






## Learning swimming escape patterns for larval fish under energy constraints

Ioannis Mandralis , Pascal Weber , Guido Novati , and Petros Koumoutsakos <sup>\*</sup>

*Computational Science and Engineering Laboratory, ETH Zürich, CH-8092, Switzerland  
and School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*

 (Received 23 November 2020; accepted 26 August 2021; published 20 September 2021)

Swimming organisms can escape their predators by creating and harnessing unsteady flow fields through their body motions. Stochastic optimization and flow simulations have identified escape patterns that are consistent with those observed in natural larval swimmers. However, these patterns have been limited by the specification of a particular cost function and depend on a prescribed functional form of the body motion. Here, we deploy reinforcement learning to discover swimmer escape patterns for larval fish under energy constraints. The identified patterns include the C-start mechanism, in addition to more energetically efficient escapes. We find that maximizing distance with limited energy requires swimming via short bursts of accelerating motion interlinked with phases of gliding. The present, data efficient, reinforcement learning algorithm results in an array of patterns that reveal practical flow optimization principles for efficient swimming and the methodology can be transferred to the control of aquatic robotic devices operating under energy constraints.

DOI: [10.1103/PhysRevFluids.6.093101](https://doi.org/10.1103/PhysRevFluids.6.093101)

### I. INTRODUCTION

Aquatic organisms involved in predator-prey interactions perform impressive feats of fluid manipulation to enhance their chances of survival [1–8]. Since early studies where prey fish were reported to rapidly accelerate from rest by bending into a C shape and unfurling their tail [9–12], impulsive locomotion patterns have been the subject of intense investigation. Studying escape strategies of prey fish has led to the discovery of sensing mechanisms [13–15], dedicated neural circuits [16–19], and biomechanic principles [20,21]. From the perspective of hydrodynamics, several studies have sought to understand the C-start escape response and how it imparts momentum to the surrounding fluid [22–27].

Despite the large volume of literature on the C-start escape response, experiments and observations indicate that swimming escapes can take a variety of forms. For example, after the initial burst from rest, many fish are seen coasting instead of swimming continuously [11,28,29]. Furthermore, theoretical and experimental studies have suggested that intermittent swimming styles, termed burst-coast swimming, can be more efficient than continuous swimming when maximizing distance given a fixed amount of energy [30–33]. This raises the question of when and why different swimming escape patterns are employed in nature, and which biophysical cost functions they optimize.

This question has been investigated by using reverse engineering methodologies to identify links between biophysical cost functions and resulting swimming patterns. For example, fast and efficient swimming motions [34,35] or the C-start escape response [36] have been reverse engineered based on the appropriate objective functions. In particular, when reverse engineering the C-start escape

---

<sup>\*</sup>petros@seas.harvard.edu

response, the parameters of a *predefined* motion sequence were optimized to maximize the escape distance. While the identified escape pattern was consistent with larval fish escapes observed in nature, the reverse engineering method employed had inherent limitations: It required an *a priori* defined objective function and a specification of the functional form for the two stages of the observed escape patterns. As a result, the full space of swimming escapes remained unexplored since reverse engineered escape patterns could only reside within the predefined design space and depended strongly on the underlying parametric assumptions.

In recent years, reinforcement learning (RL), an alternative framework for *learning* behavior based on objective functions, has emerged. RL has found an application for a variety of problems related to fluid mechanics, including learning natural swimming and flying behaviors [37–44] and optimizing and controlling engineering flow systems [45–48]. In the RL framework, decision making processes are viewed as multistage optimizations instead of one-shot optimizations that require rigidly predefined motion parameters. Due to this flexibility, RL circumvents many of the limitations of classical reverse engineering and offers a novel way through which to explore the space of swimming escape patterns.

In this paper, we introduce reinforcement learning (RL) to the study of escape responses. In particular, we employ the RL framework to understand the links between objective functions and swimming escape patterns for larval fish. The fish escape is formulated as an incremental process where a swimming agent receives information about the flow field and learns to maximize its cumulative reward autonomously. By endowing the swimming agent with limited energy and rewarding the escape distance, we find that burst-coast swimming escapes, consisting of rapid body accelerations through C bends followed by powerless gliding, maximize escape distance when the available energy is limited, in alignment with theoretical predictions [30–32]. Furthermore, we find that the RL algorithm is able to produce a wide array of different escape patterns, according to the amount of energy available, due to its inherent generalization capability. This “kaleidoscope” of escape patterns sheds light on key mechanisms which are responsible for rapid propulsion in fluids and evidences the fundamental advantage of using RL to provide links between objective functions and biological behavior.

## II. SWIMMER MODEL AND KINEMATICS

The geometry of the artificial swimmer, displayed in Fig. 1(a), is modeled after a 5-day-postfertilization zebrafish (see Appendix A for details). The swimmer propels itself by modifying its instantaneous midline curvature  $\kappa(s, t) \in \mathbb{R}$ , a quantity which imitates muscle contractions in natural anguilliform swimmers [34]. The midline curvature  $\kappa(s, t)$ , displayed in Eq. (1), is further decomposed into a baseline component  $B(s, t) \in \mathbb{R}$  and an undulatory component  $K(s, t) \in \mathbb{R}$ . This allows the swimmer to bend unilaterally and to undulate sinusoidally,

$$\kappa(s, t) = B(s, t) + K(s, t) \sin \{2\pi [t/T_{\text{prop}} - s\tau_L(t)/L] + \phi(t)\}. \quad (1)$$

The baseline curvature  $B(s, t)$  and the undulatory curvature  $K(s, t)$  are modeled as natural cubic splines defined by six control points on the swimmer midline [see Fig. 1(a)]. To imitate the stiff head, neck, and tail of larval zebrafish, the curvature is set to zero at  $s_1 = 0$ ,  $s_2 = 0.2L$ ,  $s_6 = L$ , leaving three free control points at which the curvature can be controlled. The midline curvature is parametrized by  $T_{\text{prop}} \in \mathbb{R}$ , the undulatory swimming period  $L$ , the swimmer length  $\tau_L(t)$ , the delay of the tail, and  $\phi(t)$ , the phase of the sinusoid. Given the curvature  $\kappa(s, t)$ , the midline coordinates of the swimmer are retrieved by solving the Frenet-Serret formulas [35]. As per *in vivo* observations of 5-day-postfertilization zebrafish [49], the swimmer length is set to  $L = 4.4$  mm and the propulsive swimming period to  $T_{\text{prop}} = 44$  ms. The resulting swimming Reynolds number, defined as  $\text{Re} = \frac{L^2}{T_{\text{prop}}\nu}$ , where  $\nu$  is the kinematic viscosity of water, is chosen as  $\text{Re} = 550$ . This places the swimmer in the intermediate flow regime where both viscous and inertial forces have important effects. The

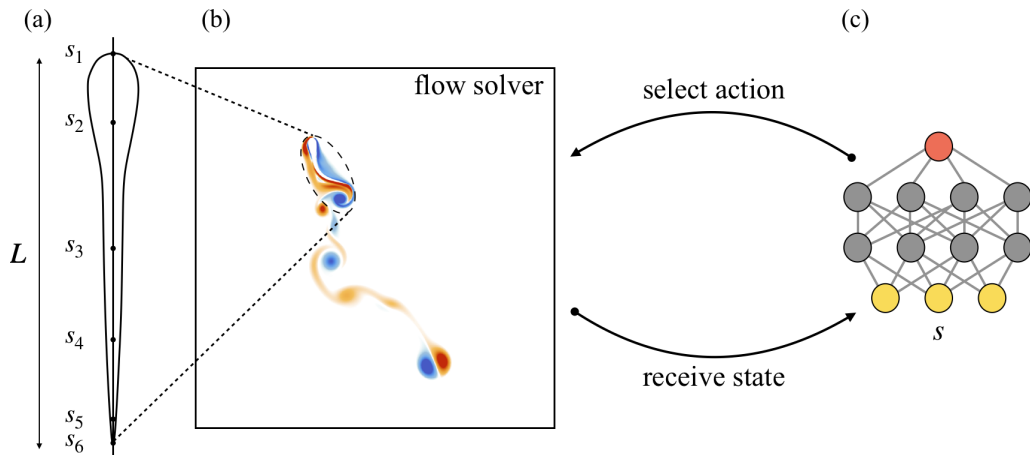


FIG. 1. Interaction between flow solver and reinforcement learning. (a) Geometrical model of a 5-day-postfertilization zebrafish larva [36]. The six curvature control points are indicated from top to bottom. (b) Simulation environment. The swimmer is placed inside a square domain and is simulated using the numerical flow solver described in Appendix A. (c) This is coupled with a deep reinforcement learning algorithm, that receives a state from the flow solver and sends back an action, deciding how the swimmer should act within the simulation.

