J. Fluid Mech. (2023), *vol*. 956, R4, doi:10.1017/jfm.2023.28

Bio-inspired variable-stiffness flaps for hybrid flow control, tuned via reinforcement learning

Nirmal J. Nair[1],† **and Andres Goza**[1]

[1]Department of Aerospace Engineering, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA

(Received 18 October 2022; revised 15 December 2022; accepted 3 January 2023)

A bio-inspired, passively deployable flap attached to an airfoil by a torsional spring of fixed stiffness can provide significant lift improvements at post-stall angles of attack. In this work, we describe a hybrid active–passive variant to this purely passive flow control paradigm, where the stiffness of the hinge is actively varied in time to yield passive fluid–structure interaction of greater aerodynamic benefit than the fixed-stiffness case. This hybrid active–passive flow control strategy could potentially be implemented using variable-stiffness actuators with less expense compared with actively prescribing the flap motion. The hinge stiffness is varied via a reinforcement-learning-trained closed-loop feedback controller. A physics-based penalty and a long–short-term training strategy for enabling fast training of the hybrid controller are introduced. The hybrid controller is shown to provide lift improvements as high as 136 % and 85 % with respect to the flapless airfoil and the best fixed-stiffness case, respectively. These lift improvements are achieved due to large-amplitude flap oscillations as the stiffness varies over four orders of magnitude, whose interplay with the flow is analysed in detail.

Key words: flow–structure interactions, machine learning

1. Introduction

For aerodynamic flows, a passive flow control device (flap) inspired by self-actuating covert feathers of birds has been shown to improve lift at post-stall angles of attack (Bramesfeld & Maughmer 2002; Duan & Wissa 2021). In particular, when the flap is mounted via a torsional spring, further aerodynamic benefits can be obtained compared with a free (zero hinge stiffness) or static configuration (Rosti, Omidyeganeh & Pinelli 2018). These added benefits arise from rich fluid–structure interaction (FSI) between the flap and vortex dynamics (Nair & Goza 2022*a*). This outcome teases a question: Could additional lift enhancement be achieved if the flap motion was controlled to yield more favourable flapping amplitudes and phase relative to key flow processes?

† Email address for correspondence: njn2@illinois.edu

To address this question, we propose a hybrid active–passive flow control method to adaptively tune the flap stiffness. That is, the flap dynamics are passively induced by the FSI, according to the actively modulated hinge stiffness. This hybrid approach could incur less expense as compared to a fully active control method where the flap deflection is controlled using a rotary actuator. Our focus is on the design of a control algorithm that can actuate the hinge stiffness to provide aerodynamic benefits without accounting for how these stiffness changes are implemented, and on explaining the physical mechanisms that drive these benefits. We note, however, that there are various ways of achieving stiffness modulation in practice via continuous variable-stiffness actuators (VSAs) (Wolf *et al.* 2015), used extensively in robotics (Ham *et al.* 2009), wing morphing (Sun *et al.* 2016), etc. A discrete VSA restricts the stiffness to vary discretely across fixed stiffness levels, but it weighs less and requires lower power (Diller, Majidi & Collins 2016).

Historically, linear approximations of fundamentally nonlinear systems are used to design optimal controllers (Kim & Bewley 2007). While these linear techniques have been effective in stabilizing separated flows at low Reynolds number, $Re \sim O(10^2)$, where the base state has a large basin of attraction, its effectiveness is compromised at larger Re (Ahuja & Rowley 2010). These challenges are exacerbated by the nonlinear FSI coupling between the flap and vortex shedding of interest here. Model predictive control (MPC) uses nonlinear models to make real-time predictions of the future states to guide the control actuations. The need for fast real-time predictions necessitates the use of reduced-order models where the control optimization problem is solved using a reduced system of equations (Peitz, Otto & Rowley 2020). Machine learning in fluid mechanics has provided further avenues of deriving more robust reduced nonlinear models to be used with MPC (Baumeister, Brunton & Kutz 2018; Mohan & Gaitonde 2018; Bieker *et al.* 2020). However, these reduced-order modelling efforts remain an area of open investigation, and would be challenging for the strongly coupled flow–airfoil–flap FSI system. We therefore utilize a model-free, reinforcement learning (RL) framework to develop our controller. RL has recently gained attention in fluid mechanics (Garnier *et al.* 2021), and is used to learn an effective control strategy by trial and error via stochastic agent–environment interactions (Sutton & Barto 2018). Unlike MPC, the control optimization problem is completely solved offline, thereby not requiring real-time predictions. RL has been successfully applied to attain drag reduction (Rabault *et al.* 2019; Fan *et al.* 2020; Paris, Beneddine & Dandois 2021; Li & Zhang 2022), for shape optimization (Viquerat *et al.* 2021) and understanding swimming patterns (Verma, Novati & Koumoutsakos 2018; Zhu *et al.* 2021).

In this work, we develop a closed-loop feedback controller using deep RL for our proposed hybrid control approach consisting of a tunable-stiffness covert-inspired flap. We train and test this controller using high-fidelity fully coupled simulations of the airfoil–flap–flow dynamics, and demonstrate the effectiveness of the variable-stiffness control paradigm compared with the highest-performing passive (single-stiffness) case. We explain the lift-enhancement mechanisms by relating the large-amplitude flap dynamics to those of the vortex formation and shedding processes around the airfoil.

