this end, a PID-based controller is implemented to actuate the cylinder to rotate sinusoidally at a normalized speed $\alpha(t) = \alpha_0 \cos(2\pi f_r t)$. The magnitude α_0 of the sinusoidal speed was kept constant at one to ensure consistency with the peak amplitude achieved by the DRL policy (see Fig. 7(b)). The frequency f_r of the rotational speed is varied in the range $0.4f_n \leq f_r \leq 1.6f_n$, where f_n is the natural frequency of the oscillating cylinder. The resulting magnitude of the steady-state VIV, A , normalized by the cylinder diameter, D , is recorded and plotted against f_r/f_n as shown in Fig. 9.

It is observed that as the normalized forcing frequency f_r/f_n increases from 0.4 to 0.8, the body begins to synchronize with the forcing frequency, resulting in a sharp reduction in the amplitude of the VIV response. This trend continues up to $f_r/f_n = 0.8$. Beyond this point, the amplitude rises rapidly with increasing f_r/f_n , reaching a peak value of $A/D = 0.65$ at $f_r/f_n = 1$. As the forcing frequency increases further, the steady-state amplitude of the VIV drops sharply, eventually falling below $A/D = 0.05$ at $f_r/f_n = 1.6$.

To evaluate the performance of the DRL policy in comparison to sinusoidal rotary forcing, a Fast Fourier Transform (FFT) is applied to the signal shown in Fig. 7(b) to determine the dominant frequency f_r of the rotational motion produced by the DRL controller. The corresponding point ($f_r/f_n, A/D$) is plotted as a blue circle on the frequency-response curve in Fig. 9. This point closely follows the sinusoidal response trend and aligns with the onset of the lock-on region, suggesting that the learned control strategy takes advantage of the lock-on mechanism, where vortex shedding synchronizes with the forcing frequency.

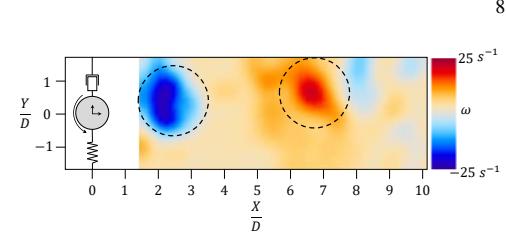


Figure 10. Instantaneous vorticity, ω , contours under the DRL-based control.

Additional evidence of this behavior is provided by PIV measurements of the wake behind the cylinder. As shown in Fig. 10, the vorticity contours display a wake pattern consisting of two oppositely-signed vortices shed per oscillation cycle, consistent with the classical 2S mode described in the literature¹. The shedding frequency matches the actuation frequency, indicating that the DRL-based controller effectively modifies the wake dynamics to achieve *lock-on* synchronization, thereby reducing the vibration amplitude.

C. Impact of actuator dynamics on DRL control

Direct implementation of the DRL policy using only state feedback results in clear attenuation of VIV, achieving approximately 80% suppression. However, this level of performance remains suboptimal when compared to the nearly complete suppression achieved by the sinusoidal rotary control strategy at higher frequency ratios (see Fig. 9). This raises the question: why does the DRL controller converge toward a low-frequency actuation policy, rather than adopting a high-frequency excitation that yields greater attenuation? We hypothesize that this discrepancy stems from actuator dynamics; specifically, delays in the motor response, which may hinder the agent's ability to effectively implement high-frequency control actions.

To address this limitation, the state vector is augmented with past motor actions, as described in Sec. IV. Training performance is evaluated for two cases: one without past actions ($n = 0$) and one with two past actions included in the state vector ($n = 2$). The corresponding reward histories are shown in Fig. 11(a), where the solid and dashed lines represent the mean values from five independent training runs for the cases without and with past actions, respectively. The shaded regions denote the standard deviation across runs.

It is evident that excluding past actions leads to lower rewards, indicating reduced capability in attenuating vibrations. In contrast, including two previous motor commands allows the agent to achieve consistently higher rewards, demonstrating improved VIV suppression performance.

