13th International Symposium on Turbulence and Shear Flow Phenomena (TSFP13)
Montreal, Canada, June 25–28, 2024

ACTIVE FLOW CONTROL OF THREE-DIMENSIONAL CYLINDERS USING DEEP REINFORCEMENT LEARNING

Pol Suárez, Francisco Alcántara-Ávila
FLOW, Engineering Mechanics
KTH Royal Institute of Technology
Stockholm, Sweden
polsm@kth.se, fraa@kth.se

Jean Rabault
Independent researcher
Oslo, Norway
jean.rblt@gmail.com

Arnau Miró
Barcelona Supercomputing Center
Barcelona, Spain
arnau.mirojane@bsc.es

Bernat Font
Faculty of Mechanical Engineering
Technische Universiteit Delft
Delft, The Netherlands
b.font@tudelft.nl

Oriol Lehmkuhl
Barcelona Supercomputing Center
Barcelona, Spain
oriol.lehmkuhl@bsc.es

Ricardo Vinuesa
FLOW, Engineering Mechanics
KTH Royal Institute of Technology
Stockholm, Sweden
rvinuesa@mech.kth.se

ABSTRACT

In this study we investigate the possibility of uncovering innovative strategies for reducing drag through deep reinforcement learning. We consider a three-dimensional cylinder, considering Reynolds numbers (Re_D) from 100 to 400. The transition to 3D wake instabilities appears in this regime. The active-flow-control (AFC) setup is based on multiple zero-net-mass-flux jets positioned on the top and bottom surfaces. A computational-fluid-dynamics solver is coupled with a multi-agent reinforcement-learning (MARL) framework based on the proximal-policy-optimization algorithm. The introduction of a MARL approach facilitates the exploitation of local invariance, adaptability of the control across different geometries, transfer learning, cross-application of agents and accelerating training. Our results demonstrate 21% and 16.5% drag reduction for $Re_D = 300$ and 400, respectively, outperforming classical periodic control, which yields up to 6% reduction for both. The current MARL-based framework marks the first instance of deep-reinforcement-learning training being carried out on 3D cylinders. This advancement opens doors for implementing AFC on increasingly more complex turbulent-flow setups.

Introduction

Flow-control systems, both passive and active, are crucial for developing sustainable solutions that can significantly cut emissions in the aviation industry (Choi *et al.*, 2008). Control devices use aerodynamics to reduce drag by manipulating pressure and viscosity. Examples include slats, flaps, winglets, and vortex generators, which improve aircraft performance and efficiency. Despite their promising potential, creating the best shapes or methods for these devices is challenging because it takes a lot of computing power to handle the complex interaction between pressure and viscosity in all flying conditions.

Alongside recent advancements in flow control, the integration of machine-learning (ML) techniques has brought significant promise to the aeronautics sector. This includes exploring fundamental issues in fluid mechanics (Vinuesa *et al.*, 2023) and developing entirely new approaches for active and passive flow control (AFC and PFC) (Le Clainche *et al.*, 2023). Deep reinforcement learning (DRL) stands out as a rapidly growing field within ML and garners substantial interest. Building on its success in board games, DRL proves effective in systems where a controller interacts with an environment to enhance a task, a characteristic relevant to most AFC scenarios. In such cases, DRL dynamically engages with the flow, receiving feedback and refining actions over time.

Designing AFC setups involves tackling complex, high-dimensional challenges that demand significant computational power to explore the vast parameter space of the control system and identify optimal values. DRL and neural networks are helpful tools that make this process easier, allowing us to develop effective control strategies without needing too much computational power.

The literature on DRL for AFC applications grows at a fast pace, exhibiting studies on flow control for two-dimensional (2D) cylinders ranging from $Re_D = 100$ and 8000 (where Re_D is the Reynolds number based on inflow velocity U_∞ and cylinder diameter D) with 17% and 33% drag reduction, respectively (Tang *et al.*, 2020; Li & Zhang, 2022; Ren *et al.*, 2021; Chatzimanolakis *et al.*, 2024), aircraft wings (Vinuesa *et al.*, 2022), fluid-structure interaction (Chen *et al.*, 2023), promising results in controlling highly turbulent flows such as (Font *et al.*, 2024, accepted), where a turbulent separation bubble reaching $Re_\tau = 750$ was successfully reduced, turbulent channels (Guastoni *et al.*, 2023) or Rayleigh–Bénard convection (Vignon *et al.*, 2023*a*). Some recent literature demonstrates the possibility of transfer learning from exploration done in 2D cylinders to 3D domains and higher Re_D (Wang *et al.*, 2023). The present work extends this state-of-the-art in 3D cylinders by using multiple actuators controlled via a novel multi-agent reinforcement learning (MARL) framework. The AFC is achieved through independent zero-net-mass-flow (ZNMF) jets aligned along slots on the top and bottom cylinder surfaces. This marks the first exploration of its kind directly within 3D cylinders.