2. Methodology

2.1. *Hybrid active–passive control*

The problem set-up is shown in figure 1, which consists of a NACA0012 airfoil of chord c at an angle of attack of 20° and $Re = 1000$, where significant flow separation and vortex shedding occur. A flap of length $0.2c$ is hinged on the upper surface of the airfoil via

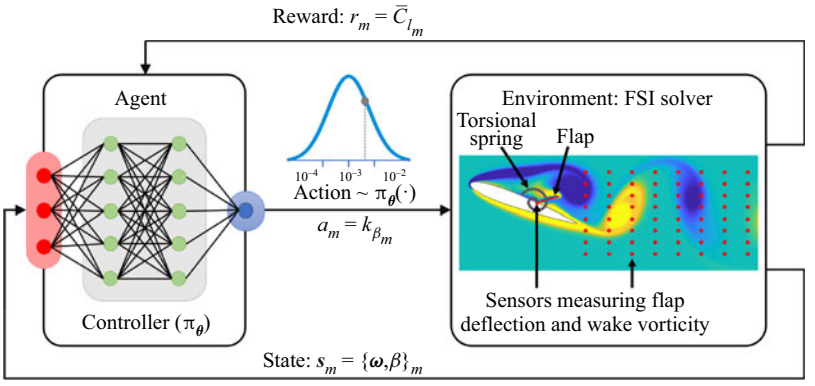

Figure 1. Schematic of the problem set-up and RL framework.

a torsional spring with stiffness k_β, where β denotes the deflection of the flap from the airfoil surface. The FSI problem considered here is two-dimensional (2-D), similar to most studies on covert-inspired flaps (Bramesfeld & Maughmer 2002; Duan & Wissa 2021) and applicable to high-aspect-ratio, bio-inspired flows largely dominated by 2-D effects (Taira & Colonius 2009; Zhang *et al.* 2020). In the passive control approach (Nair & Goza 2022*a*), k_β was fixed and maximum lift was attained at $k_\beta = 0.015$. In our hybrid active–passive control, the stiffness is a function of time, $k_\beta(t)$, determined by an RL-trained closed-loop feedback controller described in § 2.2. While the stiffness variation is allowed to take any functional form, it is restricted to vary in $k_\beta(t) \in [10^{-4}, 10^{-1}]$, similar to the range of stiffness values considered in the passive control study. The mass and location of the flap are fixed at $m_\beta = 0.01875$ and 60 % of the chord length from the leading edge, chosen here since they induced the maximal lift benefits in the passive (single-stiffness) configuration (cf. figure 5(*e–h*) for vorticity contours at four time instants in one periodic lift cycle for this highest-lift single-stiffness case).

2.2. *Reinforcement learning (RL)*

A schematic of the RL framework is shown in figure 1, where an agent (with a controller) interacts with an environment. At each time step, m, the agent observes the current state of the environment, $\boldsymbol{s}_m \in \mathbb{R}^{N_s}$ (where N_s is the number of states), implements an action, $a_m \in \mathbb{R}$, and receives a reward, $r_m \in \mathbb{R}$. The environment then advances to a new state, $\boldsymbol{s}_{m+1}$. This process is continued for M_τ time steps and the resulting sequence forms a trajectory, $\tau = \{\boldsymbol{s}_0, a_0, r_0, \ldots, \boldsymbol{s}_{M_\tau}, a_{M_\tau}, r_{M_\tau}\}$. The actions chosen by the agent to generate this trajectory are determined by the stochastic policy of the controller, $\pi_{\boldsymbol{\theta}}(a_m|\boldsymbol{s}_m)$, parametrized by weights $\boldsymbol{\theta}$. This policy outputs a probability distribution of actions, from which a_m is sampled, $a_m \sim \pi_{\boldsymbol{\theta}}(a_m|\boldsymbol{s}_m)$, as shown in figure 1.

In policy-based deep RL methods, a neural network is used as a nonlinear function approximator for the policy, as shown in figure 1. Accordingly, $\boldsymbol{\theta}$ corresponds to the weights of the neural network. The goal in RL is to learn an optimal control policy that maximizes an objective function $J(\cdot)$, defined as the expected sum of rewards as

$$J(\boldsymbol{\theta}) = \mathbb{E}_{\tau\sim\pi_{\boldsymbol{\theta}}}\left[\sum_{m=0}^{M_\tau} r_m\right]. \tag{2.1}$$

Here, the expectation is performed over N_τ different trajectories sampled using the policy, $\tau \sim \pi_{\boldsymbol{\theta}}$. The maximization problem is then solved by gradient ascent, where an approximate gradient of the objective function, $\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta})$, is obtained from the policy gradient theorem (Nota & Thomas 2019).

In this work, we use the PPO algorithm (Schulman *et al.* 2017) for computing the gradient, which is a policy-based RL method suitable for continuous control problems, as opposed to Q-learning methods for discrete problems (Sutton & Barto 2018). PPO has been used successfully to develop RL controllers for fluid flows (Rabault *et al.* 2019), and is chosen here among other policy-based methods due to its relative simplicity in implementation, better sample efficiency and ease of hyperparameter tuning.

In our hybrid control problem, the environment is the strongly coupled FSI solver of Nair & Goza (2022*b*), where the incompressible Navier–Stokes equation for the fluid coupled with Newton's equation of motion for the flap are solved numerically. The state provided as input to the controller are sensor measurements consisting of flow vorticity in the wake, $\boldsymbol{\omega}_m$, and flap deflection, β_m. The action is the time-varying stiffness, $a_m = k_{\beta_m} \in \mathbb{R}^+$. Similar to Rabault *et al.* (2019), when advancing a single control time step from m to $m+1$, the flow–airfoil–flap system is simulated for N_t numerical time steps of the FSI solver. In this duration, the chosen value of the stiffness is kept constant. The reason for introducing these two time scales – control and numerical – is to allow the FSI system to meaningfully respond to the applied stiffness and achieve faster learning.