Navier-Stokes equations are solved by performing a direct numerical simulation on a uniform grid (see Appendix A for details).

### III. REINFORCEMENT LEARNING FOR SWIMMING ESCAPES

In the RL framework, an agent learns to earn rewards through trial-and-error interaction with its environment. The agent chooses an action  $\mathbf{a}_k \in \mathcal{A}$  at discrete time instances  $k \in \mathbb{N}$  by sampling a stochastic control policy  $\pi(\mathbf{a}|s_k)$  that is conditioned on its current state  $s_k \in \mathcal{S}$ , i.e.,  $\mathbf{a}_k \sim \pi(\cdot|s_k)$ . Given the action, the environment transitions to a new state determined by the dynamics function  $D$ , i.e.,  $s_{k+1} \sim D(\cdot|\mathbf{a}_k, s_k)$ . Upon transition, the agent receives a reward signal  $r_{k+1} \in \mathbb{R}$ . The goal is to learn the optimal control policy  $\pi^*(\mathbf{a}|s)$  which maximizes the action-value function  $Q(s, \mathbf{a}) \in \mathbb{R}$ , defined as the expected long-term cumulative reward when starting from state  $s$  and taking action  $\mathbf{a}$

$$Q(s, \mathbf{a}) = \mathbb{E}_{\mathbf{a}_k \sim \pi} s_{k+1} \sim D \left[ \sum_{k=0}^N \gamma^k r_k | s_0 = s, \mathbf{a}_0 = \mathbf{a} \right]. \quad (2)$$

Here,  $\gamma \in [0, 1)$  is the discounting factor which quantifies the tradeoff between immediate and future rewards. We use a state of the art and data efficient RL algorithm (remember and forget experience replay) to identify the optimal policy [50] (see Appendix B for details).

To model the escape behavior of a prey, the swimmer is trained to maximize the distance away from its initial position after a fixed time. In addition, the work done by the swimmer on the fluid is limited based on an escape energy budget  $E_0$ . This imitates how the muscle fibers used during fish escapes can be fatigued by lactic acid buildup as glycogen stores are depleted [32]. We normalize these budgets by  $E_0$  which is the energy expended during the C-start escape sequence reported by Gazzola *et al.* [36] that closely matches larval zebrafish escapes observed in nature ( $E_0 = 14.04$ ,  $d_0 = 1.15L$ ; see Appendix B for details on computation and nondimensionalization).

Each escape episode proceeds by first sampling an energy budget uniformly in a range around  $E_0$  ( $[\frac{1}{3}E_0, 3E_0]$ ), allowing the swimmer to interact with the fluid for a fixed time period, and finally rewarding the overall distance traveled ( $r = d$ ). If the energy budget is depleted before the allocated

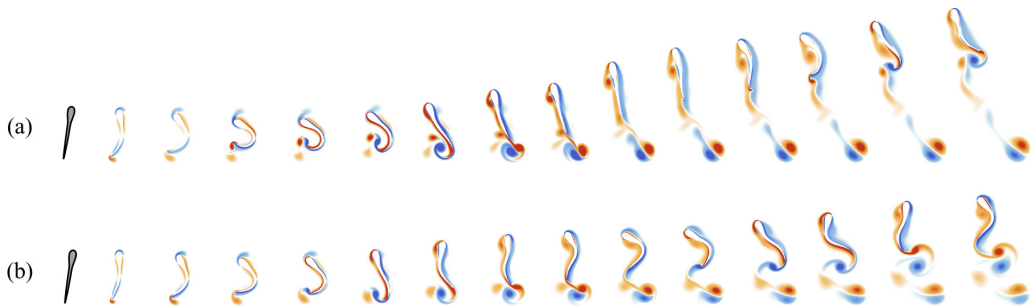


FIG. 2. Temporal evolution of the vorticity field for the optimized and learned escape patterns. (a) RL burst-coast escape pattern obtained by RL swimmer using energy budget  $E_0$ . (b) C-start escape pattern obtained by simulating the optimized parameter set reported in [36]. Orange regions represent positive flow vorticity and blue regions represent negative vorticity.

escape time is surpassed, the episode terminates prematurely (see Appendix B for training details). During each escape, the swimmer controls its body through a nine dimensional action vector that schedules changes in the shape of the curvature as well as the delay and phase of the sinusoidal motion. The swimmer senses the environment through a set of states (current distance and polar angle from starting position, body orientation, mean forward and angular velocity, as well as the remaining energy budget) and receives as reward its distance from the start, at the end of each episode.

#### IV. LEARNED VS OPTIMIZED ESCAPES

The vorticity field generated by the trained RL swimmer when escaping with energy budget  $E_0$  is displayed in Fig. 2(a). The swimmer initially forms a C bend and subsequently unfurls its tail, propelling a counter-rotating vortex pair opposite to the direction of its forward motion and coasting. Only when its speed drops significantly does the learned swimmer undulate its body further in an attempt to extend its forward motion. This learned motion sequence is termed the *burst-coast* escape pattern (see Video 1 [51] for full animation). In the following, we compare the burst-coast escape pattern obtained using RL to the *C-start* escape pattern which was found by optimizing a predefined motion sequence in [36].

The vorticity fields of the two escape sequences are displayed in Fig. 2(b). Figure 2(b) indicates that, although the RL swimmer coasts after the initial burst instead of swimming continuously, the C-bend starting pattern is remarkably close to that of the C start, without having been enforced through predefined motion parameters. The two escape patterns differ in that the C start has a continuous swimming phase which propels two counter-rotating vortex dipoles away from the swimmer body, while the burst-coast escape pattern propels only one counter-rotating vortex dipole which is created in the initial burst.

Moreover, the burst-coast strategy results in a greater escape distance than the C start while using an equal amount of energy, as shown in Fig. 3(a). Since both escapes consume equal energy and conform to natural mechanic or geometric constraints (see Appendix A), they are both feasible in a hypothetical predator-prey encounter. In fact, the use of both intermediate swimming or coasting phases are observed for various fish species in predator escapes [11,28,29].

In which situations is it advantageous to employ the C-start or the burst-coast escape patterns? We found that, at the characteristic swimming Reynolds number for larval zebrafish,  $Re = 550$  [36], the C start sets the swimmer into motion faster than the burst coast pattern [circled region in Fig. 3(a)]. In contrast, the RL swimmer uses a burst-coast pattern which more efficiently solves the task we set out for it—instead of propelling itself forward quickly, it delays the onset of its motion in order to form a more pronounced C shape, evidenced by the higher peak in midpoint curvature in Fig. 3(a), and achieves a greater overall distance using the same amount of energy. This underscores the

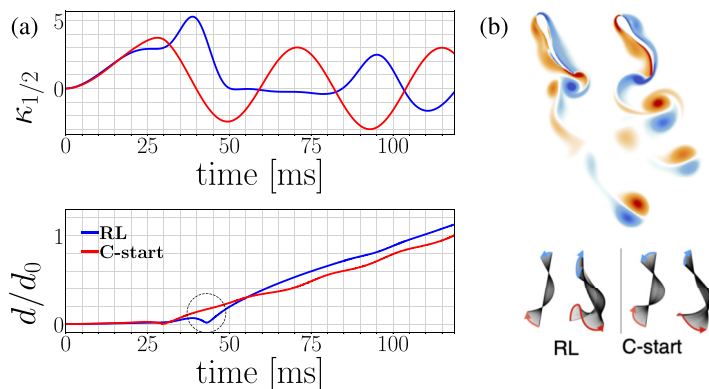


FIG. 3. Comparison of the equal-energy RL burst-coast and C-start escape patterns. (a) Swimmer midpoint curvature  $\kappa_{1/2}$  and normalized distance (in terms of swimmer length) as a function of time. The energy expended by the RL escape (blue) and C-start escape (red) is in both cases equal to  $E_0$ . (b) Vorticity fields and midline profiles for the RL escape (left) and the C-start escape (right). For both escapes, the midline of the swimmer is plotted over time within two intervals: The preparatory interval ( $t \leq 30$  ms) and the propulsive interval ( $30 \text{ ms} \leq t \leq 50$  ms)

energetic advantage of the burst-coast escape pattern as compared to continuous swimming (see Sec. V).