Initially, at approximately $Re_D \approx 40$, there are symmetric counter-rotating vortices in the near wake. Beyond $Re_D \approx$

190, laminar vortex shedding starts, forming the well-known Kármán vortex street. Between $190 < Re_D < 260$, mode-A instability prevails with dominant spanwise wavelengths of $\lambda_z = 4D$ (Williamson, 1996; Barkley & Henderson, 1996). Past $Re_D \approx 260$, mode B becomes dominant, with finer three-dimensional features having shorter wavelengths of $\lambda_z = 1D$. Beyond these stages, the cylinder wake becomes more chaotic and turbulent. Developing flow-control strategies for the transition from 2D to 3D wake around a cylinder is challenging. The MARL setup needs to utilize spanwise characteristic structures as the wake becomes 3D to create effective control methods, with implications for drag reduction. DRL maximizes rewards (R) for an agent interacting with an environment via actions A and partial observations S. Episodes of consecutive actions are used to update neural network weights, optimizing policies for maximizing expected rewards. For recent advances in flow control with MARL, we refer to Brunton & Noack (2015); Vignon *et al.* (2023*b*).

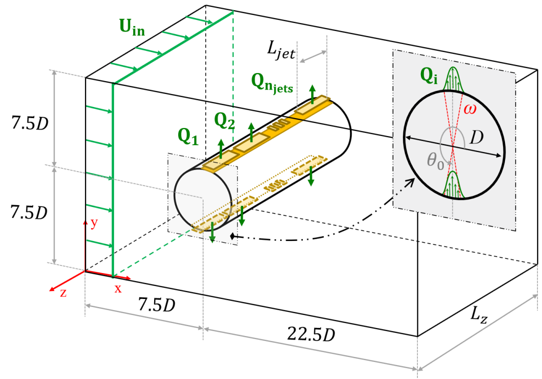

Figure 1. Schematic representation of the computational domain with cylinder diameter D as the reference length. Note that ω is the jet width and θ_0 is the angular location of each jet center. In green we show the uniform-velocity condition for the inlet and the sinusoidal profile in the jet azimuthal direction. Note that this representation is not to scale.

Methodology

Problem configuration and numerical setup

This study involves a 3D cylinder subjected to constant inflow in the streamwise direction, with all lengths non-dimensionalized using the cylinder diameter D as a reference. The computational domain, depicted in Figure 1, has dimensions $L_x = 30D$, $L_y = 15D$, and $L_z = 4D$, with the cylinder centered at $(x,y) = (7.5D, 7.5D)$. Periodic boundary conditions are applied in the cylinder spanwise direction.

At the inlet, we set a constant velocity U_∞ with a Dirichlet condition. The cylinder's surfaces follow no-slip, no-penetration conditions, while the top, bottom, and outflow surfaces of the domain act as outlets. The cylinder features two sets of $n_{\text{jet}} = 10$ aligned synthetic jets at the top and bottom (at $\theta_0^{\text{top}} = 90°$ and $\theta_0^{\text{bottom}} = 270°$, respectively) controlled externally to adjust mass-flow rate. A similar control strategy was discussed in Kim & Choi (2005), referred to as the *out-phase* approach. It involves implementing sinusoidal mass-flow distributions with different wavelengths along the spanwise direction. In this present work, the surface normal jet-velocity profile is defined in terms of the angle θ and the desired mass-flow rate Q per unit width:

$$\|U(Q,\theta)\| = Q\frac{\pi}{\rho D \omega}\cos\left(\frac{\pi}{\omega}(\theta-\theta_0)\right), \tag{1}$$

where $Q = \dot{m}/L_z$ and $|\theta - \theta_0| \in [-\omega/2, \omega/2]$, $\dot{m}$ is the mass flow rate. For each pseudo-environment, we set opposite action values within the pair of top and bottom jets, *i.e.* $Q_{90°} = -Q_{270°}$, in order to ensure the global zero net mass flux. An earlier setup was developed in Suárez *et al.* (2023).

The numerical simulations are carried out by means of the numerical solver Alya, which is described in detail in Vázquez *et al.* (2016). The spatial discretization is based on the finite-element method (FEM) and the incompressible Navier–Stokes equations are integrated numerically. For the time discretization, a semi-implicit method is used where the convective term follows a second-order Runge–Kutta scheme and a Crank–Nicholson scheme is used for the diffusive term (Crank & Nicolson, 1947). To select the appropriate time step, Alya uses an eigenvalue-based time-integration scheme (Trias & Lehmkuhl, 2011).