To compare the characteristics of the control input with and without the inclusion of past actions, we analyze the dominant frequency component of the rotational speed throughout training for both cases, as shown in Fig. 11(b). Initially, both approaches exhibit high variance in the dominant fre-

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0303905

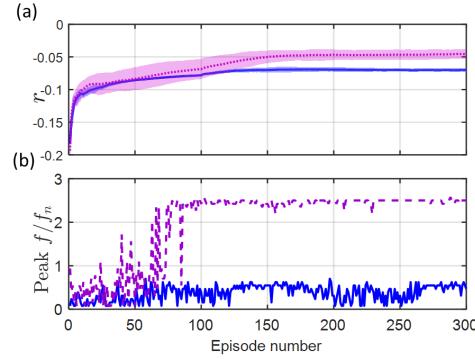


Figure 11. (a) Evaluation of the reward without augmenting past actions (solid line), and when augmenting two past actions in the state vector (dashed lines). (b) Variation of the dominant frequency of the rotary forcing with the episode number. No past actions (solid line), and two past actions (dashed lines).

frequency, reflecting the agent's exploration during early training stages. As training progresses, the variance decreases and the dominant frequencies begin to diverge. Without access to past actions, the dominant frequency ratio remains below 1. In contrast, when past motor actions are incorporated into the state, the dominant frequency ratio shifts to significantly higher values, exceeding $f_r/f_n = 2.5$. This indicates that including past actions enables the agent to explore and exploit higher-frequency control strategies, which correspond to substantially increased rewards, as shown in Fig. 11(a).

The learned control policy, which includes past actions, is deployed to attenuate the VIV of the cylinder by allowing the actor network to apply deterministic control over a 50 second duration. Figure 12(a) and (b) presents the time histories of the control actions learned by the agent and the corresponding normalized rotational speed, $\alpha(t) = \Omega(t)D/(2U)$, of the cylinder. The control action exhibits a stepwise variations, reflecting the discrete nature of the actuator commands. Compared with the case shown in Fig. 8, the actuation occurs at a considerably higher frequency, which is also evident in the oscillatory behavior of $\alpha(t)$. This rapid alternation in rotation direction enables the controller to maintain the cylinder displacement near equilibrium, thereby mitigating the VIV response. The time history of the resulting VIV response, shown in Fig. 13, clearly demonstrates that accounting for past actions leads to enhanced performance as the DRL agent is now able to achieve more than a 95% reduction in oscillation amplitude. The mitigation of vortex-induced vibration can be observed in the supplementary material video.

The corresponding wake vorticity contours in Fig. 14 reveal distinct characteristics. While the vortex shedding remains in a 2S mode, the vortices are now shed into two narrowly spaced rows due to the enhanced vibration suppression. This wake pattern closely follows mode I observed for a rigidly mounted

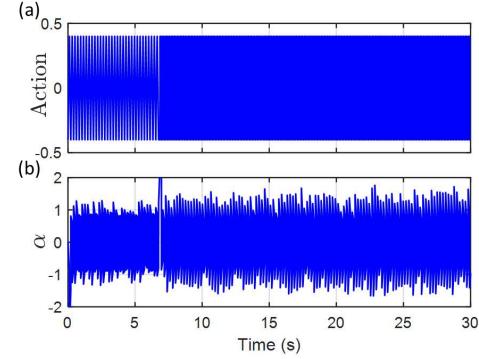


Figure 12. Results of the DRL-based deterministic control. (a) Control action prescribed by the agent, represented as the applied voltage duty cycle. (b) Resulting normalized motor rotational speed, α , as measured by the rotary encoder.

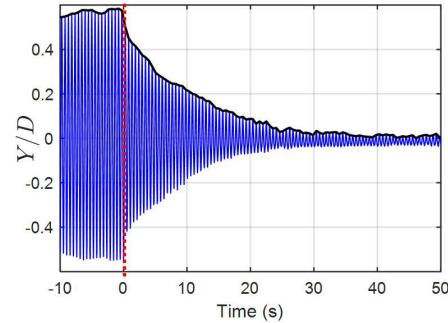


Figure 13. Time history of the non-dimensional displacement, Y/D , with past motor commands included in the state vector. Red-dashed line marks the onset of actuation.

cylinder at varying frequency ratios, as reported by Choi et al.⁵⁵.