## V. INFLUENCE OF REYNOLDS NUMBER ON ESCAPE PATTERNS

We recorded the motion sequence of the burst-coast RL escape with energy budget  $E_0$  at the original viscosity  $\nu_0$  (corresponding to Reynolds number  $\text{Re}_0 = 550$ ), and simulated it along with the C start for viscosities in the range  $\nu \in [0.1\nu_0, 5\nu_0]$ . Based on the average speed during each escape the  $\text{Re}_{\bar{v}} = \frac{\bar{v}L}{\nu}$  was computed. The normalized distance traveled per unit energy consumed  $\tilde{d}/\tilde{E}$  is plotted for each escape against the mean Reynolds number  $\text{Re}_{\bar{v}}$  in Fig. 4. This uncovers

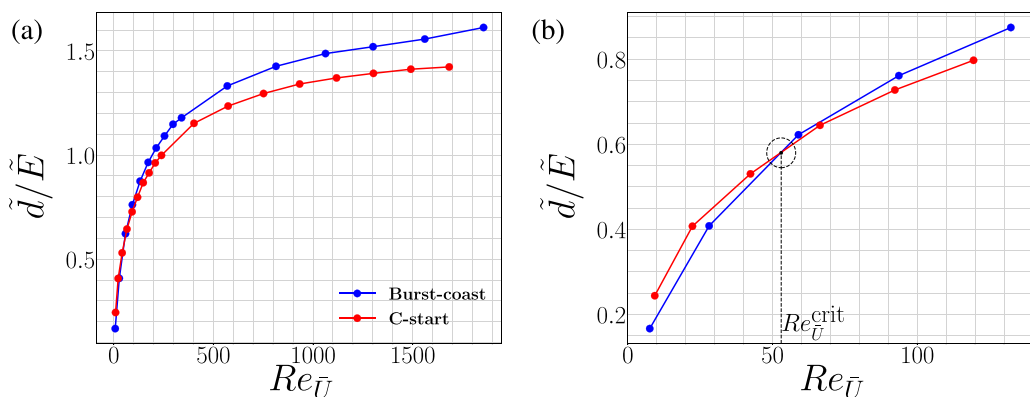


FIG. 4. Influence of Reynolds number on escape efficiency. (a) The normalized distance per unit energy  $\tilde{d}/\tilde{E}$  is plotted as a function of the mean Reynolds number  $\text{Re}_{\bar{v}}$ .  $\tilde{d} = d/d_0$  and  $\tilde{E} = E/E_0$ . In blue: The burst coast escape sequence recorded by evaluating the RL control policy with energy budget  $E_0$  at  $\nu_0$ , simulated for different viscosities. In red: The C-start escape sequence obtained by optimizing at  $\nu_0$  [36], simulated at different viscosities. (b) A zoomed-in segment of (a) on the range  $\text{Re}_{\bar{v}} \in [0, 150]$ . Displays the critical Reynolds number  $\text{Re}_{\bar{v}}^{\text{crit}}$  at which the burst coast sequence becomes more efficient than the C start.

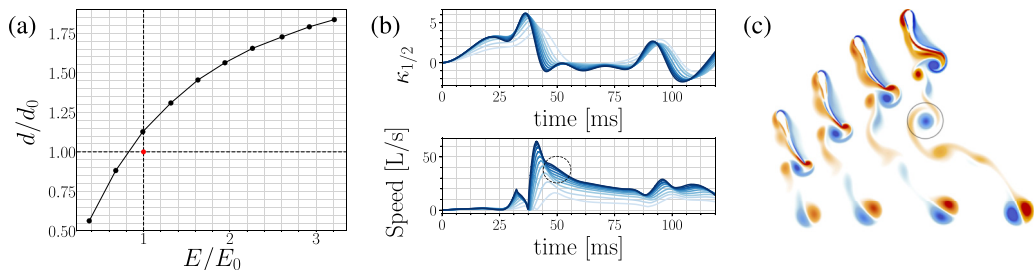


FIG. 5. Influence of energy budget on escape distance and learned strategy. (a) Normalized escape distance as a function of normalized energy expenditure for ten escapes with energy budgets linearly spaced between  $\frac{1}{3}E_0$  and  $3E_0$ . (b) Swimmer midpoint curvature  $\kappa_{1/2}$ , and swimmer speed as a function of time for the ten escapes. The higher energy escapes are represented by dark blue while the lower energy escapes are represented with light blue. The lightest blue curve corresponds to an energy budget  $\frac{1}{3}E_0$ , and the darkest blue corresponds to an energy budget  $3E_0$ . All quantities are plotted for escapes of duration 118.8 ms and are simulated using the same RL policy (for training details see Appendix B). (c) Vorticity fields in the wake of four escapes of energy budgets, from left to right,  $\frac{1}{3}E_0$ ,  $\frac{9}{10}E_0$ ,  $\frac{6}{5}E_0$ , and  $3E_0$ . Snapshots of the vorticity field are taken at  $t = 75$  ms for all four escapes. The dotted, circled region denotes a secondary vortical structure which emerges at higher energy budgets.

two distinct regimes: The low Reynolds number regime  $\text{Re}_{\bar{U}} \lesssim \text{Re}_{\bar{U}}^{\text{crit}} = 50$ , and the high Reynolds number regime  $\text{Re}_{\bar{U}} \gtrsim \text{Re}_{\bar{U}}^{\text{crit}} = 50$ . In the high Reynolds number regime the discrepancy between the burst-coast escape sequence and the C start increases. This finding is consistent with the theoretical considerations of Weihs [30–32] who predicted the existence of a transition regime,  $\text{Re}_{\bar{U}} \in [20, 200]$ , in which burst coast swimming becomes more efficient than continuous swimming.

As the Reynolds number increases further, the gap between the burst coast and the C-start escape patterns continues to grow. This transition can be attributed to the increased efficiency of the intermediate coasting phase. In fact, for  $\text{Re}_{\bar{U}} \geq 200$ , the drag coefficient of a rigidly gliding fish can be up to four times smaller than that of an actively swimming fish [33,52,53]. This is thought to drive changes in the swimming style of fish who, when growing in length, transition from the viscous hydrodynamic regime to the inertial hydrodynamic regime and replace their continuous swimming style with predominantly burst coast swimming [31]. This biological transition has been termed an *adaptive energy sparing mechanism* [31]. For the high Reynolds number regime, our analysis supports this hypothesis, suggesting that burst-coast swimming patterns convert limited energy into greater overall distance compared to patterns involving continuous swimming. Further study to corroborate this hypothesis is needed since the model used is two dimensional, and fish change their body form during growth [54].

## VI. GENERALIZING ESCAPE PATTERNS ACROSS ENERGY BUDGETS

The RL swimmer was endowed with different energy budgets and the resulting escape patterns were visualized in Fig. 5(c) (see Video 2 [51] for an example escape with energy  $3E_0$ ). We found that the learned swimmer was able to modulate its motions in order to travel further at higher energies [see Fig. 5(a)]. Which aspects of the escape strategy does it modify to achieve this? Plots of the midpoint curvature for escapes of different energies [Fig. 5(b)] evidence that, as the energy budget increases, the swimmer performs more pronounced, faster C bends. As a result, the maximum speed is attained at a higher value earlier in the escape, and the overall escape distance increases. This can be seen by the peak midpoint curvature, as well as the peak speed, moving upwards and to the left as the energy budget increases [Fig. 5(b)].