Multi-agent reinforcement learning (MARL)

We implemented a deep-reinforcement-learning (DRL) framework using Tensorforce libraries (Schaarschmidt *et al.*, 2017). DRL is very well suited for unsteady flow-control problems. It provides the possibility to dynamically interact with an environment, being able to dynamically set the actuation based on the varying flow state. We use the proximal-policy-optimization (PPO) algorithm (Schulman *et al.*, 2017), which is a policy-gradient approach based on a surrogate loss function for policy updates to prevent drastic drops in performance. This algorithm demonstrates robustness, as it is forgiving with hyperparameter initializations and can perform adequately across a diverse range of RL tasks without extensive tuning.

The neural-network architecture consists of two dense layers of 512 neurons each. The batch size, *i.e.* the total number of experiences that the PPO agent uses for each gradient-descent iteration, is set to 80, which is bigger than the values used in 2D trainings (Varela *et al.*, 2022). The limitation is that we have 10 actuators per environment and we need 10 streamed experiences which will be synchronized, so we have to work with a total of $10n_{\text{environments}}$ set of experiences. A streamed experience consists of a set of states, actions, rewards, and the predicted state that the agent expects to achieve. It is denoted as $(S,A,R,S')_{i,t}$ for each pseudo-environment, and each of the Reynolds numbers under consideration has its own agent and policy.

Previous work on 2D cylinders implemented the various training stages by means of a single-agent reinforcement learning (SARL) configuration. If the action space handles multiple jets at once, as is the case in the present 3D cylinder setup with distributed input forcing and distributed output reward (so-called DIDO scheme), SARL is not a viable option. As opposed to SARL, the MARL framework avoids the curse of dimensionality by exploiting invariances and aims to train local pseudo-environments. Doing so, the high-dimensional control becomes tractable and the agent is trained in smaller domains to maximize the local rewards. All the agents share the same neural-network weights, which is a key factor in significantly accelerating the training process. Note that each agent is coupled to a pair of jets that actuate independently from the others through the training process. The observation state S_i provided to the agent consists of partial pressure values along the domain. This information contains three slices with 99 pressure values each which are aligned with the corresponding jet. The probes or pressure values are concentrated in the wake and near-cylinder regions. This enables the agent to exploit the spanwise pressure gradients.

The total reward $R(t, i_{\text{jet}})$ defined in Equation (2) is expressed as a sum of the local, r^{local}, and global, r^{global}, rewards that correspond to each jet i_{jet}. The scalar K adjusts the values approximately within the range $[0,1]$ and β balances the local and global rewards; $\beta = 0.8$ is used in this work. The rewards r, defined in Equation (3), are functions of the aerodynamic force coefficients C_d and C_l (C_{d_b} is the averaged value for uncontrolled). The user-defined parameter α is a lift penalty and we considered $\alpha = 0.6$- The latter is essential to avoid undesired asymmetric strategies which favor a reduction of the component parallel to the incident velocity (drag) towards the perpendicular one (positive or negative lift). This is commonly referred to as the axis-switching phenomenon.

$$R(t, i_{\text{jet}}) = K \left[\beta r^{\text{local}}(t, i_{\text{jet}}) + (1-\beta) r^{\text{global}}(t) \right], \quad (2)$$

$$r(t, i_{\text{jet}}) = C_{d_b} - C_d(t, i_{\text{jet}}) - \alpha |C_l(t, i_{\text{jet}})|, \quad (3)$$

$$\text{where} \quad C_d = \frac{2F_x}{\rho A_f U_\infty^2} \quad \text{and} \quad C_l = \frac{2F_y}{\rho A_f U_\infty^2}. \quad (4)$$

The aerodynamic forces in Equation (4) involve the frontal area $A_f = DL_z$ from the local pseudo-environment surfaces for C_d^{local} and the whole cylinder for C_d^{global}.

The interactions between the agent and the physical environment are denoted as actions A, and they influence the system during T_a time units. We update the jet boundary conditions using Equation (1). The shift in time between actions, $Q_t \rightarrow Q_{t+1}$, is done by exponential functions. The smooth transition diminishes the appearance of sudden discontinuities which can spoil a training process. The DRL library requires rescaling as $Q = AQ_{\max}$ to avoid excessively large actuations. Hence, $Q_{\max} = 0.176$ was set based on our experience with DRL for flow control, and corresponds to twice the values used in the 2D cylinder setups (Varela *et al.*, 2022).

The episode duration is specifically defined to include at least six vortex-shedding periods ($T_k = 1/f_k$). We set $T_a = 0.05T_k$, based on the experience gathered with previous studies (Rabault *et al.*, 2019). Note that the vortex-shedding period is $T_k = 1/St \approx 1/0.2 = 5$, where $St = fD/U_\infty$ is the Strouhal number and f is the vortex-shedding frequency. This allows sufficient time between actions to produce an effect on the flow. If T_a is too short, there will be noise in the training process and it will become difficult to explore and correlate trajectories. On the other hand, if T_a is too large the agent will not be able to control the shorter characteristic time scales. Thus, a total of 120 actuations per episode is deemed sufficient for evaluating the cumulative reward. It is noteworthy that each episode starts from an uncontrolled converged state of the problem. This corresponds to what happens during training, but when we evaluate the DRL model in exploitation mode (also denoted as a deterministic mode). We also compare the DRL-based control with results from the classical periodic control (PC). The latter is chosen with the same jet flow rate as that of the DRL, and the frequency is chosen based on a parametric analysis of the frequency around the vortex-shedding frequency of the wake. We selected the frequency yielding the highest drag reduction.