The contrasting performance between the two strategies arises from how the agent interprets environmental information. In the first case, without memory, each state observation is linked only to its simultaneous action. Lacking the ability to explicitly model actuation delays, the policy converges to a slow, more conservative approach, producing low-frequency control inputs. In contrast, augmenting the input space with past actions provides the agent with an implicit memory, enabling it to infer the system's delayed response dynamics. This enhanced representation allows the policy to generate higher frequency control inputs that more effectively suppresses vibration.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0303905

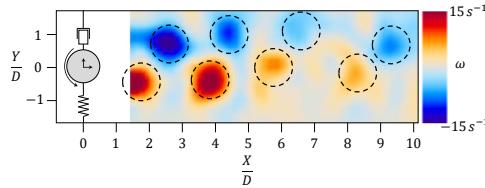


Figure 14. Instantaneous vorticity, ω , contours under the DRL-based control with past motor commands included in the state vector.

D. Comparison to a recent numerical study

An interesting distinction arises when comparing the present experimental results with recent numerical DRL studies. In our experiments, the learned controller suppresses VIV primarily through the *lock-on* mechanism, in which the actuation modifies the vortex shedding frequency and synchronizes the structural response with the forcing. In contrast, Ren *et al.*⁴⁰ reported that their DRL-guided rotary control achieved vibration suppression without altering the shedding frequency. Instead, the DRL policy stabilizes the flow by driving the growth rates of dominant modes to negative values. Several factors may account for this discrepancy. First, differences in Reynolds number (⁴⁰ simulated $Re = 100$ while our experiment is at $Re = 3000$) may influence the instability dynamics and thus the control pathways available to the DRL agent. Second, the numerical setup eliminates experimental complexities such as measurement noise or actuator delays, which are intrinsic to physical systems and may favor *lock-on* type strategies. Finally, the higher sampling and action frequencies employed in the simulations enable more rapid policy updates, which are constrained in our experimental study. These observations suggest that DRL can converge to distinct yet effective control strategies, highlighting a promising direction for future research in VIV control.

VI. CONCLUSIONS

This study presents an experimental implementation of deep reinforcement learning (DRL) for active flow control (AFC) of vortex-induced vibrations (VIV) in a circular cylinder at high Reynolds numbers ($Re = 3000$) using rotary control. Unlike previous studies which are numerical in nature and conducted at low Reynolds numbers, the present work demonstrates, for the first time, the successful deployment of DRL policies for real-time experimental control of VIV at a high Reynolds number, accounting for practical challenges such as actuation delays and limited sampling rates.

The key findings of this study can be summarized as follows:

- When incorporating state feedback only (displacement and velocity of the oscillating cylinder) into the learning algorithm, the DRL agent is able to discover a

10

low-frequency rotary control strategy that suppresses oscillations by up to 80% relative to the uncontrolled case. The learned DRL policy resembles the well-known *lock-on* phenomenon, which synchronizes vortex shedding with the actuation frequency. In the process, the DRL agent is able to directly map discrete voltage duty-cycle commands into rotary speed, thereby bypassing the need for an intermediate control model.

- Due to actuator delays, the policy learned using state feedback alone was limited to low-frequency control strategies, resulting in suboptimal performance compared to that enabled by high-frequency rotary actuation exploiting the lock-on phenomenon. To overcome this limitation, the state feedback vector is augmented with previous control actions, allowing the DRL agent to account for actuator dynamics and temporal dependencies. This augmentation enabled the agent to discover higher-frequency actuation strategies, achieving over 95% vibration suppression. These results highlight the critical role of temporal information in optimizing control performance under realistic actuation constraints.

We plan to extend the experimental DRL framework to explore energy harvesting strategies in vortex-induced vibrations. At the same time, we aim to perform experiments at higher Reynolds numbers to approach more realistic flow conditions. These efforts are intended to bridge the gap between the powerful capabilities of machine learning and the practical implementations in active flow control.