This strategy can be interpreted in terms of the duty cycle defined as  $\text{DC} = T_{\text{burst}}/T_{\text{total}}$ , where  $T_{\text{burst}}$  is the time spent accelerating before the coasting phase, and  $T_{\text{total}}$  is the total time of the escape



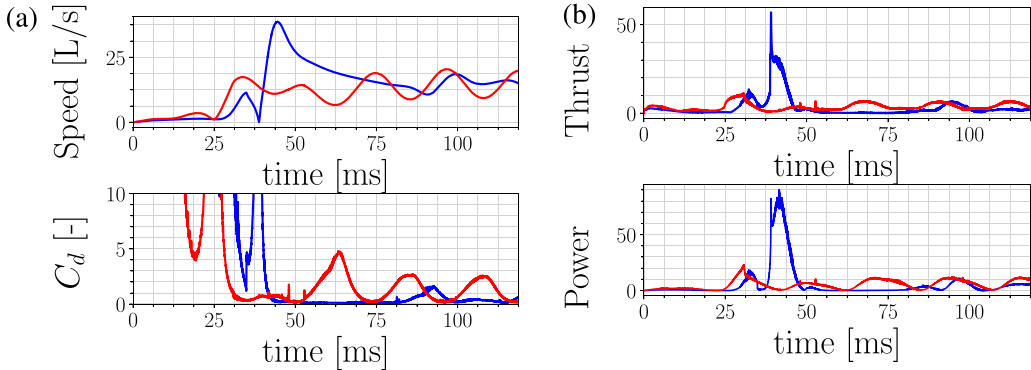


FIG. 6. Escape speed, drag coefficient, thrust force, and power consumption for equal-energy learned burst-coast pattern vs optimized C-start pattern. The energy expended by the RL escape (blue) and C-start escape (red) is in both cases equal to  $E_0$ . (a) Speed and drag coefficient of the burst-coast pattern learned through RL (blue) vs the C-start pattern (red). The speed is given in swimmer lengths per second ( $L/s$ ) and the drag coefficient is defined as  $C_d = \frac{F_d}{\rho U^2 L/2}$ , where  $F_d$  is the drag force experienced by the swimmer,  $\rho$  is the density of water,  $U$  is the swimmer speed relative to the fixed laboratory reference frame, and  $L$  is the swimmer length. (b) Thrust force and deformation power of the burst-coast pattern (blue) vs the C-start pattern (red). For details on the computation and nondimensionalization of the drag force  $F_d$ , thrust force, and deformation power, see Appendix A.

bout. In this case,  $T_{\text{burst}}$  corresponds approximately to the time where the peak speed is attained. As can be seen in Fig. 5(b), the peak speed is attained earlier for higher energies so the duty cycle decreases as the energy budget increases. In contrast, experimental data on the duty cycle of fish swimming at cruising speeds indicates that fish increase their duty cycle when higher speeds are required [55]. This suggests that, for escapes where high acceleration is required in order to increase escape distance in a fixed amount of time, the duty cycle of burst-coast strategies should be made as small as possible.

Another strategy learned by the swimmer consists in using the energy leftover from the C-bend and swim phases in order to perform slight undulations during the coasting phase. These undulations leave an imprint on the escape patterns in the form of a secondary vortical structure [circled region in Fig. 5(c)]. However, this secondary vortical structure is absent from escapes of lower energy budget. This indicates that undulatory motions during the coasting phase are second order to increasing escape distance, when compared to forming more pronounced C bends or swimming more energetically. Indeed, for high energy escapes which include undulations during the coast phase ( $E_{\text{budget}} \geq E_0$ ), the swimmer already performs C bends very close to the geometric or mechanical limit  $\kappa = 2\pi$  and undulates with curvature close to  $\kappa = 2.5$  in the final swim phase. Thus, not being able to perform a C bend past  $\kappa = 2\pi$  due to mechanical constraints, and swim with curvature past  $\kappa = 2.5$  during the final phase because of increased drag, the leftover energy is used to undulate during the coast phase. This extends the high-speed phase of the escape [circled in Fig. 5(b)] and improves the overall distance. The saturation of the C bend to the max possible curvature  $\kappa = 2\pi$  and the swim phase to  $\kappa = 2.5$  are possible factors causing the rate of conversion of energy to distance to flatten off as the energy budget increases (Fig. 5).

## VII. PROPULSIVE MECHANISMS

The escape speed, drag coefficient, thrust, and power were computed for the RL burst-coast pattern and compared to the C-start escape pattern with the optimal parameters from [36] in Fig. 6. The RL swimmer produces a significantly higher peak thrust force than the C start during the starting phase of the escape [Fig. 6(b)]. As alluded to in Sec. IV, this is a result of the RL swimmer curling

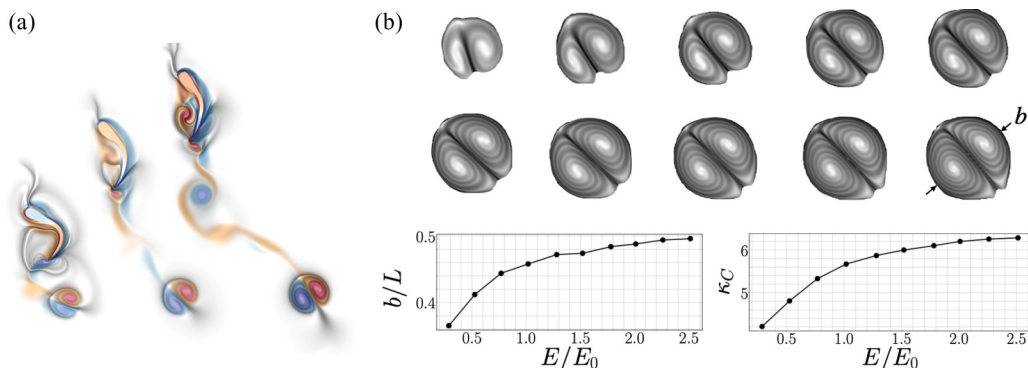


FIG. 7. Characterization of Lagrangian coherent structures at different energy budgets. Left: Vorticity fields overlaid onto the finite time Lyapunov exponent field (FTLE) of three different escapes. From left to right are the C start, the artificial swimmer with energy budget  $E_0$ , and the artificial swimmer with energy budget  $3E_0$ . The FTLE is visualized on a spectrum from white to black. The darker colors represent a higher value of the FTLE. The LCS are the “ridges” of the FTLE field, i.e., the darkest contours on the visualization. All FTLE fields are computed with the same integration length and displayed at the same time instant (see Appendix C). (a) A visualization of the lump of fluid ejected by the swimmer, the average vortex diameter  $b$ , and the peak C-bend midpoint curvature  $\kappa_C$  for escapes with energy budgets linearly spaced between  $\frac{1}{3}E_0$  and  $3E_0$ , shown to scale. The lumps of fluid are extracted by localizing the high ridge (LCS) values in the respective FTLE field. The average vortex diameter  $b$  is normalized by the swimmer length  $L$  as a function of the energy budget.

its body more than the C-start swimmer during the initial C-bend motion. As a result, a higher peak power and maximum speed are attained [Figs. 6(a) and 6(b)]. Moreover, the RL swimmer achieves a 13.5g peak acceleration, significantly greater than the 9.4g peak acceleration created by the lower thrust C-start swimmer.

Furthermore, the RL swimmer produces close to zero thrust after the initial C bend while the C-start swimmer continually produces thrust and expends power during the escape as a result of the continuous swimming pattern [Fig. 6(b)]. Thus, the learned escape pattern uses more energy during the initial propulsion but saves energy by coasting in the remaining time. Why does this result in increased escape distance using the same energy? When plotting the drag coefficient  $C_d$  against time for the two escape patterns we observe that the drag during the coasting phase of the escape is close to zero for the RL swimmer while the C-start swimmer continually experiences elevated drag due to the swimming motion. This analysis suggests that expending most energy in a strong initial C bend and then coasting is advantageous due to the decrease of fluid dynamic drag when compared to swimming.