Results

Training

An essential aspect of the training process is to leverage the physical understanding of the controlled phenomenon to evaluate anticipated reward values and physical control strategies thoughtfully. With these considerations in mind, Figure 2 shows the training curves for the four investigated cases in this study, at Reynolds numbers $Re_D = 100$, 200, 300, and 400. Commonly, sequences of actions A, states S, and rewards R are referred to as "environment episodes". However, in this case, it is more appropriate to call them "pseudo-environment episodes" due to the difference between SARL and MARL, where MARL involves multiple pseudo-environments within an environment. Hence, this Figure shows all the final rewards from the raw pseudo-environments, together with the pure drag reduction and lift-biased penalization contributions. As an example, the $Re_D = 300$ scenario closely resembles the ideal training condition. This is because the curves exhibit minimal lift bias and result in a total reward that matches the pure drag reduction. Similar patterns are obtained for the other Re_D cases indicating that the discovered policies are promising for all the cases. Note that we also observe several instances of apparent unlearning such as for $Re_D = 400$ at the episode 500 approximately. This is due to additional exploration of the agent, aimed at increasing the lift asymmetry (note the decrease of the red line), but quickly returning to exploiting what the agent has identified as a well-performing policy.

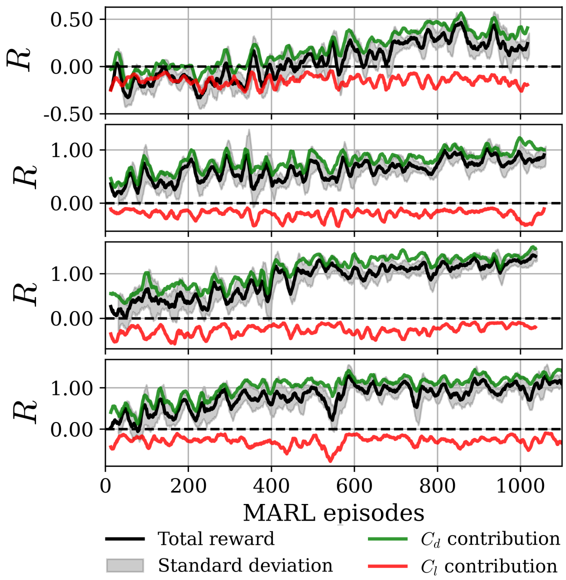

Figure 2. Evolution of the final reward R during the exploration phase as a function of the MARL episodes together with its lift-bias and pure drag-reduction contributions during training sessions. Signals are smoothed by a moving average of 15 values. From top to bottom, $Re_D = 100$, 200, 300 and 400.

In terms of computational cost, training is the most significant part. On average, each training session requires about 1200 MARL episodes, which is equivalent to running 120 numerical simulations for the entire domain. All exploration sessions were conducted on the Dardel high-performance computer in the PDC Center at KTH Royal Institute of Technology. The sessions run on 8 nodes simultaneously, each running one numerical simulation comprising 10 simultaneous pseudo-environments. Hence, 80 pseudo-environments in total. Each node has two AMD EPYC™ Zen2 2.25 GHz 64-core processors with 512 GB memory. With each batch of 8 simulations taking ideally five hours in this particular architecture, it requires less than four days of continuous operation.

Exploitation of the model

At this point, the agent policies are evaluated without any exploration. As a result, the agent calculates the most likely value of the action A within its learned probability distribution, aiming to maximize the expected reward. The DRL-based control exhibits a clear two-phase process that starts with a short transient period followed by the stationary control policy. It takes less than $4T_k$ for the DRL-based control to reach the stationary behavior. It exhibits a first suction/ejection overshoot which destabilizes the wake, and then it proceeds to restabilize it in a second phase. During the latter, the jet mass flux exhibits lower values which barely reach 75% of those in the transient overshoots. This control strategy persists until control stabilizes into stationary behavior, which is monitored by assessing mean quantities and fluctuations in aerodynamic forces. The averaged drag-reduction results for all Re_D are reported in Figure 3. It is important to note that all the cases lead to effective drag-reduction rates. The overall performance is much better than what can be obtained with the classical PC strategies. The values are averaged in time by considering an interval of at least $20T_k$, *i.e.* over 100 time units, excluding the transients obtained after applying the control. The root-mean-square of the fluctuations, $\mathrm{RMS} = \sqrt{(1/n)\sum_{i=1}^{n}(x_i - \bar{x})^2}$, minimum and maximum values provide deeper insights into the mentioned robustness. While the mean values alone may suggest a good performance of the PC, the merits of the control should not be assessed solely based on this quantity. When considering an optimal control strategy, the preferred choice typically involves selecting a control with minimal variability and few extreme values, which are characteristics exhibited as evidenced by the DRL-based control. We also study the ratio between the total fluid mass intercepted by the frontal area of the cylinder E_∞ and the total mass used by the actuators E_c. Based on the definitions used in Chatzimanolakis *et al.* (2024), we propose the following expression for the ratio E_c^*:

$$E_c^* = \frac{E_c}{E_\infty} = \frac{L_{\mathrm{jet}}}{(t_2 - t_1)Q_\infty L_z} \int_{t_1}^{t_2} \sum_{i=1}^{n_{\mathrm{jets}}} |Q_i(t)| \, \mathrm{d}t, \tag{5}$$

where t_1 and t_2 define the start and end of our time interval for evaluating the control and $Q_\infty = \rho D U_\infty$ define the mass-flow intercepted by the cylinder. Note that a complete evaluation of the mass used by the actuators would depend on the actual jets used in the experimental setup, and in this work we adopt a purely numerical approach based on modifying the Dirichlet boundary condition at the cylinder surface. Keeping this aspect in mind, the present numerical work shows that the mass cost is minimal compared to the gains achieved through drag reduction.

	Uncontrolled	Periodic control				DRL control			
Re_D	St	$Q_{\max}$	Q_{RMS}	f_c	St	$Q_{\max}$	Q_{RMS}	f_c	St
100	0.170	0.053	0.037	0.115	0.113	0.016	8.4×10^{-3}	0.153	0.139
200	0.186	0.053	0.037	0.130	0.117	0.031	9.5×10^{-3}	0.173	0.157
300	0.206	0.018	0.013	0.175	0.177	0.031	7.6×10^{-3}	0.200	0.189
400	0.202	0.012	0.008	0.172	0.194	0.025	5.7×10^{-3}	0.200	0.169

Table 1. Main characteristics of control strategies, including the Strouhal number St, maximum and fluctuations RMS of mass-flow rates per unit width ($Q_{\max}$ and Q_{RMS}) and control frequency f_c.

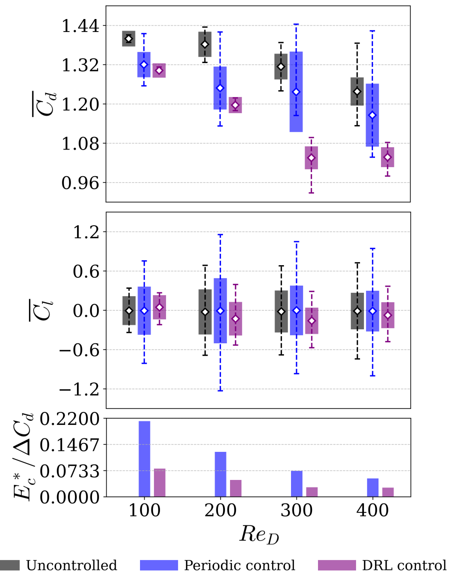

Figure 3. Summary of (top) mean drag (C_d) and (middle) lift (C_l) coefficients shown as white dots, RMS of fluctuations are represented as thick bars, and maximum-minimum shown as dashed intervals for each case. (Bottom) Non-dimensionalized cost metric per drag reduction rate $E_c^*/\Delta C_d$ (lower is better), as defined in Equation (5).

Additional physical insight has been studied by assessing the power-spectral density (PSD) of the streamwise velocity. The change of frequency impacts the wake topology after applying the various control strategies. In particular, both the DRL-based control and PC cases exhibit a reduction in St. Further insight into the various control strategies is provided in Table 1, where several characteristic variables of the various controls are shown. The first important observation is the fact that the root-mean-square (RMS) of the jet mass-flow rate (computed by averaging in time and the spanwise direction) is one order of magnitude lower in the DRL than in the PC. This indicates that the DRL-based control strategies lead to more stable and robust configurations, avoiding large peak-to-peak variations in the actuation. Although the f_c values are not dramatically different in the PC and DRL cases (in this case f_c is just the dominant frequency), the latter exhibit significantly more complex control laws than the former.