SUPPLEMENTARY MATERIAL

See the supplementary material for a video demonstrating the mitigation of vortex-induced vibration. The video illustrates the system response under the proposed control strategy and highlights the suppression of oscillatory motion compared to the uncontrolled case.

ACKNOWLEDGMENTS

The authors would like to thank Nikolas Giakoumidis (CAIR) for discussions on the setup design and Vijay Dhanvi (NYUAD Machine Shop) for apparatus fabrication.

DATA AVAILABILITY

The codes supporting the findings of this study are openly available https://github.com/Hussamsababha/Experimental_DRL-

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0303905

REFERENCES

- ¹C. H. K. Williamson and R. Govardhan, "Vortex-induced vibration," *Annual Review of Fluid Mechanics* **36**, 413–455 (2004).
- ²R. Bourguet, Y. Modarres-Sadeghi, G. E. Karniadakis, and M. S. Triantafyllou, "Wake-body resonance of long flexible structures is dominated by counterclockwise orbits," *Physical review letters* **107**, 134502 (2011).
- ³J. Shi and et al., "Numerical study of heat transfer enhancement of channel via vortex-induced vibration," *Applied Thermal Engineering* **70**, 838–845 (2014).
- ⁴A. D. Trim, H. Braaten, H. Lie, and M. A. Tognarelli, "Experimental investigation of vortex-induced vibration of long marine risers," *Journal of Fluids and Structures* **21**, 335–355 (2005).
- ⁵T. Wu and A. Kareem, "An overview of vortex-induced vibration (viv) of bridge decks," *Frontiers of Structural and Civil Engineering* **6**, 335–347 (2012).
- ⁶P. Bearman, "Circular cylinder wakes and vortex-induced vibrations," *Journal of Fluids and Structures* **27**, 648–658 (2011).
- ⁷T. Sarpkaya, "A critical review of the intrinsic nature of vortex-induced vibrations," *Journal of Fluids and Structures* **19**, 389–447 (2004).
- ⁸H. Choi, W.-P. Jeon, and J. Kim, "Control of flow over a bluff body," *Annu. Rev. Fluid Mech.* **40**, 113–139 (2008).
- ⁹S. Liang, J. Wang, and Z. Hu, "Viv and galloping response of a circular cylinder with rigid detached splitter plates," *Ocean Engineering* **162**, 176–186 (2018).
- ¹⁰F. S. Hover, H. Tvedt, and M. S. Triantafyllou, "Vortex-induced vibrations of a cylinder with tripping wires," *Journal of Fluid Mechanics* **448**, 175–197 (2001).
- ¹¹P. W. Bearman and M. Branković, "Experimental studies of passive control of vortex-induced vibration," *European Journal of Mechanics - B/Fluids* **23**, 9–27 (2004).
- ¹²Y. Yu, F. Xie, H. Yan, Y. Constantinides, O. Oakley, and G. E. Karniadakis, "Suppression of vortex-induced vibrations by fairings: a numerical study," *Journal of Fluids and Structures* **54**, 679–700 (2015).
- ¹³J. C. Owen, P. W. Bearman, and A. A. Szewczyk, "Passive control of viv with drag reduction," *Journal of Fluids and Structures* **15**, 597–605 (2001).
- ¹⁴C. Wang, H. Tang, S. Yu, and F. Duan, "Active control of vortex-induced vibrations of a circular cylinder using windward-suction-leeward-blowing actuation," *Physics of Fluids* **28** (2016).
- ¹⁵S.-J. Baek and H. J. Sung, "Numerical simulation of the flow behind a rotary oscillating circular cylinder," *Physics of Fluids* **10**, 869–876 (1998).
- ¹⁶The lock-on between the imposed sinusoidal rotary frequency and vortex shedding should not be confused with lock-in which refers to the synchronization between the body's natural frequency and vortex shedding.
- ¹⁷A. Okama, H. Takata, and T. Asanuma, "Viscous flow around a rotationally oscillating circular cylinder," (1975).
- ¹⁸S. Taneda, "Visual observations of the flow past a circular cylinder performing a rotatory oscillation," *Journal of the Physical Society of Japan* **45**, 1038–1043 (1978).
- ¹⁹P. Tokumaru and P. Dimotakis, "Rotary oscillation control of a cylinder wake," *Journal of Fluid Mechanics* **224**, 77–90 (1991).
- ²⁰L. Du and X. Sun, "Suppression of vortex-induced vibration using the rotary oscillation of a cylinder," *Physics of Fluids* **27** (2015).
- ²¹K. W. L. Wong, J. Zhao, D. L. Jacono, M. C. Thompson, and J. Sheridan, "Experimental investigation of flow-induced vibration of a sinusoidally rotating circular cylinder," *Journal of Fluid Mechanics* **848**, 430–466 (2018).
- ²²M. Zhang, L. Cheng, and Y. Zhou, "Closed-loop-controlled vortex shedding and vibration of a flexibly supported square cylinder under different schemes," *Physics of Fluids* **16**, 1439–1448 (2004).
- ²³D. Vicente-Llambay, A. Barrero-Gil, and A. Velazquez, "Flow-induced vibration control of a circular cylinder using rotational oscillation feedback," *Journal of Fluid Mechanics* **847**, 93–118 (2018).
- ²⁴J. 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).
