Efficient navigation of cargo-towing microswimmer in non-uniform flow fields

Krongtum Sankaewtong,* John J. Molina, and Ryoichi Yamamoto†
Department of Chemical Engineering, Kyoto University, Kyoto 615-8510, Japan

(Received 17 April 2024; accepted 23 August 2024; published 16 September 2024)

The vision of deploying miniature vehicles within the human body for intricate tasks holds tremendous promise across engineering and medical domains. Herein, optimal navigation of a cargo-towing swimmer under an applied zig-zag flow is studied by employing direct numerical simulations coupled with a deep reinforcement learning algorithm. Tasks include navigation in flow and shear-gradient directions. We initially explore combinations of state inputs, finding that optimal navigation necessitates swimmers to perceive hydrodynamics and alignment, surpassing reliance solely on hydrodynamic signals while considering their memories. Next, we study combinations of action spaces, allowing dynamic changes in swimming and/or rotational velocities by tuning B_1 and C_1 parameters of the squirmer model, respectively. By keeping both parameters fixed, cargo-towing swimmers demonstrate superior performance in the flow direction compared to swimmers without load due to tumbling movements influenced by shear flow. In the shear-gradient direction, swimmers without load outperform cargo-towing swimmers, with performance decreasing as load length increases. Across the combination of allowing B_1 and C_1 to change, the policies from solely dynamic B_1 actions demonstrate superior navigation. The policies are then used as a showcase against naive cargo-towing and inert colloidal chains. A t-distributed stochastic neighbor embedding analysis reveals the complex interplay between perceived hydrodynamic signals and swimmer position. In the flow direction, swimmers align effectively with regions of maximum velocity, while in the shear-gradient direction, periodic transitions from minimum to maximum state values occur. Comparing pullers, pushers, and neutral swimmers, cargo-towing swimmers show a reversal in swimming velocity trends, with pullers outpacing neutral and pusher swimmers, irrespective of load lengths.

DOI: 10.1103/PhysRevResearch.6.033305

I. INTRODUCTION

The concept of operating miniature vehicles within the human body to perform intricate tasks has been a long-standing dream that engineers today are poised to realize. In this scenario, robotic surgeons could intricately navigate the human body, targeting ailments such as arterial plaque and Alzheimer's-related protein deposits, while nanomachines could penetrate resilient materials like steel beams or airplane wings, detecting and repairing cracks to avert catastrophic failures [1]. This visionary concept holds the promise of unprecedented precision and safety across engineering and medical domains. Micro-/nanoscale swimmers, encompassing both natural and artificial entities, represent a class of objects capable of autonomous mobility by harnessing energy from their surroundings [2]. Over recent decades, these swimmers have emerged as transformative agents for navigating complex microenvironments, spanning biomedical applications such as drug delivery and gene therapy [3,4] to environmental remediation efforts [5,6]. These targeted applications require the microswimmers to carry a specific load, i.e., cargoes, towards the designated sites [7–9]. However, traditional navigation methods often face formidable challenges in traversing intricate landscapes filled with obstacles and uncertainties due to the need for extensive prior knowledge of the environment and prohibitive computational costs [10–12].

In nature, organisms, ranging from animals to humans, exhibit an inherent ability to navigate unfamiliar terrains. Much like adults effortlessly maneuvering through urban landscapes or marine zooplankton evading predators in open seas by leveraging local cues [13], these biological responses stem from a unified representation of perceptions that support memory and guide future actions—a phenomenon often referred to as a cognitive map [14]. The recent advancements in machine learning techniques, particularly reinforcement learning (RL), offer promising avenues to understand and emulate the adaptability and navigation capabilities observed in nature. Many studies have adopted these technique to study the efficient navigation of microswimmers in various types of environments [15–20], by assuming that the swimmers have privileged access to laboratory-frame information. Some further investigations have considered scenarios with partial information or restricted sensory capabilities, assuming swimmers can only detect signals in their immediate vicinity [21–24]. Recently, researchers have studied the dynamics of

*Contact author: aom@cheme.kyoto-u.ac.jp
†Contact author: ryoichi@cheme.kyoto-u.ac.jp

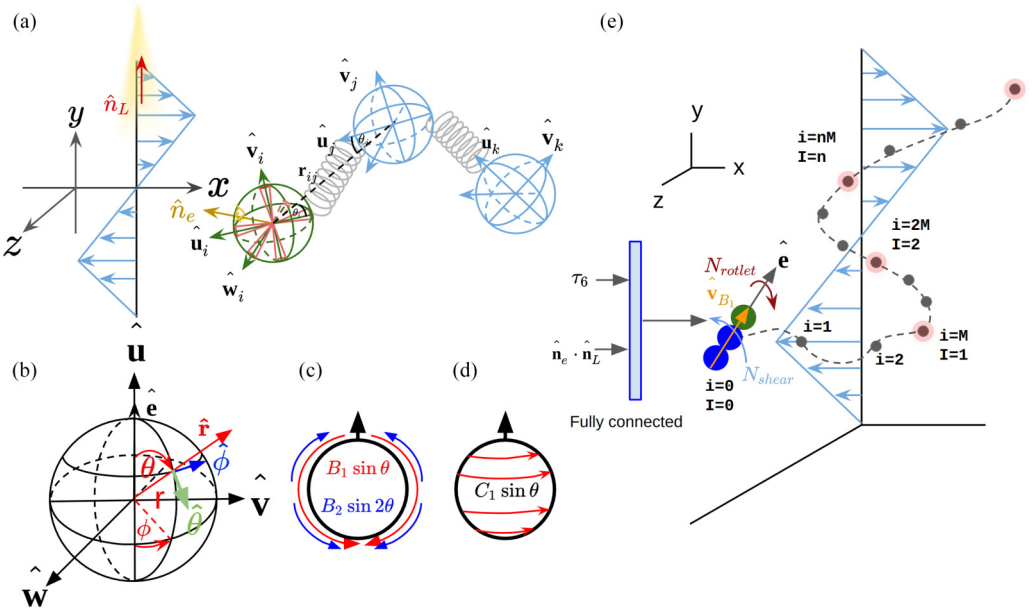

FIG. 1. (a) Depiction of a swimmer carrying a load of length $d_{\text{cargo}} = 2$ navigating a zig-zag shear flow. The principal axes of the ith bead are $\hat{\boldsymbol{u}}_i$, $\hat{\boldsymbol{v}}_i$, and $\hat{\boldsymbol{w}}_i$. The initial bead is a squirmer, while the subsequent beads are inert and distinguished by green and blue spheres. The first bead features an eye positioned at 30 deg from $\hat{\boldsymbol{u}}$, enhancing the chain's ability to detect a visual cue for a light source in $\hat{\boldsymbol{n}}_e$ direction. A springlike potential, governed by Eq. (3), connects the beads back to back. The first bead is equipped with sensors on its surface (depicted as red cones), covering an area of approximately $\pi/2a^2(1-\cos 30^\circ) \approx 0.21a^2$ (a the radius of the swimmer). These sensors enable the swimmer to perceive surface stresses exerted by the surrounding fluid. The sensor locations are situated at the poles of each principal axis. (b) Schematic representation of the squirmer model, featuring relevant unit vectors and angles in spherical coordinates $\hat{\boldsymbol{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\phi}}$. Notably, $\hat{\boldsymbol{u}}$ signifies the swimming direction. (c, d) Illustrations outlining the contributions of the source dipole ($B_1 \sin\theta$), stresslet ($B_2 \sin 2\theta$), and rotlet ($C_1 \sin\theta$) to the surface velocity, respectively. (e) Diagram depicting the learning process employed to train a cargo-towing microswimmer for navigation in a zig-zag shear flow. The input to the policy network comprises surface stresses τ_6 and alignment with the light source $\hat{\boldsymbol{n}}_e \cdot \hat{\boldsymbol{n}}_L$. The network outputs an action corresponding to a rotation ($\hat{N}_{\text{rotlet}}$) or an increase in swimming speed ($\hat{\boldsymbol{v}}_{B_1}$). The dashed black line illustrates the swimming trajectory over a learning episode, with black circles indicating the swimmer's position at discrete simulation time steps. Red circles mark action steps where a new action a_I is adopted and sustained over the subsequent M simulation steps.

microswimmers towing cargo through viscous fluids, revealing insights into propulsion mechanisms and hydrodynamic efficiencies [25–28]. However, our understanding of the efficient navigation of these cargo-carrying microswimmers remains limited.

In this paper, our objective is to showcase the feasibility of designing an intelligent cargo-towing microswimmer by coupling direct numerical simulations [29], to fully incorporate hydrodynamic interactions, with deep Q learning, a subscheme of RL. We first examine the performance of a cargo-towing microswimmer tasked with navigating a zig-zag shear flow while carrying a load of up to three inert colloidal particles. The cargo-towing microswimmers are assumed to be able to freely rotate around their body axes, as well as possess the ability to change rotational and translational velocities. The combination of actions that shows the best performance is then used to demonstrate the navigation in the shear-gradient and flow direction against both naive cargo-towing microswimmers and normal colloidal chains of the same length. Subsequently, we analyze the optimal policies for each navigation task using a nonlinear dimension reduction technique. Lastly, we investigate the influence of learning across different swimming modes—pusher, puller, and neutral—under various load lengths.

II. SIMULATION METHODS

We study a load-carrying swimmer navigating a Newtonian fluid under an applied zig-zag shear flow. The dynamics of the particles and the host fluid can be determined by simultaneously solving the Newton-Euler equations and a modified Navier-Stokes equation, using the smoothed profile (SP) method [29].