The reward for the lift maximization problem of our hybrid control approach is

$$r_m = \frac{1}{2}\bar{C}_{l_m}^2 + p_1\left(\frac{1-p_2{}^{u/u_{max}}}{p_2}\right). \tag{2.2}$$

The first term is the mean lift coefficient of the airfoil, $\bar{C}_{l_m}$, evaluated over the N_t numerical time steps. The second term, where $p_1 > 0$ and $p_2 \gg 1$ are constants whose values are given in § 3, is a physics-based penalty term that provides an exponentially growing negative contribution to the reward if the flap remains undeployed for several consecutive control time steps (intuitively, one wishes to avoid periods of prolonged zero deployment angle). Accordingly, u denotes the current count of consecutive control time steps that the flap has remained undeployed and u_{max} is the maximum number of consecutive time steps that the flap may remain undeployed. The flap is deemed undeployed if $\beta_m < \beta_{min}$.

The reward (2.2) can be augmented with additional penalty terms to achieve goals of fluctuation minimization, drag reduction or improving aerodynamic efficiency. For minimizing lift fluctuations about the mean, the penalty term could take the form of the time derivative of lift, $-\mathrm{d}\bar{C}_{l_m}/\mathrm{d}t^2$, the difference between the instantaneous lift and target lift, $-(\bar{C}_{l_m} - C_{l_{target}})^2$, or the running average of lift evaluated at previous time steps, $-(\bar{C}_{l_m} - \bar{C}_{l_{run}})^2$. For drag minimization or improving aerodynamic efficiency, the airfoil drag, $-\bar{C}_{d_m}^2$, could be incorporated into the reward. In this work, we do not consider these additional penalty terms because the hybrid control approach is inspired from covert feathers in birds, which are used as high-lift devices in stalled large-angle-of-attack scenarios. The primary aerodynamic aim in this setting is to attain increased lift, at the potential cost of drag, efficiency or fluctuations (Rosti *et al.* 2018; Nair & Goza 2022*a*).

The RL algorithm is proceeded iteratively as shown in Algorithm 1. Each iteration consists of sampling trajectories spanning a total of M_i control time steps (lines 4 and 5) and using the collected data to optimize $\boldsymbol{\theta}$ (line 19). We also define an episode as either the full set of M_i time steps or, in the case where the parameters yield an undeployed flap, the time steps until $u = u_{max}$. Note that an episode and iteration coincide only if an episode

Algorithm 1 The PPO RL algorithm applied to hybrid control

Require: Set of initial conditions, $\mathcal{S}^0$
Ensure: Optimization parameters, $\boldsymbol{\theta}$
1: Initialize state, $\boldsymbol{s}_0 \sim \mathcal{S}^0$
2: Initiate a vector of counters for each trajectory, $u[1:N_\tau] \leftarrow 0$, $count[1:N_\tau] \leftarrow 0$
3: **for** iterations $\leftarrow 1, 2, \ldots$ **do**
4: **for** window $\leftarrow 1, 2, \ldots, k$ **do** {multi-window optimization}
5: **for** $m \leftarrow 1, 2, \ldots, M_\tau (= M_i/k)$ **do** {perform trajectory sampling}
6: **for** trajectory:$j \leftarrow 1, 2, \ldots, N_\tau$ **do** {parallel trajectory sampling}
7: $a_m[j] \sim \pi_{\boldsymbol{\theta}}(a_m[j]|\boldsymbol{s}_m[j])$
8: $r_m[j], \boldsymbol{s}_{m+1}[j] = \mathrm{FSI}(\boldsymbol{s}_m[j], a_m[j])$
9: $count[j] \leftarrow count[j] + 1$
10: **if** $\beta_{m+1}[j] < \beta_{min}$ **then**
11: $u[j] \leftarrow u[j] + 1$
12: **end if**
13: **if** $u[j] = u_{max}$ || $count[j] = M_i$ **then** {episode termination}
14: Reset to initial condition, $\boldsymbol{s}_{m+1}[j] \sim \mathcal{S}^0$
15: $u[j] \leftarrow 0$, $count[j] \leftarrow 0$
16: **end if**
17: **end for**
18: **end for**
19: Optimize $\boldsymbol{\theta}$ using gradient ascent
20: **end for**
21: **end for**

is not terminated early. The state is only reset after an episode terminates (line 14), which could occur within an iteration if the episode terminates early (line 13).

We also use a modified strategy to update $\boldsymbol{\theta}$ in a given iteration. In policy-based methods, the weights update step is performed after trajectory sampling. Generally, in one iteration, trajectories are collected only once. This implies that (a) typically the weights are updated once in one training iteration and (b) the length of the trajectory is equal to the iteration length, $M_\tau = M_i$. However, in our work, we perform $k > 1$ number of weight updates in a single iteration by sampling k trajectories each of length $M_\tau = M_i/k$. This procedure is found to exhibit faster learning since the frequent weight updates sequentially cater to optimizing shorter temporal windows of the long-time horizon. We therefore refer to this procedure as the long–short-term training strategy and demonstrate its effectiveness in § 3. As shown in Algorithm 1, each iteration is divided into k optimization windows (line 4) and $\boldsymbol{\theta}$ is updated at the end of each window (line 19). Finally, for computing more accurate estimates of the expected values used in (2.1) and for evaluating gradients, the mth time advancement (line 5) is performed N_τ times independently (line 6) (Schulman *et al.* 2017). For accelerated training, this set of N_τ trajectories is sampled in parallel (Rabault & Kuhnle 2019; Pawar & Maulik 2021).