- ²⁵D. Fan, L. Yang, Z. Wang, M. S. Triantafyllou, and G. E. Karniadakis, "Reinforcement learning for bluff body active flow control in experiments and simulations," *Proceedings of the National Academy of Sciences* **117**, 26091–26098 (2020).
- ²⁶R. Paris, S. Beneddine, and J. Dandois, "Robust flow control and optimal sensor placement using deep reinforcement learning," *Journal of Fluid Mechanics* **913**, A25 (2021).
- ²⁷L. Guastoni, J. Rabault, P. Schlatter, H. Azizpour, and R. Vinuesa, "Deep reinforcement learning for turbulent drag reduction in channel flows," *The European Physical Journal E* **46**, 27 (2023).
- ²⁸P. Suárez, F. Alcántara-Ávila, A. Miró, J. Rabault, B. Font, O. Lehmkuhl, and R. Vinuesa, "Active flow control for drag reduction through multi-agent reinforcement learning on a turbulent cylinder at $re=3900$," *Flow, Turbulence and Combustion* , 1–25 (2025).
- ²⁹C. Xia, J. Zhang, E. C. Kerrigan, and G. Rigas, "Active flow control for bluff body drag reduction using reinforcement learning with partial measurements," *Journal of Fluid Mechanics* **981**, A17 (2024).
- ³⁰Z. Wang, D. Fan, X. Jiang, M. S. Triantafyllou, and G. E. Karniadakis, "Deep reinforcement transfer learning of active control for bluff body flows at high Reynolds number," *Journal of Fluid Mechanics* **973** (2023), 101017/jfm.2023.637.
- ³¹S. 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).
- ³²G. Novati, L. Mahadevan, and P. Koumoutsakos, "Controlled gliding and perching through deep-reinforcement-learning," *Physical Review Fluids* **4**, 093902 (2019).
- ³³B. Font, F. Alcántara-Ávila, J. Rabault, R. Vinuesa, and O. Lehmkuhl, "Deep reinforcement learning for active flow control in a turbulent separation bubble," *Nature communications* **16**, 1422 (2025).
- ³⁴R. Montalà, B. Font, P. Suárez, J. Rabault, O. Lehmkuhl, R. Vinuesa, and I. Rodríguez, "Deep reinforcement learning for active flow control around a three-dimensional flow-separated wing at $re=1,000$," *arXiv preprint arXiv:2509.10195* (2025).
- ³⁵J. Vasantha, J. Rabault, F. Alcántara-Ávila, M. Mortensen, and R. Vinuesa, "Multi-agent reinforcement learning for the control of three-dimensional rayleigh-bénard convection," *Flow, Turbulence and Combustion* , 1–37 (2024).
- ³⁶C. Vignon, J. Rabault, and R. Vinuesa, "Recent advances in applying deep reinforcement learning for flow control: Perspectives and future directions," *Physics of fluids* **35**, 031301 (2023).
- ³⁷F. Moslem, M. Jebelli, M. Masdari, R. Askari, and A. Ebrahimi, "Deep reinforcement learning for active flow control in bluff bodies: A state-of-the-art review," *Ocean Engineering* **327**, 120989 (2025).
- ³⁸C. Zheng, T. Ji, F. Xie, X. Zhang, H. Zheng, and Y. Zheng, "From active learning to deep reinforcement learning: Intelligent active flow control in suppressing vortex-induced vibration," *Physics of Fluids* **33** (2021).
- ³⁹W. Chen, Q. Wang, L. Yan, G. Hu, and B. R. Noack, "Deep reinforcement learning-based active flow control of vortex-induced vibration of a square cylinder," *Physics of Fluids* **35** (2023).
- ⁴⁰F. Ren, C. Wang, J. Song, and H. Tang, "Deep reinforcement learning finds a new strategy for vortex-induced vibration control," *Journal of Fluid Mechanics* **990**, A7 (2024).
- ⁴¹F. Ren, Z. Ding, Y. Zhao, and D. Song, "Active control of wake-induced vibration using deep reinforcement learning," *Physics of Fluids* **36** (2024).
- ⁴²J. Zhao, J. S. Leontini, D. L. Jacono, and J. Sheridan, "Fluid-structure interaction of a square cylinder at different angles of attack," *Journal of Fluid Mechanics* **747**, 688–721 (2014).
- ⁴³A. Khalak and C. H. Williamson, "Investigation of relative effects of mass and damping in vortex-induced vibration of a circular cylinder," *Journal of Wind Engineering and Industrial Aerodynamics* **69**, 341–350 (1997).
- ⁴⁴M. Zhao, L. Cheng, H. An, and L. Lu, "Three-dimensional numerical simulation of vortex-induced vibration of an elastically mounted rigid circular cylinder in steady current," *Journal of Fluids and Structures* **50**, 292–311 (2014).
- ⁴⁵R. Skop and O. Griffin, "A model for the vortex-excited resonant response of bluff cylinders," *Journal of Sound and Vibration* **27**, 225–233 (1973).
- ⁴⁶O. Griffin, "Vortex-excited cross-flow vibrations of a single cylindrical tube," (1980).
- ⁴⁷O. M. Griffin and S. E. Ramberg, "Vortex shedding from a cylinder vibrating in line with an incident uniform flow," *Journal of Fluid Mechanics* **75**, 257–271 (1976).
- ⁴⁸M. P. Paidoussis, S. J. Price, and E. De Langre, *Fluid-structure interactions: cross-flow-induced instabilities* (Cambridge University Press, 2010).