The cargo-towing swimmer is represented as a flexible bead-spring model consisting of N beads. The principle axes of the ith bead are $\hat{\boldsymbol{u}}_i$, $\hat{\boldsymbol{v}}_i$, and $\hat{\boldsymbol{w}}_i$, illustrated in Fig. 1(a). The interaction between beads is modeled by a potential containing a steric repulsion interaction U_{LJ} and a bonding interaction U_B (for neighboring beads on the same chain):

$$U = \sum_{i<j} U_{\text{LJ}}(r_{ij}) + \sum_{\langle i,j\rangle} U_B(\boldsymbol{r}_{ij}) \tag{1}$$

where $\sum_{\langle i,j\rangle}$ denotes a sum over bonds. The steric repulsion is given by a truncated Lennard-Jones potential with power 36–18:

$$U_{\text{LJ}}(r_{ij}) = \begin{cases} 4\beta\left[\left(\frac{\sigma}{r_{ij}}\right)^{36} - \left(\frac{\sigma}{r_{ij}}\right)^{18}\right] + \beta, & r_{ij} < 2^{\frac{1}{6}}\sigma \\ 0, & r_{ij} > 2^{\frac{1}{6}}\sigma \end{cases} \tag{2}$$

where $\boldsymbol{r}_{ij} = \boldsymbol{R}_i - \boldsymbol{R}_j$, $r_{ij} = |\boldsymbol{r}_{ij}|$, β characterizes the strength of the interactions, and σ is the diameter of the beads. The beads are connected and aligned back to back by a bonding potential of the form

$$U_B(\boldsymbol{r}_{ij}) = \frac{1}{2}k_c R_0^2 \ln\left[1 - \left(\frac{r_{ij}}{R_0}\right)^2\right] + \frac{1}{2}k_{c,u}\left(\theta_i^2 + \theta_j^2\right) + \frac{1}{2}k_{c,v}\theta_v^2 \tag{3}$$

where $\cos(\theta_i) = \hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{r}}_{ij}$, $\cos(\theta_j) = \hat{\boldsymbol{u}}_j \cdot \hat{\boldsymbol{r}}_{ij}$, $\cos(\theta_v) = \hat{\boldsymbol{v}}_i \cdot \hat{\boldsymbol{v}}_j$ [Fig. 1(a)]. The constants $k_c = 30\beta/\sigma^2$, $k_{c,u} = k_{c,v} = 50\beta$. The forces and torques on each beads due to the potential in Eq. (3) can be derived following [30]. The first bead is conceptualized as a microswimmer, with the subsequent load/cargo beads assumed to be inert, Fig. 1(a). The microswimmers are characterized using the self-propelled spherical "squirmer" model [31,32]. This model, originally devised to study the motion of a ciliated particle propelled by synchronized ciliary beating, represents the microswimmers as rigid spherical particles with a slip velocity at their surface. This slip velocity is determined by an infinite expansion of polar and azimuthal modes:

$$\mathbf{u}^s(\theta,\phi) = \sum_{n=1}^{\infty} \frac{2}{n(n+1)} B_n P_n'(\cos\theta)\sin\theta\hat{\boldsymbol{\theta}} + \sum_{n=1}^{\infty} C_n P_n'(\cos\theta)\sin(\theta)\hat{\boldsymbol{\phi}} \tag{4}$$

where $\hat{\boldsymbol{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\phi}}$ are the unit vectors in the radial, polar, and azimuthal directions at a given point (θ, ϕ) on the surface of the particle, with θ the polar angle and ϕ the azimuthal angle, as shown in Fig. 1(b). Usually, the expansion is considered up to the second order in B_n, and the azimuthal components are neglected. In this paper, we explicitly consider the azimuthal component to the first order; thus, the slip velocity at a given point on the surface of the spherical swimmer can be evaluated as

$$\mathbf{u}^s(\theta,\phi) = B_1\left(\sin\theta + \frac{\alpha}{2}\sin 2\theta\right)\hat{\boldsymbol{\theta}} + C_1 \sin\theta\hat{\boldsymbol{\phi}}. \tag{5}$$

The coefficient B_1 represents the amplitude of the primary squirming mode and is directly related to the steady-state swimming velocity of the squirmer, denoted as $U = \frac{2}{3}B_1$. The parameter $\alpha = B_2/B_1$ determines the swimmer's classification: a negative α designates a pusher, exemplified by *E. coli*; an α of zero characterizes a neutral swimmer, typified by *Paramecium*; and a positive α identifies a puller, such as *C. Reinhardtii*. The source dipole term decays proportionally to $1/r^3$, whereas the stresslet term B_2 decays as $1/r^2$ [33]. With the inclusion of the azimuthal component C_1, the squirmer gains the ability to rotate around a body-fixed axis ($\hat{\boldsymbol{u}}$, $\hat{\boldsymbol{v}}$, or $\hat{\boldsymbol{w}}$) autonomously [24], deviating from an earlier investigation that required an external torque [23]. This primary azimuthal mode C_1, termed the rotlet, is analogous to the B_1 mode responsible for self-propulsion. This mode exhibits a decay rate of r^{-2} [34]. The particle dynamics are governed by the Newton-Euler equations. For a spherical particle of radius a, with center-of-mass position $\boldsymbol{R}_i$, velocity $\boldsymbol{V}_i$, orientation matrix $\boldsymbol{Q}_i$, angular velocity $\boldsymbol{\Omega}_i$, with skew symmetric angular velocity skew($\boldsymbol{\Omega}_i$), and inertia tensor $\boldsymbol{I}_p (= 2/5 M_p a^2 \boldsymbol{I})$ and mass $M_p (= \frac{4}{3}\pi a^3 \rho_p)$ we have

$$\begin{aligned} \dot{\boldsymbol{R}}_i &= \boldsymbol{V}_i, \\ \dot{\boldsymbol{Q}}_i &= \text{skew}(\boldsymbol{\Omega}_i)\cdot\boldsymbol{Q}_i, \\ M_p\dot{\boldsymbol{V}}_i &= \boldsymbol{F}_i^H + \boldsymbol{F}_i^C, \\ \boldsymbol{I}_p\cdot\dot{\boldsymbol{\Omega}}_i &= \boldsymbol{N}_i^H + \boldsymbol{N}_i^C. \end{aligned} \tag{6}$$

The forces on the particles, as expressed in Eq. (6), comprise the hydrodynamic force $\boldsymbol{F}^H$ and the particle-particle forces $\boldsymbol{F}^C$, given by the potential in Eq. (2). Torques are similarly decomposed into hydrodynamic $\boldsymbol{N}^H$ and particle-particle $\boldsymbol{N}^C$, arising from the bonding potential in Eq. (2). The conservation of momentum guides the determination of the hydrodynamic forces and torques, ensuring a coherent coupling between the host fluid and the swimming particles. Simultaneously, the evolution of the host fluid within the SP method involves solving a modified Navier-Stokes equation for the total velocity $\boldsymbol{u} = \boldsymbol{u}_f + \boldsymbol{u}_p$, incorporating contributions from both host fluid $\boldsymbol{u}_f$ and particle $\boldsymbol{u}_p$ components. The sharp particle interface is replaced by an interface of finite thickness ξ, using a smooth and continuous particle phase field ϕ. This phase field, taking values of 1 within the particle domain, zero in the host fluid, and smoothly interpolating between these values within the interface, allows for the definition of necessary particle fields (e.g., $\boldsymbol{u}_p$). In this context, both fluid and particle domains are identified using the same phase field $\rho_f = \rho_p = \rho$ being set for this paper. The Navier-Stokes equation is expressed as

$$\rho(\partial_t + \boldsymbol{u}\cdot\nabla)\boldsymbol{u} = \nabla\cdot\boldsymbol{\sigma} + \rho(\phi\boldsymbol{f}_p + \phi\boldsymbol{f}_{\text{sq}} + \phi\boldsymbol{f}_{\text{shear}}). \tag{7}$$

The second term on the right-hand side of the equation consists of three force contributions: $\phi\boldsymbol{f}_p$, the constraint force ensuring momentum conservation; $\phi\boldsymbol{f}_{\text{sq}}$, the force arising from the squirming motion; and $\phi\boldsymbol{f}_{\text{shear}}(\boldsymbol{x}, t)$, an external force sustaining a zig-zag velocity profile:

$$v_x(y) = \begin{cases} \dot{\gamma}(-y - L_y/2), & -L_y/2 < y \leqslant -L_y/4 \\ \dot{\gamma}y, & -L_y/4 < y \leqslant L_y/4 \\ \dot{\gamma}(-y + L_y/2), & L_y/4 < y \leqslant L_y/2 \end{cases} \tag{8}$$

where $\dot{\gamma}$ is the shear rate, y represents the distance in the velocity-gradient direction, and L_y is the height of the simulation box with dimensions (L_x, L_y, L_z). Further details regarding this methodology, including its implementation, accuracy, and applications, can be found in previous works [29,35].

The cargo-towing ability of the swimmer navigating the imposed flow relies fundamentally on the learned policies, establishing a crucial mapping between sensory signals and corresponding responses. Tuned meticulously through a RL framework [36], these network/policy hyperparameters ensure the acquisition of optimal behaviors. Specifically, these policies empower the swimmer to select precise actions that yield optimal long-term rewards. Employing a deep Q-learning strategy, combined with prioritized experience replay and n-step learning [37–39], enables the acquisition of navigation strategies for a given task. In the RL framework, two main actors interact: the agent (swimmer) and the environment. The agent processes information from the environment

to determine its current *state*, leading to an *action*. After each action, the environment provides feedback in the form of a *reward* and the revelation of a *new state*.