3. Results

3.1. *RL and FSI parameters*

The parameters of the FSI environment are the same as in Nair & Goza (2022*b*), which contains the numerical convergence details. The spatial grid and time-step sizes are $\Delta x/c = 0.00349$ and $\Delta t/(c/U_\infty) = 0.0004375$, respectively. For the multi-domain approach for far-field boundary conditions, five grids of increasing coarseness are used, where the finest and coarsest grids are $[-0.5, 2.5]c \times [-1.5, 1.5]c$ and $[-23, 25]c \times [-24, 24]c$, respectively. The airfoil leading edge is located at the origin. For the subdomain approach, a rectangular subdomain that bounds the physical limits of flap

displacements, $[0.23, 0.7]c \times [-0.24, 0.1]c$, is utilized. The FSI solver is parallelized and simulated across six processors.

For the states, $N_s = 65$ sensor measurements are used, which measure vorticity at 64 locations distributed evenly across $[1, 2.4]c \times [-0.6, 0.1]c$ and flap deflection as denoted by the red markers in figure 1. To ensure unbiased stiffness sampling across $[10^{-4}, 10^{-1}]$, a transformation between stiffness and action is introduced: $k_{\beta_m} = 10^{a_m}$. Accordingly, a_m is sampled from a normal distribution, $\mathcal{N}(a_m, \sigma)$ in the range $[-4, -1]$, so that k_{β_m} is sampled from a log-normal distribution. The neural network consists of fully connected layers with two hidden layers. The size of each hidden layer is 64 and the hyperbolic tangent (tanh) function is used for nonlinear activations. The parameters in the reward function (2.2) are $p_1 = 0.845$, $p_2 = 10\,000$, $u_{max} = 20$ and $\beta_{min} = 5^\circ$.

The initial state for initializing every episode corresponds to the limit cycle oscillation solution obtained at the end of a simulation with a constant $k_\beta = 0.015$ spanning 40 convective time units ($t = 0$ in this work denotes the instant at the end of this simulation). In advancing one control time step, $N_t = 195$ numerical time steps of the FSI solver are performed, or approximately 0.085 convective time units. That is, the control actuation is provided every 5 % of the vortex-shedding cycle. The various PPO-related parameters are the discount factor $\gamma = 0.9$, learning rate $\alpha = 0.0003$, $n_e = 10$ epochs and clipping fraction $\epsilon = 0.2$. Refer to Schulman *et al.* (2017) for the details of these parameters. The $N_\tau = 3$ trajectories are sampled in parallel. The PPO RL algorithm is implemented by using the Stable-Baselines3 library (Raffin *et al.* 2021) in Python.

We test the utility of our long–short-term strategy described in § 2.2 against the traditional long-term strategy. In the latter, the controller is optimized for a long-time horizon of 10 convective time units ($M_i = 120$) spanning approximately six vortex-shedding cycles, and weights are updated traditionally, i.e. only once in an iteration ($M_\tau = 120, k = 1$ window). On the other hand, in the long–short-term strategy, while the long-time horizon is kept the same ($M_i = 120$), the optimization is performed on two shorter optimization windows ($M_\tau = 60,\ k = 2$). For all cases, a single iteration comprising $M_i = 120$ control time steps, and $N_\tau = 3$ simultaneously running FSI simulations, each parallelized across six processors, takes approximately 2.7 hours to run on a Illinois Campus Cluster node consisting of 20 cores and 128 GB memory.

3.2. *Implementation, results and mechanisms*

To demonstrate the effectiveness of the long–short-term strategy, the evolution of the mean reward (sum of rewards divided by episode length) versus iterations for the two learning strategies as well as the passive case of $k_\beta = 0.015$ (for reference) are shown in figure 2. Firstly, we note that the evolution is oscillatory because of the stochasticity in stiffness sampling during training. Next, it can be seen that, with increasing iterations, for both cases, the controller gradually learns an effective policy as the mean reward increases beyond the passive reference case. However, the long–short-term strategy is found to exhibit faster learning as well as attain a larger reward at the end of 90 iterations as compared to the long-term one. This is because splitting the long-time horizon into two shorter windows and sequentially updating the weights for each window alleviates the burden of learning an effective policy for the entire long horizon as compared to learning via a single weight update in the long-term strategy. The remainder of the results focus on the performance of the control policy obtained after the 90th iteration of the long–short strategy. A deterministic policy is used for evaluating the true performance of

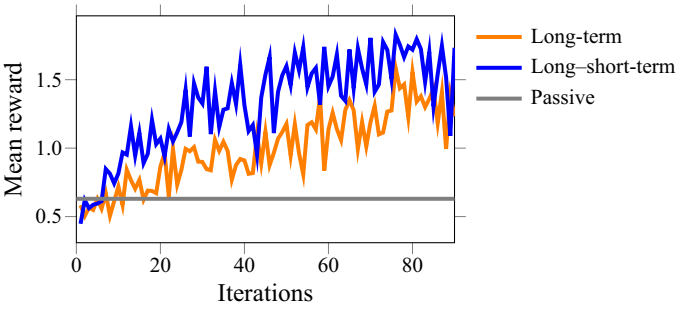

Figure 2. Evolution of mean rewards with iterations for different training strategies.