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: [10.1063/5.0303905](https://doi.org/10.1063/5.0303905)

- ⁴⁹J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv preprint arXiv:1707.06347 (2017).
- ⁵⁰S. P. Singh, T. Jaakkola, and M. I. Jordan, "Learning without state-estimation in partially observable markovian decision processes," in *Machine Learning Proceedings 1994* (Elsevier, 1994) pp. 284–292.
- ⁵¹A. R. Cassandra, "A survey of pomdp applications," in *Working notes of AAAI 1998 fall symposium on planning with partially observable Markov decision processes*, Vol. 1724 (1998).
- ⁵²D. Bertsekas, *Dynamic programming and optimal control: Volume I*, Vol. 4 (Athena scientific, 2012).
- ⁵³G. Dulac-Arnold, N. Levine, D. J. Mankowitz, J. Li, C. Paduraru, S. Gowal, and T. Hester, "Challenges of real-world reinforcement learning: definitions, benchmarks and analysis," *Machine Learning* **110**, 2419–2468 (2021).
- ⁵⁴M. Weissenbacher, A. Borovykh, and G. Rigas, "Reinforcement learning of chaotic systems control in partially observable environments," *Flow, Turbulence and Combustion* , 1–22 (2025).
- ⁵⁵S. Choi, H. Choi, and S. Kang, "Characteristics of flow over a rotationally oscillating cylinder at low reynolds number," *Physics of Fluids* **14**, 2767–2777 (2002).