During training, the swimmer's goal is maximize expected rewards over predefined episodes. These episodes are further discretized into N_s action segments, each consisting of M simulation steps of length Δt [see Fig. 1(e)]. The total duration of an episode ($T_{\text{episode}} = N_s \Delta T$), where $\Delta T = M\Delta t$, marks the duration of an action segment. At the beginning of action segment I, corresponding to simulation time step $i = IM$ and time $T_I = (IM) \cdot \Delta t \equiv t_{i=IM}$, the swimmer perceives the current state s_I, choosing an action a_I dictated by a policy function π. This chosen action will be carried over the next M simulation steps (one action segment). At the end of the action segment, the swimmer advances s_I to a new state $s'_I = s_{I+1}$, with a corresponding reward r_I that depends on the initial and final states. For the purposes of the learning, the states of the system at the intermediate simulation time steps $t_{IM} < t < t_{(I+1)M}$ are irrelevant; only the initial and final states are needed. The tuple of this set of experiences (action-reward) can be written as (s_I, a_I, s'_I, r_I).

The training regimen compiles rewards at the end of each episode, by summing up the individual rewards for each action step $r = \sum_{I=0}^{N_s/M} r_I$. The expected cumulative reward after an action at time step T_I can be evaluated from the action-value function (Q function) $Q_\pi(s_I, a_I) = r_I + \gamma r_{I+1} + \gamma^2 r_{I+2} + \ldots$. Here $\gamma \in [0, 1)$ is the discount factor, and π is the policy function mapping between states and action. The optimal policy π^* that maximizes the long-term reward should fulfill the Bellman equation $Q_{\pi^*}(s_t, a_t) = r_{t+1} + \gamma \max_a Q_{\pi^*}(s_{t+1}, a)$ [36]. This Q function is represented by a neural network, which is trained with a Bellman-informed loss function (i.e., the loss is zero if the Bellman equation is satisfied). To maximize long-term rewards, the agent selects actions using an ϵ-greedy scheme, balancing exploration and exploitation. Actions are determined by maximizing the current value function, $a_I = \text{argmax}_a Q\pi(s_I, a)$, with probability $1 - \epsilon$, otherwise chosen randomly with probability ϵ, resulting in an ϵ-greedy selection scheme. Initially set to 1, ϵ exponentially decays (decay rate k) to 0.015 to encourage exploration in early training. As the policy approaches optimality, actions align with the trained policy. The final ϵ value is set to a small nonzero positive number to foster adaptability.

During each episode, the swimmer's positions and orientations are randomly initialized, and the agents execute a total of 2×10^3 action steps. At each action step, the state s_I, action a_I, next state s'_I, and corresponding reward r_I are stored in a "replay memory," with a maximum size of $N_{P_{\max}}$. Subsequently, a batch of these stored experiences (size N_b) is drawn during each action step and input into the optimizer to adjust the hyperparameters of the Q function neural network. The update rule is expressed as

$$Q(s_I, a_I) \leftarrow Q(s_I, a_I) + \alpha_I [r_I + \gamma \max_a Q(s_{I+1}, a) - Q(s_I, a_I)] \tag{9}$$

where α_I represents the learning rate. Details of this update rule can be found in [37]. The loss function utilized for the Q network training is defined as

$$L_I(s_I, a_I; \boldsymbol{\theta}_I) = [Y_I - Q(s_I, a_I; \boldsymbol{\theta}_I)]^2. \tag{10}$$

Here, $\boldsymbol{\theta}$ represents the Q function network parameters, and $Y_I = r_I + \gamma \max_a Q(s_{I+1}, a; \boldsymbol{\theta}_I^-)$ is the target reward at learning step I. We note that the the target reward Y_I is evaluated using another neural network called the target network, whose parameters $\boldsymbol{\theta}_I^-$ are adjusted to mirror those of the predicting Q network ($\boldsymbol{\theta}_I$) every C steps and are kept constant in between intervals. To enhance learning stability, the target accumulated reward estimation Y employs the target network (with parameters $\boldsymbol{\theta}^-$) instead of the Q network (with parameters $\boldsymbol{\theta}$), improving the prediction of the expected accumulated reward. Notably, both networks (predicting and target) are identical; the difference lies only in the network parameters. Finally, the gradient with respect to the weights $\boldsymbol{\theta}$ is given by

$$\nabla_{\boldsymbol{\theta}_I} L(\boldsymbol{\theta}_I) = [Y_I - Q(\boldsymbol{\theta}_I)] \nabla_{\boldsymbol{\theta}_I} Q(\boldsymbol{\theta}_I). \tag{11}$$

In this paper, the cargo-towing swimmer is endowed with the capability to execute two distinct categories of actions: rotations and variation in the translational and/or rotational velocities [related to B_1 and C_1 in Eq. (5) respectively]. Rotations, originating in the azimuthal C_1 component of the surface velocity, occur about one of the principal body axes, with the rotational velocity proportional to $\Omega_{C_1} = C_1/a$. Consequently, there exist seven total actions associated with rotation: six potential rotational actions [see Fig. 1(a)], for rotations around $\{\pm u, \pm v, \pm w\}$ with respect to body frames, in addition to the passive action of "no rotation" ($C_1 = 0$). At the onset of each episode, the initial values are set as $B_1^0 = 0.1$ and $C_1^0 = 0.6$. At each action time step I, the swimmer possesses the ability to accelerate/decelerate its translational velocity ($B_1^I = B_1^{I-1} \pm \Delta B_1$) and/or rotational velocity ($C_1^I = C_1^{I-1} \pm \Delta C_1$), where $\Delta B_1 = 0.02$, $\Delta C_1 = 0.1$, and B_1^I and C_1^I are constrained within $0.02 \leqslant B_1^I \leqslant 0.2$ and $0.1 \leqslant C_1^I \leqslant 1.0$ respectively. The former constraint ensures that the swimmer maintains a minimum velocity above the lower limit, allowing it to continue moving with respect to the background fluid, while the latter ensures the existence of finite rotational velocity. The reward per action time step is defined by $r_I^\mu = R_{I+1}^\mu - R_I^\mu$, where μ is the desired swimming direction, i.e., along the flow direction x or in the shear-gradient direction y. Note that we do not consider the reward in the vorticity direction due to its decoupling from the shear flow [23,24]. The definition of the reward outlined above can also be regarded as a simple approximation for a swimmer that is able to detect gradients in a given scalar field (e.g., gradient in nutrient concentration), with the reward expressed as $r_I^{\nabla C} = C(\boldsymbol{R}_{I+1}) - C(\boldsymbol{R}_I)$. The active swimmer is presumed capable of detecting the surface stress $\tau_i (0 \leqslant i < N_\tau)$, with N_τ representing the total number of sensors strategically positioned at antipodal points along the intersection of its three principal body axes [as depicted in Fig. 1(a)]. The three-dimensional (3D) signals from each sensor contribute to the state representation alongside the visual signal $\hat{\boldsymbol{n}}_e \cdot \hat{\boldsymbol{n}}_L$.

For the simulation parameters, we adopt as characteristic units the grid spacing Δ, viscosity η, and density of the host fluid ρ_f. These units establish the time and mass units as $\rho_f \Delta^2/\eta$ and $\rho_f \Delta^3$ respectively. The bead radius a and

interface thickness ξ are set to 5Δ and 2Δ, respectively. Our simulation employs a box dimension of $32\Delta \times 64\Delta \times 32\Delta$ with full periodic boundary conditions across all dimensions. The applied shear rate is denoted as $\dot{\gamma} = 0.04\eta/(\rho_f \Delta^2)$, corresponding to a Reynolds number $\mathrm{Re} \approx 1$. We explore scenarios where the number of load particles carried by the swimmer d_{cargo} falls within the set $\{1, 2, 3\}$. For most of the cases presented here, and unless stated otherwise, the swimmer is set to be neutral with $\alpha = 0$.

Regarding the learning parameters, we use as discount rate $\gamma = 0.99$, learning rate $\alpha_l = 0.000\,25$, batch size $N_b = 128$, and maximum size of the replay memory $N_{P_{\max}} = 10^6$. The parameters of the target network are updated with those of the Q network every $C = 100$ action steps, and the greedy decay rate for ϵ is set as $k = 0.981$. Each Q network comprises an input layer with a number of neurons corresponding to the number of three-dimensional surface stress signals obtained from the sensors, i.e., 18, in addition to the visual signal $\hat{\boldsymbol{n}}_e \cdot \hat{\boldsymbol{n}}_L$, three hidden layers with 100 neurons each, and an output layer with neurons matching the size of the action space. A learning episode spans $N_s = 2 \times 10^3$ action steps, with an action interval of $M = 10$ simulation time steps. The policy network is trained over 1000 episodes. It should be noted that the choice of action time step duration is purely due to computational limitations. Higher frequency updates would lead to latency issues with our current computational resources. By setting $M = 1$, one can expect qualitatively similar trained policies from the training protocol. At steady state, by setting $M = 10$, $B_1 = 0.1$, and $C_1 = 0.6$ the swimmer is able to translate $2/3B_1 M\Delta t \approx 0.0095$ times of its diameter and rotate approximately $C_1/aM\Delta t \approx 0.09$ rad due to the C_1 mode. The shear rate strength is calibrated to ensure that the swimmers move slowly compared with the background fluid speed. Concurrently, this setup permits the examination of regions with varying shear gradients throughout the simulation time steps. Additionally, the selection of learning hyperparameters is designed to expedite the convergence to the optimal policy, as elaborated in our prior study [23].