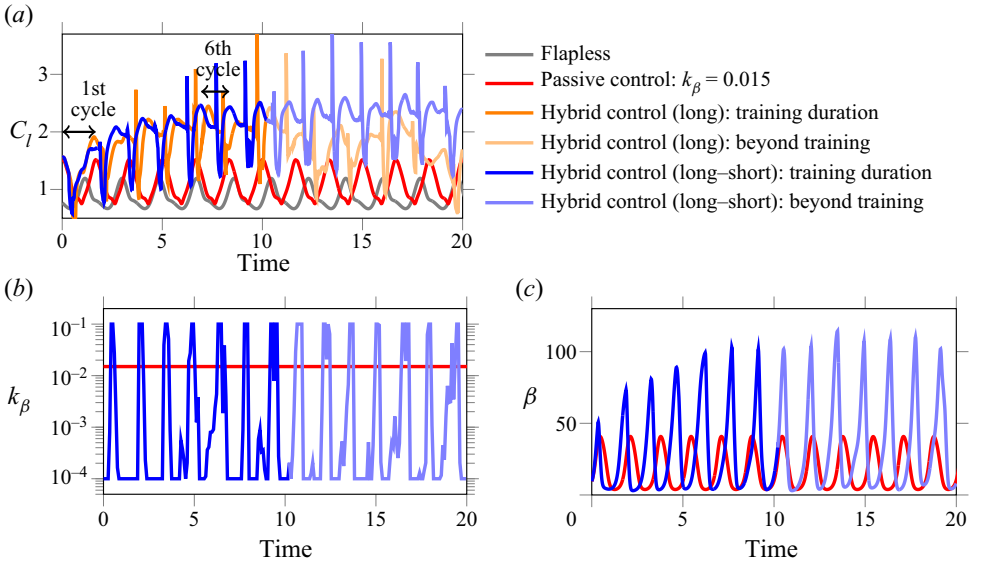

Figure 3. Temporal plots for flapless (no control), passive and hybrid flow control cases: (*a*) airfoil lift coefficient, (*b*) hinge stiffness and (*c*) flap deflection.

the controller, where the actuation provided by the neural network is directly used as the stiffness instead of stochastically sampling a suboptimal stiffness in training.

The airfoil lift of the flapless, maximal passive control and hybrid control cases are plotted in figure 3(*a*). For hybrid control, the lift is plotted not only for $t \in [0, 10]$, for which the controller has been trained, but also for $t \in [10, 20]$. It can be seen that the hybrid controller trained using the long–short-term strategy is able to significantly increase the lift in the training duration and beyond. Overall, in $t \in [0, 20]$, a significant lift improvement of 136.43 % is achieved as compared to the flapless case. For comparison, the corresponding lift improvement of the best passive case is 27 % (Nair & Goza 2022*a*). We note that this large lift improvement is, however, attained at the cost of increased lift fluctuations, the cause for which will be discussed at the end of this section. We emphasize that, while the maximum peak-to-trough fluctuation in a vortex-shedding cycle for the hybrid case has increased by 376 % as compared to the baseline flapless case, the standard deviation of lift in the pseudo-steady regime, $t > 10$, has increased by a relatively smaller amount of 104 %. For reference, the corresponding maximum lift fluctuations and standard deviation

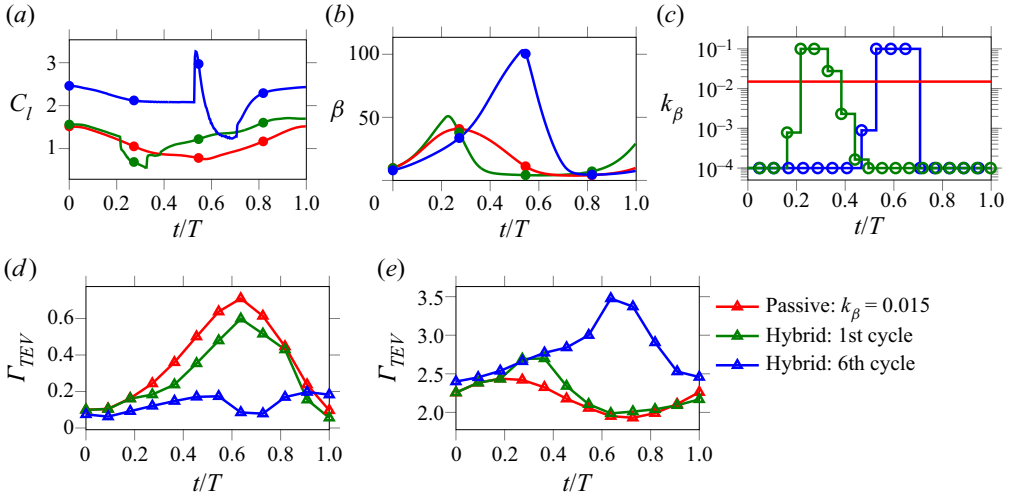

Figure 4. Time variation in various quantities during the lone periodic cycle for the passive case, and the first and sixth cycles of the hybrid case (highlighted in figure 3*a*): (*a*) lift coefficient, (*b*) flap deflection, (*c*) stiffness, (*d*) TEV strength and (*e*) LEV strength (magnitude).

for the passive case are 43.8 % and 44.2 %, respectively. The standard deviation provides a more representative measure of the magnitude of fluctuations since the spikes occur in a smaller time scale. When lift fluctuation minimization is of importance, additional penalty terms can be incorporated into the reward, as described in § 2.2.