Finally, the hydrodynamic mechanisms exploited by the artificial swimmer were analyzed using Lagrangian coherent structures (LCSs) [56]. Well defined LCSs are characterized by negligible flux across their surface, acting as transport barriers within the flow [57]. In Fig. 7(a) the resulting LCSs are superimposed onto the vorticity fields for the C start and two RL escapes of different energy. The LCSs evidence one predominant coherent structure: A lump of fluid carried by the counter-rotating vortex pair. This has previously been observed in [58], and is a common structure found in fish escape sequences.

Furthermore, we notice that the lump of fluid ejected by the C-start escape motion [Fig. 7(a), left] is less symmetric, and smaller than that of the RL escape [Fig. 7(a), middle]. Thus, the RL swimmer uses more of the allocated energy budget to trap and accelerate a large volume of water opposite to the escape direction, while the C-start swimmer uses less energy to propel the initial ball of water, but compensates by using the leftover energy to swim continuously for the rest of the escape. Moreover in the case of RL, the vortex dipole and its lump of fluid are more aligned with the direction of the motion than what is observed in C start. This continuous alignment minimizes



the time derivative of the vorticity linear impulse and as such the drag experienced by the swimmer [59]. Finally, we found that as the energy budget increased, the average diameter of the lump of fluid ejected by the swimmer began to flatten off close to  $b \approx \frac{1}{2}L$  [see Fig. 7(b)]. Since the peak curvature of the C bend cannot increase beyond  $2\pi$  ( $\kappa_C \leq 2\pi$ ), the swimmer cannot convert higher energy into more escape distance indefinitely by forming more pronounced C bends [Fig. 5(b)], thus limiting the maximum escape distance attainable at a given Reynolds number and morphology.

## VIII. CONCLUSIONS

This study explores the use of deep reinforcement learning to discover swimming escape patterns which maximize distance given a fixed amount of energy. In contrast to an optimization process with an overarching goal, RL explores an array of incremental processes allowing it to learn a range of escape patterns. Our results indicate that maximum swimming distance can be achieved through short bursts of accelerating motion interlinked by phases of powerless gliding. In the context of fish escaping from disturbances, we find that, at higher Reynolds numbers, burst-coast escape patterns result in greater escape distances than burst-swim escape patterns (C starts), but C starts propel the swimmer away from the initial position faster. This suggests that larval zebrafish performing C starts may not only be aiming to maximize escape distance, but their internal reward function may further include a notion of urgency to distance themselves from their initial position as quickly as possible. Future studies may benefit from encoding urgency into the reward function and observing how the resulting swimming escape patterns change.

Contrary to an optimization setting which requires using domain-specific knowledge to pre-define stages of motion, the reinforcement learning setting is free from prior bias on the functional form of the escape pattern. This additional freedom results in escape patterns that outperform those obtained by optimization, for the same energy budget. Moreover, we find that training the swimmer to control its motions as a function of the energy budget produces a “kaleidoscope” of escape patterns that reveal practical flow optimization principles for efficient swimming. Studying the learned strategies indicates that the formation of a C shape, a coasting phase, and a final swim motion are necessary components of distance maximizing escapes across energy budgets. Other strategies, such as slight undulations during the coasting phase, are eliminated as the energy budget decreases, indicating their second-order, but non-negligible, importance for achieving rapid propulsion from rest. This suggests that reinforcement learning can more robustly discover swimming escape patterns than methods based on reverse engineering via optimization.

Finally, we emphasize that RL is a data efficient learning methodology. In the current problem setup, each action is determined by nine real valued parameters, and each escape consists of around seven actions. If this problem were to be solved with a stochastic optimizer, this amounts to solving a  $n = 63$  dimensional constrained optimization problem. Stochastic optimizers like CMA-ES find global optima for a variety of functions using  $300n-500n^2$  function evaluations [60]. Thus, the computational cost can be anywhere between 18 900 and 1 984 500 simulations. Furthermore, the RL policy can be controlled according to a real-valued energy budget, producing a wide array of swimming escapes. Approximating this level of fine-grained control with an optimization approach would require repeating the optimization for many different energy budgets (e.g.,  $N$  times)—further increasing the computational cost to anywhere between 18 900 $N$  and 1 984 500 $N$  simulations. Compounded with the fact that the optimal number of actions to be taken given the energy and time constraints is *not known in advance* and should be learned (as done by reinforcement learning) renders the problem totally impractical if approached with optimization-based reverse engineering. On the contrary, we find that solving the full problem via RL is tractable using only around 150 000 simulations.

In summary we find that RL is a powerful tool for the discovery of swimming escape patterns under energy constraints. The identified escape patterns deepen our understanding of the hydrodynamic mechanisms that are exploited by natural swimmers.

**APPENDIX A: NUMERICAL METHOD**

We use a simplified two-dimensional (2D) geometric model of a 5-day-postfertilization zebrafish [36]. The swimmer shape is described by the body half-width  $w(s) \in \mathbb{R}$  along the curvilinear coordinate  $s \in [0, L]$  for a body length  $L \in \mathbb{R}$ ,

$$w(s) = \begin{cases} w_h \sqrt{1 - \left(\frac{s_b - s}{s_b}\right)^2} & 0 \leq s < s_b \\ (-2\Delta w - w_t \Delta s) \delta s_b^3 \\ + (3\Delta w + w_t \Delta s) \delta s_b^2 & s_b \leq s < s_t \\ + w_h & \\ w_t - w_t \left(\frac{s - s_t}{L - s_t}\right)^2 & s_t \leq s < L \end{cases} \quad (\text{A1})$$

Above,  $\Delta w = w_t - w_h$ ,  $\Delta s = s_t - s_b$ , and  $\delta s_b = \frac{s - s_b}{\Delta s}$ , where  $s_b = 0.0862L$ ,  $s_t = 0.3448L$ ,  $w_h = 0.0635L$ ,  $w_t = 0.0254L$ . In order to resolve the head and the tail of the swimmer, the spacing along the midline in the first and last 10% of the body is linearly increased from  $\Delta x/8$  to  $\Delta x/\sqrt{2}$ , where  $\Delta x$  denotes the uniform resolution of the computational grid.

The flow field generated by the motion of the artificial swimmer is simulated by solving the two-dimensional, incompressible Navier-Stokes equations with volume penalisation [61–63]:

$$\nabla \cdot \mathbf{u} = 0, \quad \frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{\nabla p}{\rho} + \nu \nabla^2 \mathbf{u} + \lambda \chi(\mathbf{u}_s - \mathbf{u}). \quad (\text{A2})$$

Here,  $\mathbf{u}(x, t) \in \mathbb{R}^2$  corresponds to the fluid velocity and  $p(x, t) \in \mathbb{R}$  to the pressure. The fluid properties are determined by the viscosity  $\nu \in \mathbb{R}$  and the fluid density  $\rho \in \mathbb{R}$ . The fluid-structure interaction is modeled by the penalty term  $\lambda \chi(\mathbf{u}_s - \mathbf{u})$ , where  $\mathbf{u}_s \in \mathbb{R}^2$  denotes the combined translational, rotational, and deformation velocity of the swimmer. The characteristic function  $\chi(x, t) \in \mathbb{R}$  is 1 inside the swimmer, 0 elsewhere. The equation is solved in a two-dimensional domain  $\mathbf{x} \in \Omega \subset \mathbb{R}^2$ , over a time interval  $t \in [0, T] \subset \mathbb{R}$ . The domain was chosen to be four times the length of the swimmer  $\Omega = [4L, 4L] \subseteq \mathbb{R}^2$  and we ran the solver up to time  $T_{\max} = 118.8$  ms.

In order to solve Eq. (A2) we discretized the equation on a uniform grid in space. On this grid, the spatial derivatives were approximated using second-order centered finite differences. For this purpose we used the uniform grid library Cubism [64]. The time stepping is performed using explicit Euler, where the time step  $\Delta t$  was adopted such that the CFL number was constrained to 0.1. To ensure momentum conservation and stability the penalty parameter is set to  $\lambda = 1/\Delta t$  [62]. For the characteristic function we use a second order approximation of the Heaviside function [65]. During the reinforcement learning, we use  $512 \times 512$  grid points. All quantities of interest are subsequently computed at the higher resolution of  $1024 \times 1024$  grid points.