III. RESULTS AND DISCUSSIONS

We consider a microswimmer both with and without cargo, that is trained to perform several swimming tasks. Given the imposed zig-zag flow and the periodic boundary conditions, the chain can swim up, enter the negative flow velocity region, u^x_{f-}, or descend toward the positive flow region, u^x_{f+}. Initially, we explore the navigation performance across various combinations of state inputs, i.e., surface stresses and alignment with the light source, along with the corresponding memories. Subsequently, we analyze different combinations of actions, allowing for dynamic adjustments in swimming and rotational velocities through the tuning of B_1 and C_1 parameters, respectively. The set of actions demonstrating the best performance will be showcased against naive cargo-towing and inert colloid chains, where all beads are inert. Finally, we compare the navigation performance across different types of swimmers. First, three types of input signals are perceived by the swimmer: the current time step surface stresses and alignment with the light source $\{\tau_i^I, (\hat{\boldsymbol{n}}_e \cdot \hat{\boldsymbol{n}}_L)^I\}$, the current and the previous time step surface stresses $\{\tau_i^I, \tau_i^{I-1}\}$, and the current and the previous time steps of both surface stresses and alignment with the light source $\{\tau_i^I, \tau_i^{I-1}, (\hat{\boldsymbol{n}}_e \cdot \hat{\boldsymbol{n}}_L)^I, (\hat{\boldsymbol{n}}_e \cdot \hat{\boldsymbol{n}}_L)^{I-1}\}$. The velocities of the swimmer tasked with navigating in the flow and shear-gradient directions using the aforementioned signals are shown in Fig. 2 with solid-circle, dashed, and solid-triangular lines respectively. Across all examined cases, it becomes evident that providing $\{\tau_i^I, \hat{\boldsymbol{n}}_e \cdot \hat{\boldsymbol{n}}_L\}$ results in the best performance, followed by providing a combination of surface stresses and light alignment (with their respective memories); only providing surface stresses (and their memory) yields the worst performance. This indicates that the combination of hydrodynamics and the alignment to the light source is essential for efficient navigation. It is noted that although the input including the memories of both surface stresses and light alignment achieves the same level of performance as the input without memories, the convergence toward optimal behavior is slower for the former. This is due to the increased complexity of the input/action space, which requires more time to reach optimality [36].

Subsequently, we investigate various action combinations, employing both fixed values of B_1 and C_1, as well allowing for dynamic variations in their magnitudes. Figures 2(a) and 2(e) show the results obtained for a swimmer with a static $B_1 = 0.1$, which is able to rotate with a rotational velocity proportional to $C_1 = 0.6$. Additionally, the velocity of a swimmer without a load, which is comparable to the case studied in [24], is depicted in red. The value of B_1 is selected to ensure that the swimmer moves slowly compared to the fluid speed, while still allowing it to explore regions with different shear gradients within the time step of an episode [23]. Conversely, C_1 is chosen based on its demonstrated optimal performance for a swimmer without cargo. Specifically, $C_1 = 0.6$ enables the swimmer to actively rotate roughly six times faster than the rotation induced by the shear flow [24]. For navigation in the flow direction [Fig. 2(a)], the x component of the velocity, u_p^x, shows that the swimmer without the load exhibits inferior performance compared to the load-carrying swimmer. This discrepancy arises from the nonspherical shape of the swimmer with cargo, which has a propensity for tumbling under shear flow, thereby promoting reorientation and realignment in the flow direction [40,41]. Conversely, navigation in the shear-gradient direction heavily relies on the swimming velocity, as evidenced by the decline in the y component of the swimmer's velocity, u_p^y, with an increase in the length of load (Appendix A). It is noteworthy that the swimming velocity of a spherical swimmer is $U = 2/3B_1 = 0.06$, comparable to the speed of a swimmer without any load indicated by the red line.

Now we consider the case where the parameter B_1 dynamically adjusts at each action time step I, following the expression $B_1^I = B_1^{I-1} \pm \Delta B_1$, subject to the constraint $0.02 \leqslant B_1^I \leqslant 0.2$, while C_1^I remains static at 0.6, while still allowing the axis of rotation to be switched. In Fig. 2(b), the u_p^x component of the load-carrying swimmer ($d_{\mathrm{cargo}} \geqslant 1$) exhibits an optimal value comparable to the velocity of the shear flow at its peak, i.e., $u_f^x = \dot{\gamma} L_y/4 = 0.64$ [Eq. (8)], surpassing that of the loadless swimmer, thereby underscoring the cargo-towing swimmer's superior reorientation and alignment with the flow. Across all load lengths examined in

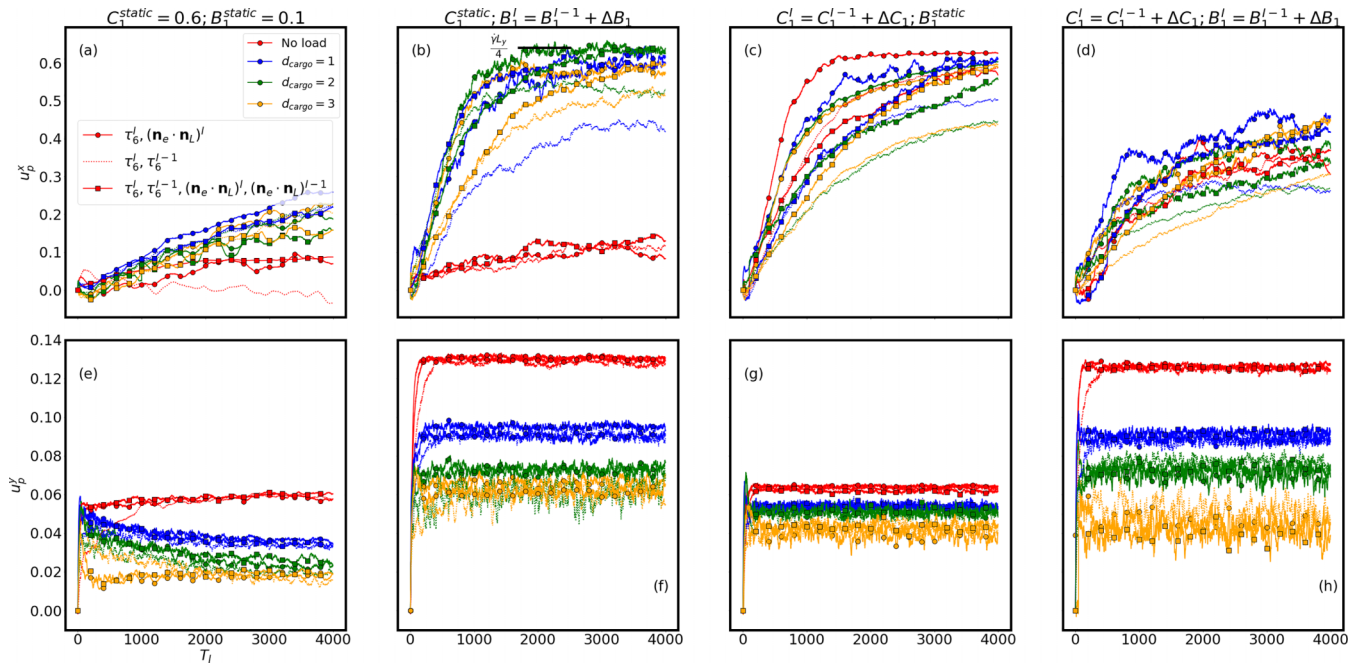

FIG. 2. The velocities of a trained cargo-towing swimmer, with the load length of $d_{\text{cargo}} \in \{1, 2, 3\}$ observed over a navigation course comprising a total of $T_I = 10^4$ action time steps. Here we consider three sets of input signals: $\{\tau_i^I, (\hat{\boldsymbol{n}}_e \cdot \hat{\boldsymbol{n}}_L)^I\}$, $\{\tau_i^I, \tau_i^{I-1}\}$, and $\{\tau_i^I, \tau_i^{I-1}, (\hat{\boldsymbol{n}}_e \cdot \hat{\boldsymbol{n}}_L)^I, (\hat{\boldsymbol{n}}_e \cdot \hat{\boldsymbol{n}}_L)^{I-1}\}$. (a)–(d) The x component of the velocity of a swimmer trained to navigate in the flow direction. (e)–(h) The y component of the velocity of a swimmer trained to navigate in the shear-gradient direction. Each column showcases navigation outcomes using constant $C_1 = 0.6$ and $B_1 = 0.1$; constant $C_1 = 0.6$ and dynamic $B_1^I = B_1^{I-1} + \Delta B_1$, where B_1^I represents B_1 at action time step I constrained between $0.02 \leqslant B_1 \leqslant 0.2$; constant $B_1 = 0.1$ and dynamic $C_1^I = C_1^{I-1} + \Delta C_1$, where C_1^I denotes C_1 at action time step I capped between $0.1 \leqslant C_1 \leqslant 1.0$; and both dynamic $C_1^I = C_1^{I-1} + \Delta C_1$ and $B_1^I = B_1^{I-1} + \Delta B_1$, respectively. Note that the velocities of a swimmer without any load (no load) are also plotted for each case for comparison.

this paper, the u_p^x component demonstrates that the swimmer adeptly aligns with the maximum flow region irrespective of its length. Conversely, for navigation in the shear-gradient direction [Fig. 2(f)], the loadless swimmer demonstrates the highest velocity equal to $2/3B_1 = 0.133$ when $B_1 = 0.2$, its upper limit. As previously noted, the velocity during shear-gradient navigation is heavily contingent upon the swimming velocity of the cargo-towing swimmer, as evident in Fig. 2(e), where u_p^y decreases with an increasing in the length of the load. The augmented navigation speed in this scenario directly results from the heightened maximum swimming velocity, compared to the static $B_1 = 0.1$ condition depicted in Fig. 2(e).