Figure 4 demonstrates the biggest advantage of a MARL implementation: the control policy can act locally, exploiting wake vortical structures and distributing the flow of the jets in the spanwise direction to maximize the global objective of minimizing the overall drag. The agent utilizes less than $10\% Q_{\max}$. While cases up to $Re_D = 200$ exhibit a uniform distribution in the spanwise direction, beyond this Reynolds number, the control exhibits spanwise variations. As mentioned in the Introduction, for $Re_D \geq 250$ the wake displays three-dimensional features, effectively utilized by the DRL control to maximize the achieved drag-reduction rates. During the exploration stage, for $Re_D = 100$ and 200, the agent could not find any strategy with spanwise variations leading to better performance, indicating that the wake is two-dimensional in these cases, favoring spanwise-uniform control strategies. On the other hand, at $Re_D = 300$ and 400, we see patterns in the flow related to the transitional Re_D, including spanwise structures of approximately one cylinder diameter ($\lambda_z = 1D$) associated with mode-B instabilities. In Figure 5 we illustrate how the flow topology is influenced by the various drag-reduction

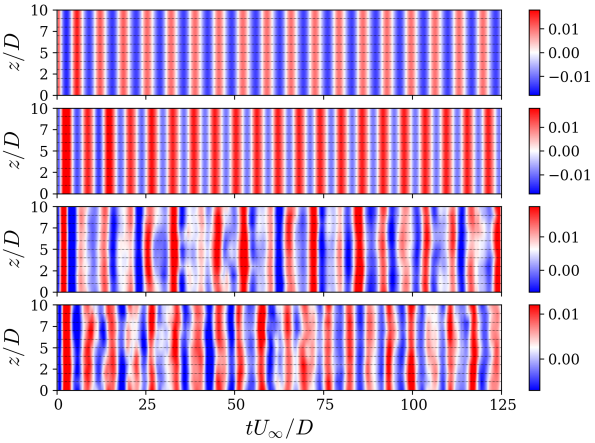

Figure 4. Evolution in time of the mass-flow rate per unit width Q for all jets individually, showing also their spanwise distribution for the DRL cases. From top to bottom: $Re_D = 100, 200, 300$ and 400.

strategies, on three representative phases: uncontrolled, transient, and stabilized control. The flow visualizations indicate that the control strategies based on DRL aim to enhance the spacing between successive vortical structures, resulting in a reduction of the vortex-shedding period T_k. Hence, mode-B instabilities are diminished when the control is applied, and the intensity of the vortex shedding is attenuated. These changes lead to a more organized wake structure, resembling the characteristic two-dimensional laminar wake. Figure 4 corroborates these findings, illustrating diminished oscillations during the controlled phase. Studying flow statistics offers deeper insight into the mechanisms employed by the DRL agent to discover flow-control strategies, particularly when analyzing the mean flow and the Reynolds stresses. To compute the latter, the Reynolds decomposition is used to decompose the flow variables (u) into time-averaged mean ($\overline{u}$) and fluctuating (u') components, $u = \overline{u} + u'$. In Figures 6, we observe the impact of the DRL-based control: the wake nearly doubles the recirculation-bubble length, delaying the wake-stagnation point by approximately one diameter in the streamwise direction. We only present $Re_D = 400$ as a representative case. The back pressure increases by $\Delta\overline{C_p^b} \approx 0.4$, which is directly related to the drag reduction mechanism. The Reynolds stresses are presented in Figure 6, which shows that the peaks move downwards in the streamwise direction after applying the control, with only small changes in the vertical location. DRL-based control generally leads to the reduction of the peak magnitude in almost all the fluctuating quantities. All Re_D cases elucidate that the same behavior occurs within this regime range.

Discussion

In this study a multi-agent reinforcement-learning (MARL) framework is coupled with a numerical solver to discover effective drag-reduction strategies by controlling multiple jets placed along the span of three-dimensional cylinders. We study cases at $Re_D = 100, 200, 300$, and 400, where wake transition from 2D to 3D is observed. All DRL-based control policies outperform the classical periodic control in this Re_D range. Such a range is characterized by the emergence of spanwise instabilities, which the DRL agent can exploit to discover effective drag-reduction strategies. This is achieved by taking advantage of exploiting the underlying physics within pseudo-environments and optimizing the global problem involving multiple interactions in parallel. One of the main advantages of employing MARL is the possibility to deploy trained agents across various cylinder lengths and numbers of actuators while ensuring consistency in the spanwise width of the jets (L_{jet}) and their corresponding pressure values as observation states (S). Note that the training focuses on symmetries and invariant structures. This would not be possible with SARL, which is restricted to a certain number of actuators (and also the corresponding algorithm limitations). MARL allows cheaper training sessions in smaller and simplified computational domains, thereby speeding up the process, which is required to perform flow control in high-fidelity simulations.

These findings highlight the effectiveness of the DRL approach, which can discover flow-control strategies more sophisticated than those obtained with the classical periodic control, spanning wide ranges of frequencies and tackling different flow features in the wake. DRL-based control achieves a remarkable performance, reducing drag by 21% and 16.5% for $Re = 300$ and 400 respectively, outperforming PC strategies which only achieve around 6% reduction for both Re. Furthermore, the results presented here represent the first training conducted in 3D cylinders. This sets a new benchmark for the DRL community, which may motivate its use in future applications for DIDO schemes.