In the following we describe the operator splitting formalism used to compute the time step. Starting with the velocity field  $\mathbf{u}^t$  at time step  $t$  we computed an intermediate velocity  $\mathbf{u}^*$  by performing advection and diffusion,

$$\mathbf{u}^* = \mathbf{u}^t + \Delta t(\nu \nabla^2 \mathbf{u} - \mathbf{u} \cdot \nabla \mathbf{u}). \quad (\text{A3})$$

The resulting velocity field is nondivergence free and we used pressure projection [66]

$$\mathbf{u}^{**} = \mathbf{u}^* - \Delta t \frac{\nabla p^{t+1}}{\rho}. \quad (\text{A4})$$

The pressure field was computed by solving the Poisson equation that results when taking the divergence of Eq. (A4) and using that the obstacle velocities can be nondivergence free due to

the deformation component  $\nabla \cdot \mathbf{u}^{**} = \chi \nabla \cdot \mathbf{u}_s^{t+1}$ ,

$$\Delta p^{t+1} = \frac{\rho}{\Delta t} \left[ \nabla \cdot \mathbf{u}^* - \sum_{s=1}^{N_s} \chi \nabla \cdot \mathbf{u}_s^{t+1} \right]. \quad (\text{A5})$$

We conclude the time step by employing the penalization force on the field,

$$\mathbf{u}^{t+1} = \mathbf{u}^{**} + \chi (\mathbf{u}_s^{t+1} - \mathbf{u}^{**}), \quad (\text{A6})$$

where we used  $\lambda = 1/\Delta t$ .

Using the numerical solution of the Navier-Stokes equation we can compute the work done by the deformation of the swimmer body  $\mathbf{u}_{\text{def}}$  on the surrounding flow, which can be thought of as the muscle input power, as

$$E(t) = \int_0^t \left[ \int_{\partial \Sigma} \mathbf{u}_{\text{def}} \cdot d\mathbf{F} \right] dt. \quad (\text{A7})$$

Note that the computational model used does not account for muscle dynamics so the muscle input power cannot directly be computed, only approximated.

The energy values reported are nondimensionalized by  $ML^2/T_{\text{prop}}^2$ , where  $M$  is the swimmer mass,  $L$  is the swimmer length, and  $T_{\text{prop}}$  is the propulsive swimming period. The deformation power follows from Eq. (A7), computed as  $P_{\text{def}} = \frac{dE(t)}{dt}$ , and is nondimensionalized by  $ML^2/T_{\text{prop}}^3$ . The propulsive thrust  $F_t$  and drag force  $F_d$  are computed as displayed in Eqs. (A8) and (A9) are nondimensionalized by  $ML^2/T_{\text{prop}}^2$ ,

$$F_t = \int_{\partial \Sigma} (\mathbf{u} \cdot d\mathbf{F} + |\mathbf{u} \cdot d\mathbf{F}|)/(2|\mathbf{u}|), \quad (\text{A8})$$

$$F_d = \int_{\partial \Sigma} (\mathbf{u} \cdot d\mathbf{F} - |\mathbf{u} \cdot d\mathbf{F}|)/(2|\mathbf{u}|). \quad (\text{A9})$$

In Eqs. (A7)–(A9),  $\partial \Sigma$  denotes the swimmer surface, and  $d\mathbf{F}$  is the force acting on the swimmer comprised of viscous and pressure-based forces  $d\mathbf{F} = d\mathbf{F}_p + d\mathbf{F}_v = 2\mu \mathbf{D}ndS - PndS$ . Here,  $\mathbf{D} = \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^T)$  is the strain-rate tensor,  $P$  is the surface pressure,  $\mu$  is the dynamic viscosity,  $\mathbf{n}$  is the surface normal, and  $dS$  is the infinitesimal surface element.

## APPENDIX B: REINFORCEMENT LEARNING

During the escape, the swimmer senses its cylindrical coordinates  $(d, \phi) \in \mathbb{R}^2$ , its center-of-mass velocity  $\mathbf{v} \in \mathbb{R}^2$ , as well as its orientation and angular velocity  $\theta, \dot{\theta} \in \mathbb{R}$  relative to the fixed laboratory frame. Furthermore, it has access to the remaining energy available for the escape  $E_{\text{to-go}} \in \mathbb{R}$ , and a memory of its action from two previous time steps  $\mathbf{a}_t, \mathbf{a}_{t-1}$ . These perceptive abilities form the state vector  $\mathbf{s} = (d, \phi, \mathbf{v}, \theta, \dot{\theta}, E_{\text{to-go}}, \mathbf{a}_t, \mathbf{a}_{t-1}) \in \mathbb{R}^{25}$ . The energy available for the escape,  $E_{\text{to-go}} \in \mathbb{R}$ , is computed as the difference between the available energy  $E_{\text{budget}}$  and the work done by the swimmer on the surrounding flow.

Given each current state, the swimmer is able to influence its mid-line configuration by scheduling changes in curvature. In particular it can select the amount of monolateral (baseline) curvature  $B(s, t) \in \mathbb{R}$ , undulatory curvature  $K(s, t) \in \mathbb{R}$ , the traveling wave phase  $\tau_L \in \mathbb{R}$ , the overall phase  $\phi \in \mathbb{R}$ , and the duration of each transition  $\Delta t \in \mathbb{R}$ . Since the baseline and undulatory curvature are each parametrized by six control points, of which three are free, the actions are given as  $\mathbf{a} = (B_1, B_2, B_3, K_1, K_2, K_3, \tau_L, \phi, \Delta t) \in \mathbb{R}^9$ . Here,  $B_i, K_i \in \mathbb{R}$  for  $i = 1, 2, 3$  denote the three controllable baseline and three controllable undulatory curvatures on the swimmer midline. As per experimental observations of zebrafish performing fast starts [49], we constrain the maximum curvature of the artificial swimmer to  $|\kappa(s, t)| \leq 2\pi/L$  and the action duration to  $\Delta t \in [0.5T_{\text{prop}}, T_{\text{prop}}]$ . The phases are constrained to  $\tau_L, \phi_L \in [0, 2\pi]$ . The transition between actions takes place in time by cubic interpolation with derivatives equal on both sides of extrema to ensure continuity.

The swimmer starts in a zero curvature configuration at the center of a square simulation domain and is assigned a random energy budget  $E_{\text{budget}}$  sampled uniformly between  $\frac{1}{3}E_0$  and  $3E_0$ . The episode terminates if the swimmer has depleted its energy budget, the allocated time is surpassed, or spatial constraints are violated. The maximum time for the escape is set to  $T_{\text{max}} = 118$  ms, equal to the time length of the C-start escape [36]. Furthermore, the episode is terminated and the swimmer is penalized with  $r = -10$  if the swimmer changes orientation by  $|\Delta\theta| \geq \pi/2$ . Since a typical zebrafish only changes its orientation by approximately  $44^\circ$  during a fast start [49], this constraint restricts the exploration space, but is sufficiently relaxed to not significantly influence the escape pattern. To model the scenario of a predator approaching from behind the zebrafish, we enforce that the swimmer center-of-mass remains in an infinite triangular region starting  $0.2L$  behind the swimmer tail end point and with  $40^\circ$  aperture. If the swimmer exits the allowable region the episode is terminated and the swimmer is penalized with  $r = -10$ .

The off-policy actor-critic reinforcement learning algorithm V-RACER [50] implemented in smarties [67] is employed for 1 000 000 state-action-reward observations. The neural network used to approximate the policy network and state value function has three hidden layers of 32 parameters each. A discount factor of  $\gamma = 0.99$  is used, the batch size is set to 128, the exploration noise probability is set to 0.2, and the learning rate for stochastic gradient descent is set to 0.0001. The other hyperparameters are left as described in the original publication.