Next, we investigate the scenario where the parameter C_1 dynamically adjusts according to the expression $C_1^I = C_1^{I-1} \pm \Delta C_1$, constrained within $0.1 \leqslant C_1^I \leqslant 1.0$, while B_1^I remains fixed at 0.1. In Fig. 2(c), when navigating in the flow direction, both the swimmer with and without load optimally align at the position of the maximum background shear flow $\dot{\gamma} L_y/4$. Prior studies have highlighted the challenge in achieving this task with a spherical swimmer due to limitations in adequately exploring the action/state space with full 3D rotations [23,24], a challenge which we successfully overcome. In Fig. 2(g), when navigating in the shear-gradient direction, the performance of the swimmer with the load surpasses that of the static case depicted in Fig. 2(e). However, compared to the dynamic B_1 cases, this performance is inferior due to differences in the maximum swimming velocities.

Lastly, we consider the scenario where both parameters B_1 and C_1 are dynamically adjusted at each action time step. When swimming in the flow direction, enabling dynamic changes for both parameters results in suboptimal performance compared to enabling changes for only one parameter. This is evidenced by the decrease in u_p^x depicted in Fig. 2(d) compared to Figs. 2(b) and 2(c). This decline may stem from the expanded action space. As discussed above, the task of swimming in the flow direction is more sensitive to the sampling/exploration of the action/state space, necessitating a longer training period using the same network structure to achieve equivalent performance [42], as obtained in the other two cases, i.e., with dynamical change in either B_1 or C_1, discussed above. Conversely, the expanded action space does not significantly impact training for swimming in the shear-gradient direction. The performance of the trained swimmer remains comparable to that achieved with the dynamic B_1 parameter, underscoring the high dependency of this task on swimming velocity.

Next, we will demonstrate the navigation capabilities in both flow and shear-gradient directions, employing the policies obtained for the cargo-towing swimmer with load length $d_{\text{cargo}} = 2$. In this scenario, only the B_1 parameter is allowed to change at each action time step, as this action set exhibited the best performance compared to the other sets considered. In Fig. 3, the x component of the velocity for three types of the three-bead chains, i.e., normal-three-bead chain (red), naive cargo-towing swimmer with $d_{\text{cargo}} = 2$ (blue), and smart

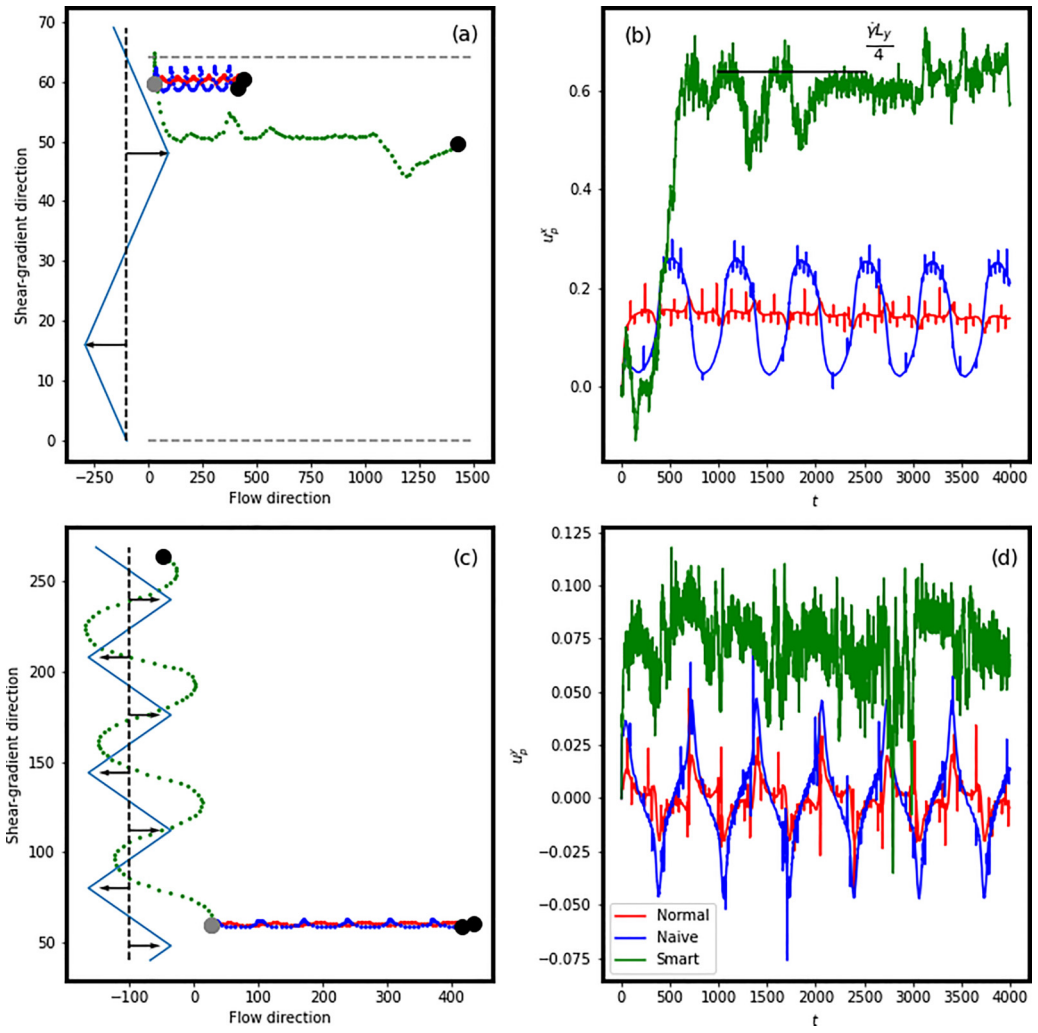

FIG. 3. The velocities and the trajectories of a trained (green)/naive (blue) cargo-towing swimmer, with the load length $d_{\text{cargo}} = 2$, and an inert chain (red) comprised of three beads. Trajectories are shown for the tasks of swimming in the flow (a), (b) and shear-gradient (c), (d) directions over $T_l = 10^4$ action time steps. Here $C_1 = 0.6$ and $B_1^l = B_1^{l-1} + \Delta B_1$ ($0.02 \leqslant B_1 \leqslant 0.2$).

cargo-towing swimmer with $d_{\text{cargo}} = 2$ (green), along with their trajectories along the shear plane, for the tasks of navigating in the flow and shear-gradient directions, are shown. In the context of swimming in the flow direction [Figs. 3(a) and 3(b)], all chains are initially placed near the top of the simulation box, where the background flow velocity is zero. Note that the choice of the initial location of the swimmer is arbitrary and does not affect the navigation performance in either direction for the trained agent (Appendix B). The trajectories depicted in Fig. 3(a) clearly demonstrate that the smart cargo-towing swimmer effectively navigates toward the maximum flow region while maintaining alignment with the flow. In contrast, the naive cargo-towing swimmer and the normal chain become ensnared between the positive and negative vorticity regimes. The x component velocity plot presented in Fig. 3(b) underscores the distinction in navigation performance. The velocity of the smart cargo-towing swimmer attains the maximum velocity of the background shear flow, $\dot{\gamma} L_y/4$, while the naive cargo-towing swimmer and the normal chain exhibit periodic rotation, indicative of being trapped between regions with opposite vorticities. Notably, the velocity magnitude of the naive cargo-towing swimmer surpasses that of the normal chain due solely to its activity. For the task of navigation in the shear-gradient direction [Figs. 3(c) and 3(d)], all chains are initially positioned in a region where the background flow velocity is close to zero and positive, i.e., at $L_y = 60$ with an initial orientation of 45° with respect to the flow direction, the same as the initial setup for the navigation task in the flow direction. The trajectories depicted in Fig. 3(c) illustrate the optimal navigation of the smart cargo-carrying swimmer through the shear gradient, while the naive cargo-towing swimmer and the normal chain are drawn in the flow direction and experience periodic rotations influenced by the background shear flow. The y component velocities plotted in Fig. 3(d) confirm that the smart cargo-towing swimmer has reached its maximum swimming speed (Appendix A), while the naive cargo-towing swimmer and the normal chain exhibit periodic rotations.