REFERENCES

Barkley, Dwight & Henderson, Ronald D. 1996 Three-dimensional floquet stability analysis of the wake of a circular cylinder. *Journal of Fluid Mechanics* **322**, 215–241.

Brunton, S. L. & Noack, B. R. 2015 Closed-Loop Turbulence Control: Progress and Challenges. *Applied Mechanics Reviews* **67** (5).

Chatzimanolakis, Michail, Weber, Pascal & Koumoutsakos, Petros 2024 Learning in two dimensions and controlling in three: Generalizable drag reduction strategies for flows past circular cylinders through deep reinforcement learning. *Phys. Rev. Fluids* **9**, 043902.

Chen, Wenjie, Wang, Qiulei, Yan, Lei, Hu, Gang & Noack, Bernd R 2023 Deep reinforcement learning-based active flow control of vortex-induced vibration of a square cylinder. *Physics of Fluids* **35** (5).

Choi, Haecheon, Jeon, Woo-Pyung & Kim, Jinsung 2008 Control of flow over a bluff body. *Annual Review of Fluid Mechanics* **40** (1), 113–139.

Crank, John & Nicolson, Phyllis 1947 A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type **43**, 50–67.

Font, B., Alcántara-Ávila, F., Rabault, J., Vinuesa, R. & Lehmkuhl, O. 2024, accepted Active flow control of a turbulent separation bubble through deep reinforcement learning. *Journal of Physics: Conference Series* .

Guastoni, L., Rabault, J., Schlatter, P., Azizpour, H. & Vinuesa, R. 2023 Deep reinforcement learning for turbulent drag reduction in channel flows. *European Physical Journal E* **46** (4), 27.

Jeong, Jinhee & Hussain, Fazle 1995 On the identification of a vortex. *Journal of Fluid Mechanics* **285**, 69–94.

Kim, Jinsung & Choi, Haecheon 2005 Distributed forcing of flow over a circular cylinder. *Physics of Fluids* **17** (3), 033103.

Le Clainche, Soledad, Ferrer, Esteban, Gibson, Sam, Cross, Elisabeth, Parente, Alessandro & Vinuesa, Ricardo 2023 Improving aircraft performance using machine learning: A

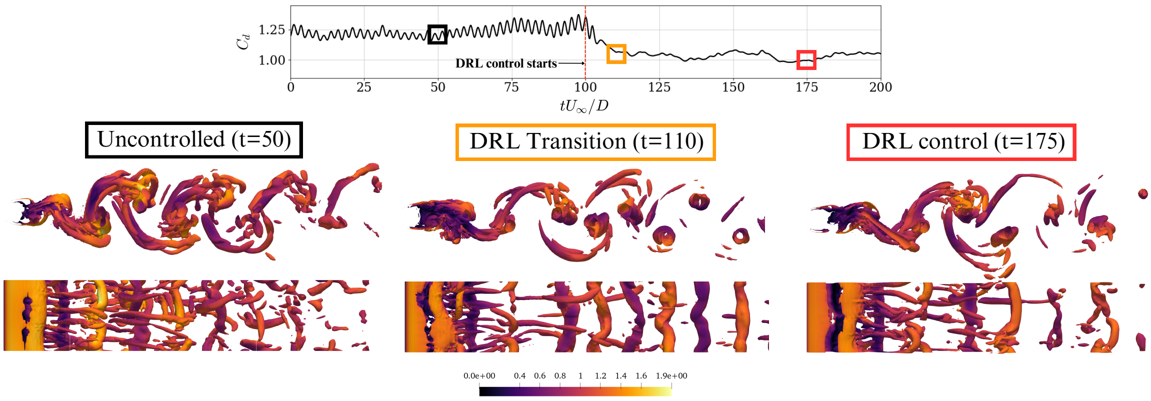

Figure 5. Temporal evolution of flow coherent structures at $Re_D = 400$ during DRL-based strategy exploitation, from uncontrolled state to a stable DRL control. (Top) Drag coefficient C_d as a function of time and (bottom) snapshots showing vortical structures identified with the λ_2 criterion defined by Jeong & Hussain (1995), where the isosurface $\lambda_2 D^2/U_\infty^2 = -0.5$ is shown.