### APPENDIX C: LAGRANGIAN COHERENT STRUCTURES

The Lagrangian coherent structures (LCSs) are the ridges of the finite time Lyapunov exponent field (FTLE). The FTLE is a scalar field  $\sigma(\mathbf{X}_0)$  which characterizes the amount of stretching about the trajectory of a passive flow particle, from the location  $\mathbf{X}_0$  to the location  $\mathbf{X}$  during time  $T$  [68]. The trajectory  $\mathbf{X}(t)$  of a particle located at position  $\mathbf{X}_0 = \mathbf{X}(t_0)$  at time  $t_0$  can be obtained by integrating

$$\frac{d}{dt}\mathbf{X}(t) = \mathbf{u}[\mathbf{X}(t), t], \quad (\text{C1})$$

where  $\mathbf{u}$  is the flow velocity field. By following the particle trajectories for a time length  $T$ , we obtain the particle flow map which gives the particle position at later times  $\mathbf{X}_0 \rightarrow \mathbf{X}(t_0 + T) = \phi(\mathbf{X}_0, T)$ . From the flow map we can obtain the (right) Cauchy-Green deformation tensor  $\Delta$ , which quantifies the stretching of an infinitesimal material line,

$$\Delta(\mathbf{X}_0) = \left[ \frac{\partial \phi(\mathbf{X}_0, T)}{\partial \mathbf{X}_0} \right]^\top \frac{\partial \phi(\mathbf{X}_0, T)}{\partial \mathbf{X}_0}. \quad (\text{C2})$$

The finite-time Lyapunov exponent (FTLE) is then defined as the square root of the logarithm of the maximum eigenvalue  $\lambda_{\text{max}}$  of the Cauchy-Green deformation tensor  $\Delta$  normalized by the integration time  $T$ ,

$$\sigma(\mathbf{X}_0) = \frac{1}{T} \sqrt{\ln\{\lambda_{\text{max}}[\Delta(\mathbf{X}_0)]\}}. \quad (\text{C3})$$

The FTLE fields reported in this study are calculated by integrating in forward time using the open source software FTLE2D [68]. We use a set of 158 velocity fields of the 118.8-ms escape, spaced at equal time intervals. The FTLE is computed for the first 100 velocity fields, thus the receding integration time horizon is 58 time steps. The time horizon is chosen to be sufficiently long in order to adequately locate the LCS on the FTLE fields.

---

[1] M. S. Triantafyllou, Survival hydrodynamics, *J. Fluid Mech.* **698**, 1 (2012).

[2] J. Peng and J. O. Dabiri, Transport of inertial particles by Lagrangian coherent structures: Application to predator-prey interaction in jellyfish feeding, *J. Fluid Mech.* **623**, 75 (2009).

- [3] A. Nair, K. Changsing, W. J. Stewart, and M. J. McHenry, Fish prey change strategy with the direction of a threat, *Proc. R. Soc. B: Biological Sci.* **284**, 20170393 (2017).
- [4] M. J. McHenry, J. L. Johansen, A. P. Soto, B. A. Free, D. A. Paley, and J. C. Liao, The pursuit strategy of predatory bluefish (*Pomatomus saltatrix*), *Proc. Biological Sci.* **286**, 20182934 (2019).
- [5] M. A. Borla, B. Palecek, S. Budick, and D. M. O'Malley, Prey capture by larval zebrafish: Evidence for fine axial motor control, *Brain, Behav. Evol.* **60**, 207 (2002).
- [6] A. Soto, W. J. Stewart, and M. J. McHenry, When optimal strategy matters to prey fish, *Integr. Comp. Biol.* **55**, 110 (2015).
- [7] S. A. Budick and D. M. O'Malley, Locomotor repertoire of the larval zebrafish: Swimming, turning and prey capture, *J. Exp. Biol.* **203**, 2565 (2000).
- [8] S. P. Colin, J. H. Costello, L. J. Hansson, J. Titelman, and J. O. Dabiri, Stealth predation and the predatory success of the invasive ctenophore *Mnemiopsis leidyi*, *Proc. Natl. Acad. Sci. USA* **107**, 17223 (2010).
- [9] J. Gray, Directional control of fish movement, *Proc. R. Soc. London. Series B* **113**, 115 (1933).
- [10] P. W. Webb, Acceleration performance of rainbow trout *salmo gairdneri* and green sunfish *lepomis cyanellus*, *J. Exp. Biol.* **63**, 451 (1975).
- [11] D. Weihs, The mechanism of rapid starting of slender fish, *Biorheology* **10**, 343 (1973).
- [12] P. Domenici and R. Blake, The kinematics and performance of fish fast-start swimming, *J. Exp. Biol.* **200**, 1165 (1997).
- [13] J. H. Bollmann, The zebrafish visual system: From circuits to behavior, *Ann. Rev. Vision Sci.* **5**, 269 (2019).
- [14] A. Carrillo, D. V. Le, M. Byron, H. Jiang, and M. J. McHenry, Canal neuromasts enhance foraging in zebrafish (*Danio rerio*), *Bioinspiration Biomimetics* **14**, 035003 (2019).
- [15] W. J. Stewart, A. Nair, H. Jiang, and M. J. McHenry, Prey fish escape by sensing the bow wave of a predator, *J. Exp. Biol.* **217**, 4328 (2014).
- [16] I. Bianco and F. Engert, Visuomotor transformations underlying hunting behavior in zebrafish, *Curr. Biol.* **25**, 831 (2015).
- [17] I. H. Bianco, A. R. Kampff, and F. Engert, English, Prey capture behavior evoked by simple visual stimuli in larval zebrafish, *Front. Syst. Neurosci.* **5**, 1 (2011).
- [18] T. W. Dunn, Y. Mu, S. Narayan, O. Randlett, E. A. Naumann, C.-T. Yang, A. F. Schier, J. Freeman, F. Engert, and M. B. Ahrens, Brain-wide mapping of neural activity controlling zebrafish exploratory locomotion, *eLife* **5**, e12741 (2016).
- [19] T. Dunn, C. Gebhardt, E. Naumann, C. Riegler, M. Ahrens, F. Engert, and F. Del Bene, Neural circuits underlying visually evoked escapes in larval zebrafish, *Neuron* **89**, 613 (2016).
- [20] B. C. Jayne and G. V. Lauder, Red and white muscle activity and kinematics of the escape response of the bluegill sunfish during swimming, *J. Comp. Phys. A* **173**, 495 (1993).
- [21] M. A. B. Schwalbe, A. L. Boden, T. N. Wise, and E. D. Tytell, Red muscle activity in bluegill sunfish *Lepomis macrochirus* during forward accelerations, *Sci. Rep.* **9**, 8088 (2019).
- [22] W. C. Witt, L. Wen, and G. V. Lauder, Hydrodynamics of C-start escape responses of fish as studied with simple physical models, *Integr. Comp. Biol.* **55**, 728 (2015).
- [23] B. P. Epps and A. H. Techet, Impulse generated during unsteady maneuvering of swimming fish, *Exp. Fluids* **43**, 691 (2007).
- [24] E. D. Tytell and G. V. Lauder, Hydrodynamics of the escape response in bluegill sunfish, *Lepomis macrochirus*, *J. Exp. Biol.* **211**, 3359 (2008).
- [25] G. Li, U. K. Müller, J. L. van Leeuwen, and H. Liu, Escape trajectories are deflected when fish larvae intercept their own C-start wake, *J. R. Soc., Interface* **11**, 20140848 (2014).
- [26] I. Borazjani, The functional role of caudal and anal/dorsal fins during the C-start of a bluegill sunfish, *J. Exp. Biol.* **216**, 1658 (2013).
- [27] I. Borazjani, F. Sotiropoulos, E. D. Tytell, and G. V. Lauder, Hydrodynamics of the bluegill sunfish C-start escape response: Three-dimensional simulations and comparison with experimental data, *J. Exp. Biol.* **215**, 671 (2012).
- [28] D. L. Kramer and R. L. Mclaughlin, The Behavioral Ecology of Intermittent Locomotion, *American Zoologist* **41**, 137 (2001).