To characterize the learned policies, we examine their representations to better understand the successful execution of the navigation tasks. Our focus lies in unraveling the intricate mapping between high-dimensional input signals and their corresponding actions given by the neural network. In pursuit of this understanding, we adopt a

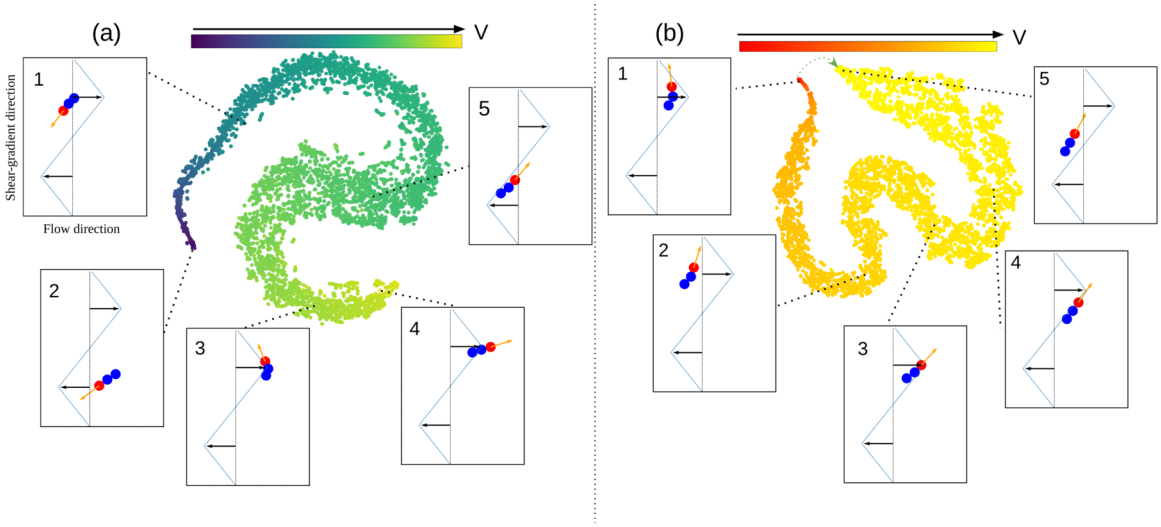

FIG. 4. The 2D-projection representations of the output last hidden layer, i.e., state values using t-SNE for the policies governing the task of (a) navigation in the flow direction and (b) navigation in the shear-gradient direction for a cargo-towing swimmer with $C_1 = 0.6$ and dynamic $B_1^I = B_1^{I-1} + \Delta B_1$. The cargo length is $d_{\text{cargo}} = 2$ and the swimmer performs the navigation over a course of $T_I = 10^4$ action time steps. Each data point within the visualization corresponds to a 2D representation of perceived hydrodynamic signals at a specific state, color coded based on the state value $V_I(s) = \max_V Q_{\pi^*}(s_I, a_I)$ predicted by the learned policies.

nonlinear dimensional reduction technique developed for the visualization of high-dimensional data called t-distributed stochastic neighbor embedding (t-SNE) [43]. This technique maps high-dimensional data to a lower-dimensional space, typically two-dimensional (2D) or 3D, while preserving the relative distances between data points. We consider the same navigation tasks as shown in Fig. 3 and extract the acquired representations from the final hidden layer, i.e., the output of the neural network, onto a 2D plane as shown in Fig. 4. Each point is color coded based on its state value, defined by

$$V(s) = \max_a Q_{\pi^*}(s, a) \tag{12}$$

where $V_I(s)$ is the expected maximum total reward that a swimmer can achieve at action time step I, upon perceiving the state s_I and following the optimal policy π^*. Thus, a higher (lower) value of $V_I(s)$ indicates that starting navigation from this state s will result in more (less) efficient navigation.

It is imperative to emphasize that the spatial arrangement of these plotted points holds no significance. Instead, our focus rests solely upon the close juxtaposition of representations corresponding to perceptually akin states and the utilization of color codes, indicative of the expected maximum total reward associated with each state. For the task of navigation in the flow direction, Fig. 4(a), when the swimmer swims in the region of negative flow velocity $u^x_{f^-}$ with the swimming direction towards the negative flow direction, the state value reaches its minimum [Fig. 4(a2)]. On the other hand, when the swimmer swims in the region of positive flow velocity $u^x_{f^+}$ with the aligned swimming direction, the state value peaks [Fig. 4(a4)]. For the swimmer that swims in the $u^x_{f^+}$ region with an orientation pointing away from the region [Fig. 4(a1)], the state value is lower compared to the swimmer that swims in the $u^x_{f^-}$ region with a swimming direction towards the $u^x_{f^+}$ region [Fig. 4(a5)]. For the task of navigation in the shear-gradient direction, Fig. 4(b), the state value shows the periodicity, i.e., when the swimmer swims in the vicinity of the zero velocity region, $u^x_{f^0}$, and the swimming direction aligns with the background shear streamline, the state value reach its maximum. In contrast, when the swimmer enters the nonzero velocity region, the state value will be lower and reach its minimum when the swimmer arrives at the highest velocity region, regardless of the flow direction. Notably, swimmer reorientation by shear necessitates a perpendicular swimming direction, facilitating alignment with subsequent shear streamlines, resulting in a shift from minimum to maximum state value, as indicated by the dotted green arrow in Fig. 4(b).

In addition to examining the neutral swimmer ($\alpha = 0$) as discussed throughout this paper, we extend our analysis to include other swimmer types characterized by squirming parameters $\alpha = -2$ (pusher) and $\alpha = 2$ (puller). In Fig. 5, the x and y components of the velocity of the load-carrying swimmer with length $d_{\text{cargo}} \in \{1, 2, 3\}$, along with that of a swimmer with no load, tasked with swimming in the flow and shear-gradient directions are shown, respectively. The rotational velocity of the swimmer remains fixed and proportional to $C_1 = 0.6$, while the swimming velocity dynamically adjusts, as set by B_1 within the range of $0.02 \leqslant B_1 \leqslant 0.2$. In the context of navigation in the flow direction, regardless of swimmer type, the swimmer without load exhibits the lowest swimming velocity [Fig. 5(a)]. However, with the addition of cargo, the x component of the swimmer velocity reveals an effective alignment with regions of maximum velocity, irrespective of swimmer type. Particularly intriguing is the observation across load lengths, where swimmers with a load length of $d_{\text{cargo}} = 2$ demonstrate the swiftest convergence to the maximum flow regime. Conversely, in navigating the shear-gradient direction, for the loadless swimmer [Fig. 5(e)],

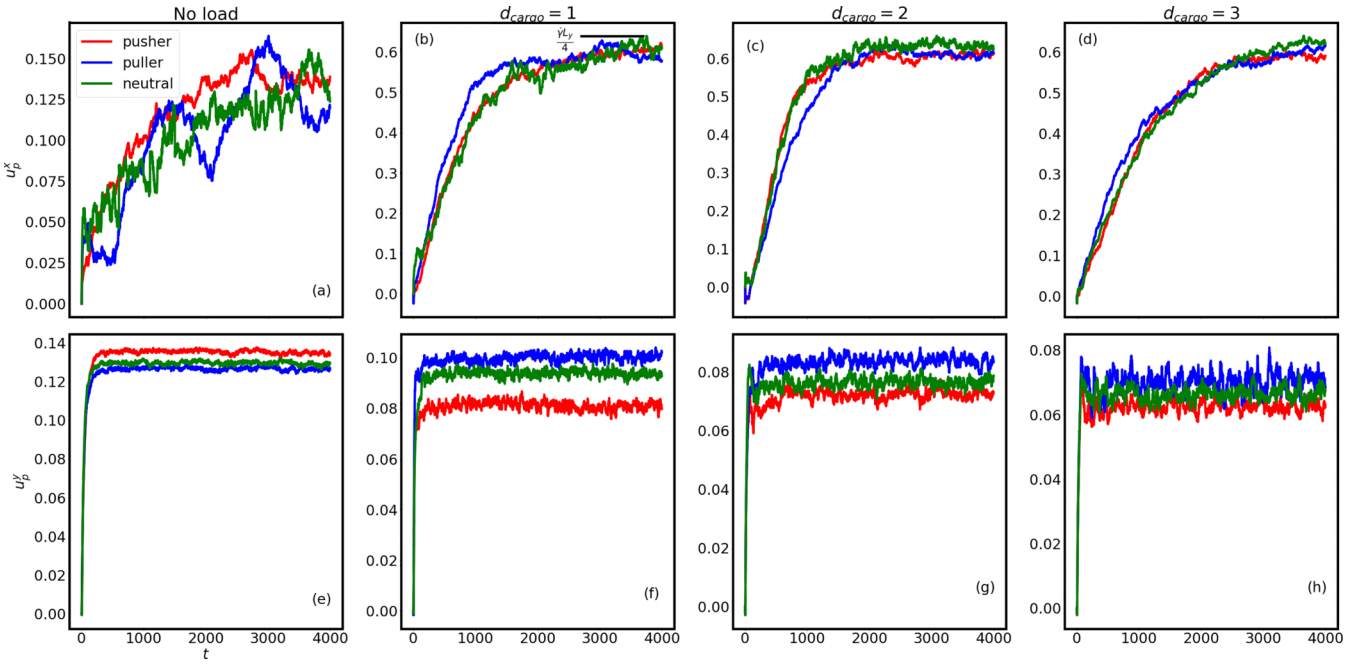

FIG. 5. The velocities of a trained cargo-towing swimmer, with the cargo length $d_{\text{cargo}} \in \{1, 2, 3\}$ along with the swimmer with no load (a), (e) observed over a navigation course comprising a total of $T_l = 10^4$ action time steps. (a)–(d) The x component of the velocity of a cargo-towing microswimmer trained to navigate in the flow direction. (e)–(h) The y component of the velocity of a microswimmer carries the load trained to navigate in the shear-gradient direction. The rotation of the swimmer is fixed with $C_1 = 0.6$ while the translation velocity is dynamically changed with respect to $B_1^l = B_1^{l-1} + \Delta B_1$ constrained between $0.02 \leqslant B_1 \leqslant 0.2$.

pushers exhibit the highest swimming velocity, followed by neutral swimmers and pullers, consistent with previous findings utilizing fixed translational and rotational velocities [23]. However, with the addition of cargo, a notable reversal in swimming velocity trends is observed. Pullers now outpace neutral swimmers and pushers [Fig. 5(f)], a pattern consistently observed across varying load lengths [Figs. 5(g) and 5(h)]. This phenomenon can be attributed to the differential decay in flow velocity ahead of the puller-type swimmer carrying a load, which occurs at a faster rate compared to that of the cargo-towing pusher [44]. This differential decay plays a crucial role in reversing the swimming velocities of both types of cargo-towing swimmers, particularly influencing tasks requiring navigation in the shear-gradient direction.