We also show the lift of the hybrid controller trained using the long-term strategy. While it is able to provide similar large lift improvements as compared to the long–short-term strategy in the training duration, the lift drops considerably in the duration beyond training, implying that more training iterations are required for robustness. Hereafter, we discuss the results obtained from the long–short-term strategy only. The stiffness actuations outputted by the controller and the resulting flap deflection are plotted in figure 3(*b*) and 3(*c*), respectively. It can be observed for hybrid control that the stiffness varies across four orders of magnitude (as compared to fixed $k_\beta = 0.015$ in passive control) and often reaching its bounding values of $k_\beta = 10^{-4}$ and 10^{-1}. Owing to these large stiffness variations, the flap oscillates with an amplitude that is more than twice that in passive control, indicating that larger-amplitude flap oscillations can yield larger lift benefits when timed appropriately with key flow processes.

To understand the physical mechanisms driving lift benefits for the hybrid case, we show various quantities during the first and sixth vortex-shedding cycles (cf. figure 3*a*). These distinct cycles allow for a comparison of the transient and quasi-periodic regimes. The perfectly periodic dynamics of the passive $k_\beta = 0.015$ case are also shown for reference. Firstly, from figure 4(*c*), we can see that initially, in $t/T \in [0, 0.16]$, the controller actuation is lower ($k_\beta = 10^{-4}$) than the constant passive actuation ($k_\beta = 0.015$). This low stiffness prompts the hybrid flap in the first cycle to undergo a slightly larger deflection until $\beta \approx 50^\circ$ in figure 4(*b*). The decisive actuation occurs at $t/T \approx 0.21$ when the largest $k_\beta = 10^{-1}$ is prescribed (cf. figure 4*c*), which forces the flap to oscillate downwards within a short time span until $t/T = 0.4$ (cf. figure 4*b*). The flap then begins to rise only after the actuation is reduced back to $k_\beta = 10^{-4}$ by $t/T = 0.5$ (cf. figure 4*c*). For comparison, the

rising and falling of the single-stiffness flap in the same duration of $t/T \in [0, 0.5]$ occurs gradually (cf. figure 4*b*).

To understand the effect of such an aggressive flapping mechanism on airfoil lift, we plot the circulation strengths of the trailing- and leading-edge vortices (TEV and LEV, respectively) in figures 4(*d*) and 4(*e*), respectively. Here, Γ_{TEV} and Γ_{LEV} are the magnitudes of positive and negative circulation strengths evaluated in bounding boxes, $[0.85, 1.1]c \times [-0.35, -0.1]c$ and $[0, 1.1]c \times [-0.35, 0.2]c$, respectively. It can be observed that, after $t/T \approx 0.18$ when the flap strongly oscillates downwards in the first cycle, Γ_{TEV} and Γ_{LEV} for the hybrid case are decreased and increased, respectively, as compared to the passive case. The overall effect on performance is that the lift of the hybrid case in the first cycle begins to increase at $t/T \approx 0.4$ after an initial dip, as seen in figure 4(*a*).

Now, as time progresses, this aggressive flapping mechanism continues, but with increasing amplitude, as observed in figure 3(*c*). Eventually, by the sixth cycle, the switching in stiffness occurs at delayed time instants of $t/T \approx 0.53$ and 0.7 (cf. figure 4*c*). As a consequence, the flap oscillates upwards until $t/T = 0.5$ and attains a large deflection of $\beta \approx 100^\circ$ (cf. figure 4*b*). Then, as the stiffness switches to high $k_\beta = 10^{-1}$, the flap deflection suddenly drops to $\beta \approx 5^\circ$ within a short time span. Similar to the first cycle, this strong downward motion mitigates the TEV and enhances the LEV, as seen in figures 4(*d*) and 4(*e*), respectively, but now to a much stronger degree.

To visualize these effects, vorticity contours at four time instants in the sixth cycle are plotted in figures 5(*a*)–5(*d*) and compared to passive control in figures 5(*e*)–5(*h*). The TEV, which is clearly decipherable for the passive case in figure 5(*g*), is now limited to a much smaller size in the hybrid case in figure 5(*c*). This is because the strong angular velocity of the downward-oscillating flap sheds away the TEV quickly and restricts its growth. This downward motion further contributes to a reduced width of the separated recirculation region in the airfoil-normal direction in figure 5(*a*) as compared to passive control in figure 5(*e*). Owing to the TEV mitigating, LEV enhancing and separation width reducing mechanisms of hybrid flow control, the airfoil attains a much higher lift by the sixth cycle as compared to the passive case (cf. figure 4*a*).

Now, we briefly discuss the occurrence of the spikes observed in the lift signal in figure 3(*a*) by focusing on the spike in the sixth cycle in figure 4(*a*), visualized via C_p contours at four time instants surrounding the occurrence of the spike in figures 6(*a*)–6(*d*). We note that the spike initiates at $t/T \approx 0.53$ when the flap attains its highest deflection (cf. figures 4*a* and 4*b*). Preceding this instant, as the flap oscillates upwards, it induces the fluid in the vicinity to move upstream due to its no-slip condition. When the flap comes to a sudden stop, as the stiffness switches from 10^{-4} to 10^{-1} at $t/T \approx 0.53$ (cf. figure 4*c*), the moving fluid in the region post-flap abruptly loses its momentum. This momentum loss is manifested as a strong rise in post-flap pressure (cf. figure 6*b*). The fluid in the pre-flap region, however, does not experience a barrier in its upstream motion, and instead continues to roll up and builds up the suction pressure. The flap dividing the high-pressure post-flap and low-pressure pre-flap regions can be clearly seen in figures 6(*a*) and 6(*c*), respectively. This large pressure difference across the flap contributes to the large spikes in the airfoil lift.