- [29] B. A. Chadwell, E. M. Standen, G. V. Lauder, and M. A. Ashley-Ross, Median fin function during the escape response of bluegill sunfish (*Lepomis macrochirus*). I: Fin-ray orientation and movement, *J. Exp. Biol.* **215**, 2869 (2012).
- [30] D. Weihs, Energetic advantages of burst swimming of fish, *J. Theor. Biol.* **48**, 215 (1974).
- [31] D. Weihs, Energetic significance of changes in swimming modes during growth of larval anchovy, *Engraulis mordax*, *Natl. Mar. Fish. Serv. Fish. Bull.* **77**, 507 (1980).
- [32] J. J. Videler and D. Weihs, en Energetic advantages of burst-and-coast swimming of fish at high speeds, *J. Exp. Biol.* **97**, 169 (1982).
- [33] G. Wu, Y. Yang, and L. Zeng, Kinematics, hydrodynamics and energetic advantages of burst-and-coast swimming of koi carps (*Cyprinus carpio koi*), *J. Exp. Biol.* **210**, 2181 (2007).
- [34] S. Kern and P. Koumoutsakos, Simulations of optimized anguilliform swimming, *J. Exp. Biol.* **209**, 4841 (2006).
- [35] S. Kern, P. Chatelain, and P. Koumoutsakos, Modeling, simulation and optimization of anguilliform swimmers, *Bio-mechanisms of Swimming and Flying: Fluid Dynamics, Biomimetic Robots, and Sports Science* (2008), pp. 167–178.
- [36] M. Gazzola, W. M. V. Rees, and P. Koumoutsakos, C-start: Optimal start of larval fish, *J. Fluid Mech.* **698**, 5 (2012).
- [37] M. Gazzola, B. Hejazialhosseini, and P. Koumoutsakos, Reinforcement learning and wavelet adapted vortex methods for simulations of self-propelled swimmers, *SIAM J. Sci. Comput.* **36**, B622 (2014).
- [38] G. Reddy, A. Celani, T. J. Sejnowski, and M. Vergassola, Learning to soar in turbulent environments, *Proc. Natl. Acad. Sci. USA* **113**, E4877 (2016).
- [39] M. Gazzola, A. A. Tchieu, D. Alexeev, A. d. Brauer, and P. Koumoutsakos, Learning to school in the presence of hydrodynamic interactions, *J. Fluid Mech.* **789**, 726 (2016).
- [40] G. Novati, S. Verma, D. Alexeev, D. Rossinelli, W. M. v. Rees, and P. Koumoutsakos, Synchronisation through learning for two self-propelled swimmers, *Bioinspiration Biomimetics* **12**, 036001 (2017).
- [41] S. Colabrese, K. Gustavsson, A. Celani, and L. Biferale, Flow Navigation by Smart Microswimmers via Reinforcement Learning, *Phys. Rev. Lett.* **118**, 158004 (2017).
- [42] S. Verma, G. Novati, and P. Koumoutsakos, Efficient collective swimming by harnessing vortices through deep reinforcement learning, *Proc. Natl. Acad. Sci. USA* **115**, 5849 (2018).
- [43] G. Reddy, J. Wong-Ng, A. Celani, T. J. Sejnowski, and M. Vergassola, Glider soaring via reinforcement learning in the field, *Nature (London)* **562**, 236 (2018).
- [44] G. Novati, L. Mahadevan, and P. Koumoutsakos, Controlled gliding and perching through deep-reinforcement-learning, *Phys. Rev. Fluids* **4**, 093902 (2019).
- [45] F. Guéniat, L. Mathelin, and M. Y. Hussaini, A statistical learning strategy for closed-loop control of fluid flows, *Theor. Comput. Fluid Dyn.* **30**, 497 (2016).
- [46] J. Rabault, M. Kuchta, A. Jensen, U. Réglade, and N. Cerardi, Artificial neural networks trained through deep reinforcement learning discover control strategies for active flow control, *J. Fluid Mech.* **865**, 281 (2019).
- [47] S. L. Brunton, B. R. Noack, and P. Koumoutsakos, Machine learning for fluid mechanics, *Annu. Rev. Fluid Mech.* **52**, 477 (2020).
- [48] D. Fan, L. Yang, Z. Wang, M. S. Triantafyllou, and G. E. Karniadakis, Reinforcement learning for bluff body active flow control in experiments and simulations, *Proc. Natl. Acad. Sci. USA* **117**, 26091 (2020).
- [49] U. K. Müller, J. G. M. v. d. Boogaart, and J. L. v. Leeuwen, Flow patterns of larval fish: Undulatory swimming in the intermediate flow regime, *J. Exp. Biol.* **211**, 196 (2008).
- [50] G. Novati and P. Koumoutsakos, Remember and Forget for Experience Replay, in *International Conference on Machine Learning* (PMLR, Long Beach, California, 2019), pp. 4851–4860.
- [51] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevFluids.6.093101> for Video 1: An example of the learned escape pattern when the swimmer is endowed with energy budget  $E_0$  and Video 2: An example of the learned escape pattern when the swimmer is endowed with energy budget  $3E_0$ .
- [52] M. J. Lighthill, Large-amplitude elongated-body theory of fish locomotion, *Proc. R. Soc. London, Ser. B* **179**, 125 (1971).



- [53] P. W. Webb, The swimming energetics of trout: I. Thrust and power output at cruising speeds, *J. Exp. Biol.* **55**, 489 (1971).
- [54] C. J. Voesenek, F. T. Muijres, and J. L. V. Leeuwen, Biomechanics of swimming in developing larval fish, *J. Exp. Biol.* **221** (2018).
- [55] G. Li, I. Ashraf, B. François, D. Kolomenskiy, F. Lechenault, R. Godoy-Diana, and B. Thiria, Burst-and-coast swimmers optimize gait by adapting unique intrinsic cycle, *Commun. Biol.* **4**, 40 (2021).
- [56] G. Haller and G. Yuan, Lagrangian coherent structures and mixing in two-dimensional turbulence, *Phys. D (Amsterdam, Neth.)* **147**, 352 (2000).
- [57] S. C. Shadden, F. Lekien, and J. E. Marsden, Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows, *Phys. D (Amsterdam, Neth.)* **212**, 271 (2005).
- [58] F. Huhn, W. M. van Rees, M. Gazzola, D. Rossinelli, G. Haller, and P. Koumoutsakos, Quantitative flow analysis of swimming dynamics with coherent Lagrangian vortices, *Chaos* **25**, 087405 (2015).
- [59] P. Koumoutsakos and A. Leonard, High-resolution simulations of the flow around an impulsively started cylinder using vortex methods, *J. Fluid Mech.* **296**, 1 (1995).
- [60] N. Hansen and S. Kern, Evaluating the CMA evolution strategy on multimodal test functions, *Parallel Problem Solving from Nature -PPSN VIII, Lecture Notes in Computer Science (Springer, Berlin, Heidelberg, 2004)*, Vol. 3242.
- [61] P. Angot, C. H. Bruneau, and P. Fabrie, A penalization method to take into account obstacles in incompressible viscous flows, *Numer. Math.* **81**, 497 (1999).
- [62] M. Coquerelle and G.-H. Cottet, A vortex level set method for the two-way coupling of an incompressible fluid with colliding rigid bodies, *J. Comput. Phys.* **227**, 9121 (2008).
- [63] M. Gazzola, P. Chatelain, W. M. van Rees, and P. Koumoutsakos, Simulations of single and multiple swimmers with non-divergence free deforming geometries, *J. Comput. Phys.* **230**, 7093 (2011).
- [64] D. Rossinelli, B. Hejazialhosseini, P. Hadjidoukas, C. Bekas, A. Curioni, A. Bertsch, S. Futral, S. J. Schmidt, N. A. Adams, and P. Koumoutsakos, 11 PFLOP/s simulations of cloud cavitation collapse, in *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*, SC '13 (Association for Computing Machinery, Denver, CO, 2013), pp. 1–13.
- [65] J. D. Towers, Finite difference methods for approximating Heaviside functions, *J. Comput. Phys.* **228**, 3478 (2009).
- [66] A. J. Chorin, Numerical solution of the navier-stokes equations, *Math. Comput.* **22**, 745 (1968).
- [67] <https://github.com/cselab/smarties>.
- [68] C. Conti, D. Rossinelli, and P. Koumoutsakos, GPU and APU computations of Finite Time Lyapunov Exponent fields, *J. Comput. Phys.* **231**, 2229 (2012).