IV. CONCLUSIONS

In conclusion, we have conducted direct numerical simulations, using the smoothed profile method, coupled with a deep reinforcement learning algorithm to study the optimal navigation of a load-carrying swimmer, where the load is represented by inert spherical particles, under an applied zig-zag shear flow. We consider loads of up to three colloids ($d_{\text{cargo}} \in \{1, 2, 3\}$). This flow was selected for its simplified and deterministic characteristics, making it well suited for elucidating fundamental mechanisms and validating theoretical models. Nonetheless, we acknowledge the importance of evaluating more complex and realistic flows, such as Kolmogorov flows, which could be considered in future work. The swimmer is designated to perform the tasks of navigation in flow and shear-gradient directions. The task of navigation in the shear-vorticity direction is not considered as it has been found to be decoupled from the shear flow. A neutral swimmer is initially considered with and without load. Several combinations of state inputs were examined, and we found that the optimal navigation requires the swimmer be able to perceive hydrodynamic forces and alignment to a light source that is perpendicular to the flow direction. This resulted in better performance compared to a swimmer that relies solely on hydrodynamic signals.

Next, combinations of actions were considered, such that the swimmer can dynamically adapt its translational and/or rotational velocities, by tuning the B_1 and C_1 parameters, respectively. In the case where both B_1 and C_1 are constant, the loadless swimmer showed the worse performance for the task of navigation in the flow direction, compared to the cargo-towing swimmer. This is due to the tumbling movements of the nonspherical swimmer, the swimmer with the load, under the influence of the shear flow. On the other hand, for the task of navigating in the shear-gradient direction, the loadless swimmer outperforms the load-carrying swimmer, whose performance decreases upon increasing the load, showing the strong dependence on the swimming velocity. When the parameter B_1 is allowed to dynamically change and C_1 is fixed, the load-carrying swimmer exhibits optimal navigation in the flow direction, as the swimmer can align with the stream at the point of maximum velocity, and thus outperform the loadless swimmer, regardless of the load length. For navigation in the shear-gradient direction, the swimmer without load shows the best performance, as it swims with the highest swimming velocity, as given by the upper bound of the capped B_1. The navigation performance drops as the length of the load increases due to the decrease in the swimming velocity. On the other hand, when C_1 is dynamically changed,

while B_1 is fixed, both types of swimmers, with or without load, show optimal swimming in the flow direction, with the loadless swimmer slightly outperforming the load-carrying swimmer. For the loadless swimmer in particular, this task has been found to be difficult to achieve [23], and by allowing the rotational velocity to change over time, we have successfully overcome this challenge. For navigating in the shear-gradient direction, the cargo-towing swimmer outperforms the load-carrying swimmer in the case of static B_1 and C_1, but is inferior compared to the case where B_1 is dynamically changed due to the differences in maximum swimming velocities. Once both B_1 and C_1 are allowed to change, the performance for the navigation in the flow direction is worse compared to the case where either of the parameters can be varied, due to the increase in the complexity in action space, while the navigation in the shear-gradient direction shows the optimal behavior.

The policy from the action space allowing only changes in B_1 is used to compare the performance for the navigation tasks against a naive cargo-towing swimmer, with a load length of $d_{\text{cargo}} = 2$, and a chain of three inert colloidal beads. The trained swimmer demonstrates superior navigation abilities compared to the naive cargo-towing swimmer and the inner colloid chain, which experience periodic rotations driven by the background shear flow. The obtained policies are then analyzed using the t-SNE technique which offer insight into the complex interplay between the perceived multidimensional hydrodynamic signals and their representations upon the swimmer's position and orientation within the shear plane. In the flow direction, the state value is at its minimum when the swimmer moves in the region of negative flow velocity with the swimming direction towards the negative flow. Conversely, it peaks when the swimmer swims in the region of positive flow velocity with an aligned swimming direction. Notably, the state value is lower when the swimmer swims in the positive flow velocity region with the orientation heading away from it, compared to swimming in the negative flow velocity region towards the positive flow velocity region. In the shear-gradient direction, the state value exhibits periodicity. It reaches its maximum when the swimmer swims near the zero velocity region with the swimming direction aligned with the background shear streamline. Conversely, it decreases as the swimmer enters the nonzero velocity region, reaching its minimum when the swimmer arrives at the highest velocity region, regardless of flow direction. Shear-induced swimmer reorientation necessitates a perpendicular swimming direction, facilitating alignment with subsequent shear streamlines, resulting in a transition from minimum to maximum state value. Finally, we examined the performance of the navigation across the types of swimmer by comparing pullers, pushers, and neutral swimmers. In the context of navigating along the flow direction, our findings indicate that swimmers without any load demonstrate inferior performance compared to a load-carrying swimmer, irrespective of swimmer type. In navigating the shear-gradient direction, the loadless swimmer exhibits varying swimming velocities, depending on the swimmer type, with pushers achieving the highest velocities, followed by neutral swimmers and pullers. These results align with prior studies that employed fixed translational and rotational velocities [23]. However, for the cargo-towing swimmer, a notable reversal in swimming velocity trends is observed. Pullers now outpace neutral and pusher swimmers, a pattern consistently observed across varying load lengths.

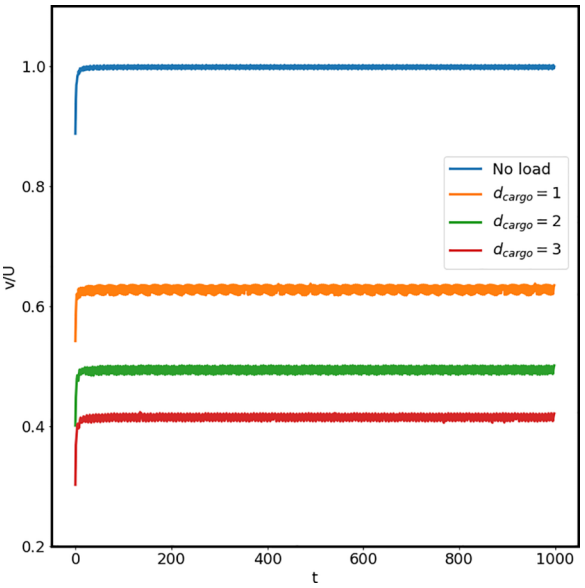

FIG. 6. The normalized swimming velocity of a neutral swimmer with and without the load. The load length is in the range of $d_{\text{cargo}} \in \{1, 2, 3\}$. The swimming velocity is normalized relative to the steady-state swimming velocity of $2/3B_1$.

ACKNOWLEDGMENTS

This work was supported by the Grants-in-Aid for Scientific Research (Japan Society for the Promotion of Science (JSPS) KAKENHI) under Grants No. 20H05619 and No. 23H04508 and the JSPS Core-to-Core Program "Advanced core-to-core network for the physics of self-organizing active matter (Grant No. JPJSCCA20230002)." We acknowledge the Joint Usage/Research Center for Interdisciplinary Large-Scale Information Infrastructures and the High Performance Computing Infrastructure in Japan (Projects No. jh230061 and No. jh240063) for providing the computational resources of the Wisteria/BDEC-01 at the Information Technology Center, The University of Tokyo.

APPENDIX A: DEPENDENCY OF THE SWIMMING VELOCITY ON THE LENGTH OF THE CARGO

In the main text, we discussed the performance of navigation of tasks of a neutral type swimmer traversing in the direction of shear gradients, noting a discernible decrease in performance concomitant with increasing load lengths. The efficacy of navigation in this specific direction is markedly contingent upon the swimmer's velocity. This dependency is visually elucidated in Fig. 6, wherein the swimming velocities of a natural swimmer with and without the load are shown. As a free swimmer (no load), the swimmer attains the steady-state velocity $2/3B_1$, as depicted. When the cargo is loaded to the swimmer, a noticeable reduction in swimming speed ensues. Moreover, as the length of the cargo is incrementally