Finally, we remark on the effectiveness of the control strategy learned in 2-D conditions on three-dimensional (3-D) flows past a wing. At similar Reynolds numbers, for sufficiently high aspect ratio ($AR \geq 3$) and a large angle of attack of 20°, the separation process is similar to that in 2-D, where vortex shedding from the leading and trailing edges is more dominant than the wing-tip vortices (Taira & Colonius 2009; Zhang *et al.* 2020).

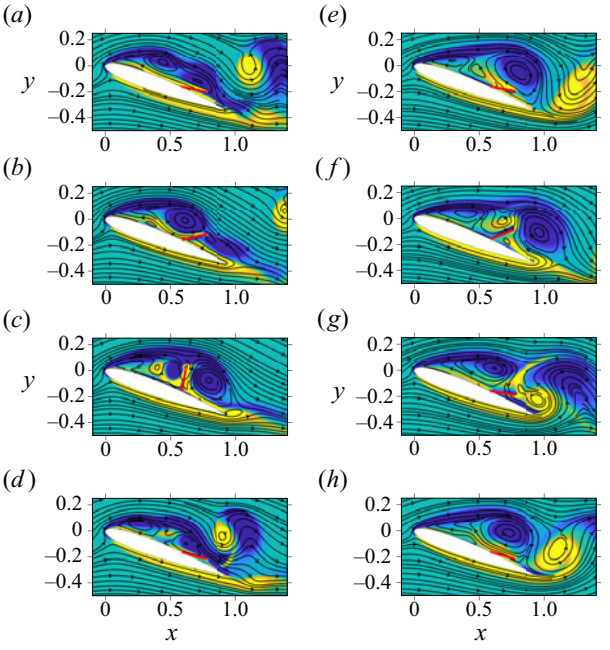

Figure 5. Vorticity contours for hybrid control during the sixth cycle and passive single-stiffness control at four instants indicated by the markers in figure 4(*a*). Hybrid: (*a*) $t = 0T$, (*b*) $t = 0.27T$, (*c*) $t = 0.55T$ and (*d*) $t = 0.82T$. Passive: (*e*) $t = 0T$, (*f*) $t = 0.27T$, (*g*) $t = 0.55T$ and (*h*) $t = 0.82T$.

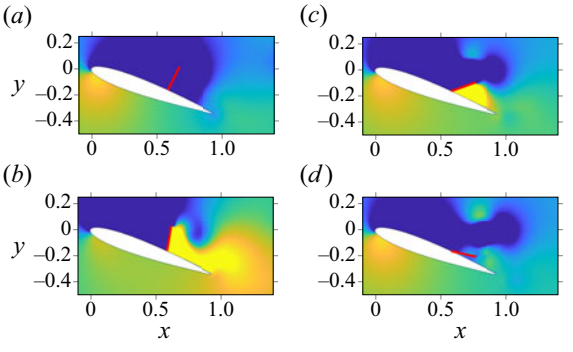

Figure 6. The C_p contours at four time instants in the sixth cycle of hybrid control. Hybrid: (*a*) $t = 0.45T$, (*b*) $t = 0.55T$, (*c*) $t = 0.64T$ and (*d*) $t = 0.73T$.

In this regime, it is plausible that the actuation strategy learnt in 2-D would provide lift benefits in the 3-D flow, though these would presumably be alleviated by tip-vortex effects. For smaller aspect ratios ($AR = 1$), the tip vortex dominates and qualitatively changes the separated flow behaviour. In this case, the 2-D-trained controller is unlikely to yield lift benefits. However, such aspect ratios are significantly smaller than the bio-inspired flows of interest. Moreover, actuation with flaps has the potential to be a useful paradigm in this regime as well (Arivoli & Singh 2016), due to similar length scales of flow separation, vortex formation and interaction processes between 2-D and 3-D. We also emphasize that the RL framework is agnostic to the environment, and provided that a 3-D solver were

developed to simulate the dynamics, the same training procedure could be performed without modification.

4. Conclusions

A hybrid active–passive flow control method was introduced as an extension of the covert-inspired passive flow control method, consisting of a torsionally mounted flap on an airfoil at post-stall conditions involving vortex shedding. This hybrid strategy consisted of actively actuating the hinge stiffness to passively control the dynamics of the flap. A closed-loop feedback controller trained using deep RL was used to provide effective stiffness actuations to maximize lift. The RL framework was described, including modifications to the traditional RL methodology that enabled faster training for our hybrid control problem. The hybrid controller provided lift improvements as high as 136 % and 85 % with respect to the flapless airfoil and the maximal passive control (single-stiffness) cases, respectively. These lift improvements were attributed to large flap oscillations due to stiffness variations occurring over four orders of magnitude. Detailed flow analysis revealed an aggressive flapping mechanism that led to significant TEV mitigation, LEV enhancement and reduction of separation region width. We remark that, since the stiffness changes can be well approximated by a small number of finite jumps, a discrete VSA could be a pathway to realizing this actuation strategy.

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

Author ORCIDs.
Nirmal J. Nair https://orcid.org/0000-0003-4431-4020;
Andres Goza https://orcid.org/0000-0002-9372-7713.

References

Ahuja, S. & Rowley, C.W. 2010 Feedback control of unstable steady states of flow past a flat plate using reduced-order estimators. *J. Fluid Mech.* **645**, 447–478.

Arivoli, D & Singh, I. 2016 Self-adaptive flaps on low aspect ratio wings at low Reynolds numbers. *Aerosp. Sci. Technol.* **59**, 78–93.