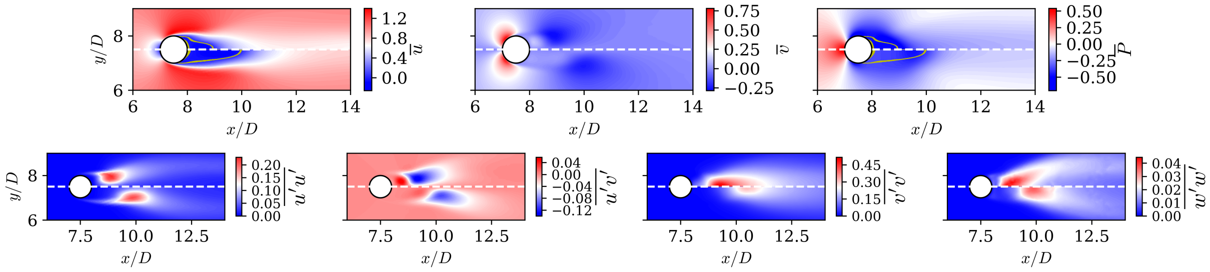

Figure 6. (Top) Mean velocities and pressure fields for $Re_D = 400$, where the upper-half domain is uncontrolled and the bottom half is DRL-controlled flow. Regions where $\overline{u} = 0$ are denoted by yellow color and indicate the wake-stagnation. (Bottom) Reynolds stresses $\overline{u'u'}, \overline{u'v'}, \overline{v'v'}$ and $\overline{w'w'}$, where the upper-half domain is uncontrolled and the bottom half is DRL-controlled flow.

review. *Aerospace Science and Technology* **138**.

Li, Jichao & Zhang, Mengqi 2022 Reinforcement-learning-based control of confined cylinder wakes with stability analyses. *Journal of Fluid Mechanics* **932**, A44.

Rabault, J., Kuchta, M., Jensen, A., Reglade, U. & Cerardi, N. 2019 Artificial Neural Networks trained through Deep Reinforcement Learning discover control strategies for active flow control. *Journal of Fluid Mechanics* **865**, 281–302.

Ren, Feng, Rabault, Jean & Tang, Hui 2021 Applying deep reinforcement learning to active flow control in weakly turbulent conditions. *Physics of Fluids* **33** (3).

Schaarschmidt, M., Kuhnle, A. & Fricke, K. 2017 TensorForce: A TensorFlow library for applied reinforcement learning Published: Web page.

Schulman, John, Wolski, Filip, Dhariwal, Prafulla, Radford, Alec & Klimov, Oleg 2017 Proximal policy optimization algorithms. *Preprint arXiv: 1707.06347* .

Suárez, Pol, Alcántara-Ávila, Francisco, Miró, Arnau, Rabault, Jean, Font, Bernat, Lehmkuhl, Oriol & Vinuesa, R 2023 Active flow control for three-dimensional cylinders through deep reinforcement learning. *Preprint arXiv: 2309.02462* .

Tang, H., Rabault, J., Kuhnle, A., Wang, Y. & Wang, T. 2020 Robust active flow control over a range of Reynolds numbers using an artificial neural network trained through deep reinforcement learning. *Physics of Fluids* **32** (5).

Trias, FX & Lehmkuhl, O 2011 A self-adaptive strategy for the time integration of navier-stokes equations. *Numerical Heat Transfer, Part B: Fundamentals* **60** (2), 116–134.

Varela, P., Suárez, P., Alcántara-Ávila, F., Miró, A., Rabault, J., Font, B., García-Cuevas, L. M., Lehmkuhl, O. & Vinuesa, R. 2022 Deep reinforcement learning for flow control exploits different physics for increasing Reynolds number regimes. *Actuators, MDPI* **11** (12), 359.

Vignon, C., Rabault, J., Vasanth, J., Alcántara-Ávila, F., Mortensen, M. & Vinuesa, R. 2023*a* Effective control of two-dimensional Rayleigh–Bénard convection: invariant multi-agent reinforcement learning is all you need. *Physics of Fluids* **35** (6).

Vignon, C., Rabault, J. & Vinuesa, R. 2023*b* Recent advances in applying deep reinforcement learning for flow control: Perspectives and future directions. *Physics of Fluids* **35** (3).

Vinuesa, Ricardo, Brunton, Steven L & McKeon, Beverley J 2023 The transformative potential of machine learning for experiments in fluid mechanics. *Nature Reviews Physics* **5** (9), 536–545.

Vinuesa, Ricardo, Lehmkuhl, Oriol, Lozano-Durán, Adrian & Rabault, Jean 2022 Flow control in wings and discovery of novel approaches via deep reinforcement learning. *Fluids* **7** (2), 62.

Vázquez, M., Houzeaux, G., Koric, S., Artigues, A., Aguado-Sierra, J., Arís, R., Mira, D., Calmet, H., Cucchietti, F., Owen, H., Taha, A., Burness, E. D., Cela, J. M. & Valero, M. 2016 Alya: Multiphysics engineering simulation toward exascale. *Journal of Computational Science* **14**, 15–27.

Wang, Zhicheng, Fan, Dixia, Jiang, Xiaomo, Triantafyllou, Michael S. & Karniadakis, George Em 2023 Deep reinforcement transfer learning of active control for bluff body flows at high reynolds number. *Journal of Fluid Mechanics* **973**, A32.

Williamson, C.H.K 1996 Vortex Dynamics in the Cylinder Wake. *Annual Review of Fluid Mechanics* **28**, 477–539.