Baumeister, T., Brunton, S.L. & Kutz, J.N. 2018 Deep learning and model predictive control for self-tuning mode-locked lasers. *J. Opt. Soc. Am.* B **35** (3), 617–626.

Bieker, K., Peitz, S., Brunton, S.L., Kutz, J.N. & Dellnitz, M. 2020 Deep model predictive flow control with limited sensor data and online learning. *Theor. Comput. Fluid Dyn.* **34**, 1–15.

Bramesfeld, G. & Maughmer, M.D 2002 Experimental investigation of self-actuating, upper-surface, high-lift-enhancing effectors. *J. Aircraft* **39** (1), 120–124.

Diller, S., Majidi, C. & Collins, S.H. 2016 A lightweight, low-power electroadhesive clutch and spring for exoskeleton actuation. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 682–689. IEEE.

Duan, C. & Wissa, A. 2021 Covert-inspired flaps for lift enhancement and stall mitigation. *Bioinspir. Biomim.* **16**, 046020.

Fan, D., Yang, L., Wang, Z., Triantafyllou, M.S. & Karniadakis, G.E. 2020 Reinforcement learning for bluff body active flow control in experiments and simulations. *Proc. Natl Acad. Sci.* USA **117** (42), 26091–26098.

Garnier, P., Viquerat, J., Rabault, J., Larcher, A., Kuhnle, A. & Hachem, E. 2021 A review on deep reinforcement learning for fluid mechanics. *Comput. Fluids* **225**, 104973.

Ham, R., Sugar, T., Vanderborght, B., Hollander, K. & Lefeber, D. 2009 Compliant actuator designs. *IEEE J. Rob. Autom. Mag.* **3** (16), 81–94.

Kim, J. & Bewley, T.R. 2007 A linear systems approach to flow control. *Annu. Rev. Fluid Mech.* **39**, 383–417.

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

Mohan, A.T. & Gaitonde, D.V. 2018 A deep learning based approach to reduced order modeling for turbulent flow control using LSTM neural networks. arXiv:1804.09269.

Nair, N.J. & Goza, A. 2022*a* Fluid–structure interaction of a bio-inspired passively deployable flap for lift enhancement. *Phys. Rev. Fluids* **7**, 064701.

Nair, N.J. & Goza, A. 2022*b* A strongly coupled immersed boundary method for fluid–structure interaction that mimics the efficiency of stationary body methods. *J. Comput. Phys.* **454**, 110897.

Nota, C. & Thomas, P.S. 2019 Is the policy gradient a gradient? arXiv:1906.07073.

Paris, R., Beneddine, S. & Dandois, J. 2021 Robust flow control and optimal sensor placement using deep reinforcement learning. *J. Fluid Mech.* **913**, A25.

Pawar, S. & Maulik, R. 2021 Distributed deep reinforcement learning for simulation control. *Mach. Learn. Sci. Technol.* **2** (2), 025029.

Peitz, S., Otto, S.E. & Rowley, C.W. 2020 Data-driven model predictive control using interpolated koopman generators. *SIAM J. Appl. Dyn. Syst.* **19** (3), 2162–2193.

Rabault, J., Kuchta, M., Jensen, A., Réglade, U. & Cerardi, N. 2019 Artificial neural networks trained through deep reinforcement learning discover control strategies for active flow control. *J. Fluid Mech.* **865**, 281–302.

Rabault, J. & Kuhnle, A. 2019 Accelerating deep reinforcement learning strategies of flow control through a multi-environment approach. *Phys. Fluids* **31** (9), 094105.

Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M. & Dormann, N. 2021 Stable-baselines3: reliable reinforcement learning implementations. *J. Mach. Learn. Res.* **22** (268), 1–8.

Rosti, M.E., Omidyeganeh, M. & Pinelli, A. 2018 Passive control of the flow around unsteady aerofoils using a self-activated deployable flap. *J. Turbul.* **19** (3), 204–228.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A. & Klimov, O. 2017 Proximal policy optimization algorithms. arXiv:1707.06347.

Sun, J., Guan, Q., Liu, Y. & Leng, J. 2016 Morphing aircraft based on smart materials and structures: a state-of-the-art review. *J. Intell. Mater. Syst. Struct.* **27** (17), 2289–2312.

Sutton, R.S. & Barto, A.G. 2018 *Reinforcement Learning: An Introduction*. MIT Press.

Taira, K. & Colonius, T.I.M 2009 Three-dimensional flows around low-aspect-ratio flat-plate wings at low reynolds numbers. *J. Fluid Mech.* **623**, 187–207.

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

Viquerat, J., Rabault, J., Kuhnle, A., Ghraieb, H., Larcher, A. & Hachem, E. 2021 Direct shape optimization through deep reinforcement learning. *J. Comput. Phys.* **428**, 110080.

Wolf, S., *et al.* 2015 Variable stiffness actuators: review on design and components. *IEEE/ASME Trans. Mech.* **21** (5), 2418–2430.

Zhang, K., Hayostek, S., Amitay, M., He, W., Theofilis, V. & Taira, K. 2020 On the formation of three-dimensional separated flows over wings under tip effects. *J. Fluid Mech.* **895**, A9.

Zhu, Y., Tian, F.-B., Young, J., Liao, J.C. & Lai, J.C.S. 2021 A numerical study of fish adaption behaviors in complex environments with a deep reinforcement learning and immersed boundary–lattice boltzmann method. *Sci. Rep.* **11** (1), 1–20.