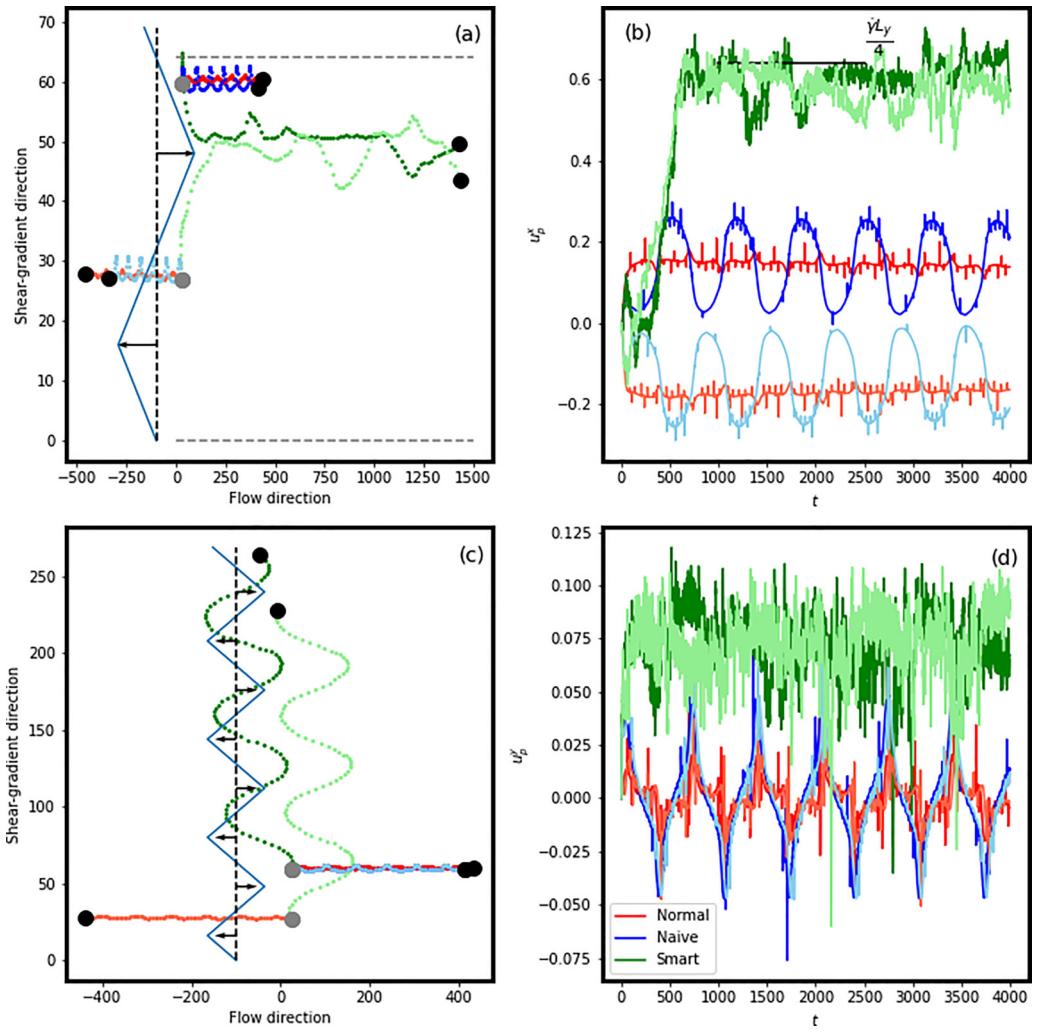

FIG. 7. The velocities and the trajectories of a trained (green and light-green)/naive (blue and light-blue) cargo-towing swimmer, with the load length $d_{\text{cargo}} = 2$, and a normal chain (red and light-red) comprised of three inert beads, starting from different initial positions, observed over a navigation course for the task of swimming in the flow (a), (b) and shear-gradient (c), (d) directions comprising a total of $T_l = 10^4$ action time steps. Here the constant $C_1 = 0.6$ and dynamic $B_1^l = B_1^{l-1} + \Delta B_1$ constrained between $0.02 \leqslant B_1 \leqslant 0.2$ are considered.

augmented, a commensurate decrease in swimming speed is observed. Notably, it is apparent that the swimmer bearing a load with a length of $d_{\text{cargo}} = 3$ manifests the most sluggish swimming velocity among the assessed scenarios.

APPENDIX B: DEPENDENCY OF THE PARTICLE INITIALIZATION

In Fig. 3 of the main text, we discuss the navigation performance of different types of cargo-towing swimmers with the cargo length $d_{\text{cargo}} = 2$: trained, naive, and normal swimmers wherein only the B_1 parameter can change at each action time step in both flow and shear-gradient direction. The initial position of the swimmer is set where the background flow velocity is near zero and positive. Here, we compare this with cases where the swimmers' initial positions are located in regions where the background flow is negative. The results are shown in Fig. 7.

For the task of navigating in the flow direction, the trained (smart) swimmer can navigate towards and swim efficiently along the maximum flow regimes. Although there are deviations from the maximum regime due to imperfect alignment with the flow, the smart swimmer can realign its direction when deviations occur. In contrast, the naive and normal swimmers are dragged towards the negative flow direction with periodic rotation, indicating they are trapped amidst vorticities.

For the task of navigating in the shear-gradient direction, the smart swimmer starting in the negative flow region achieves the same performance as the smart swimmer starting from the zero flow region. Similar to the flow direction task, the naive and normal swimmers are dragged in the negative flow direction, as they lack the ability to align their swimming direction towards the shear-gradient direction.

[1] T. E. Mallouk and A. Sen, Powering nanorobots, Sci. Am. **300**, 72 (2009).

[2] S. Ramaswamy, The mechanics and statistics of active matter, Annu. Rev. Condens. Matter Phys. **1**, 323 (2010).

[3] N. Buss, O. Yasa, Y. Alapan, M. B. Akolpoglu, and M. Sitti, Nanoerythrosome-functionalized biohybrid microswimmers, APL Bioengineering **4**, 026103 (2020).

[4] S. K. Srivastava, M. Medina-Sánchez, B. Koch, and O. G. Schmidt, Medibots: Dual-action biogenic microdaggers for single-cell surgery and drug release, Adv. Mater. **28**, 832 (2016).

[5] C. Richard, J. Simmchen, and A. Eychmüller, Photocatalytic iron oxide micro-swimmers for environmental remediation, Z. Phys. Chem. **232**, 747 (2018).

[6] W. Gao and J. Wang, The environmental impact of micro/nanomachines: A review, ACS Nano **8**, 3170 (2014).

[7] J. Li, B. Esteban-Fernández de Ávila, W. Gao, L. Zhang, and J. Wang, Micro/Nanorobots for biomedicine: Delivery, surgery, sensing, and detoxification, Sci. Robot. **2**, eaam6431 (2017).

[8] X. Ma and S. Sánchez, Self-propelling micro-nanorobots: Challenges and future perspectives in nanomedicine, Nanomedicine **12**, 1363 (2017).

[9] M. Luo, Y. Feng, T. Wang, and J. Guan, Micro-/nanorobots at work in active drug delivery, Adv. Funct. Mater. **28**, 1706100 (2018).

[10] K. K. Dey, X. Zhao, B. M. Tansi, W. J. Méndez-Ortiz, U. M. Córdova-Figueroa, R. Golestanian, and A. Sen, Micromotors powered by enzyme catalysis, Nano Lett. **15**, 8311 (2015).

[11] S. Sánchez, L. Soler, and J. Katuri, Chemically powered micro- and nanomotors, Angew. Chem., Int. Ed. **54**, 1414 (2015).

[12] H. Ceylan, I. C. Yasa, O. Yasa, A. F. Tabak, J. Giltinan, and M. Sitti, 3D-printed biodegradable microswimmer for theranostic cargo delivery and release, ACS Nano **13**, 3353 (2019).

[13] R. Almeda, H. van Someren Gréve, and T. Kiørboe, Behavior is a major determinant of predation risk in zooplankton, Ecosphere **8**, e01668 (2017).

[14] R. A. Epstein, E. Z. Patai, J. B. Julian, and H. J. Spiers, The cognitive map in humans: Spatial navigation and beyond, Nat. Neurosci. **20**, 1504 (2017).

[15] S. Colabrese, K. Gustavsson, A. Celani, and L. Biferale, Flow navigation by smart microswimmers via reinforcement learning, Phys. Rev. Lett. **118**, 158004 (2017).

[16] Z. Zou, Y. Liu, Y.-N. Young, O. S. Pak, and A. C. H. Tsang, Gait switching and targeted navigation of microswimmers via deep reinforcement learning, Commun. Phys. **5**, 158 (2022).

[17] S. Muiños-Landin, A. Fischer, V. Holubec, and F. Cichos, Reinforcement learning with artificial microswimmers, Science Robotics **6**, eabd9285 (2021).

[18] P. Gunnarson, I. Mandralis, G. Novati, P. Koumoutsakos, and J. O. Dabiri, Learning efficient navigation in vortical flow fields, Nat. Commun. **12**, 7143 (2021).

[19] M. Nasiri and B. Liebchen, Reinforcement learning of optimal active particle navigation, New J. Phys. **24**, 073042 (2022).

[20] K. Gustavsson, L. Biferale, A. Celani, and S. Colabrese, Finding efficient swimming strategies in a three-dimensional chaotic flow by reinforcement learning, Eur. Phys. J. E **40**, 110 (2017).

[21] J. Qiu, N. Mousavi, K. Gustavsson, C. Xu, B. Mehlig, and L. Zhao, Navigation of micro-swimmers in steady flow: The importance of symmetries, J. Fluid Mech. **932**, A10 (2022).

[22] J. Qiu, N. Mousavi, L. Zhao, and K. Gustavsson, Active gyrotactic stability of microswimmers using hydromechanical signals, Phys. Rev. Fluids **7**, 014311 (2022).

[23] K. Sankaewtong, J. J. Molina, M. S. Turner, and R. Yamamoto, Learning to swim efficiently in a nonuniform flow field, Phys. Rev. E **107**, 065102 (2023).

[24] K. Sankaewtong, J. J. Molina, and R. Yamamoto, Autonomous navigation of smart microswimmers in non-uniform flow fields, Phys. Fluids **36**, 041902 (2024).

[25] O. Raz and A. M. Leshansky, Efficiency of cargo towing by a microswimmer, Phys. Rev. E **77**, 055305(R) (2008).

[26] W. Gao, D. Kagan, O. S. Pak, C. Clawson, S. Campuzano, E. Chuluun-Erdene, E. Shipton, E. E. Fullerton, L. Zhang, E. Lauga, and J. Wang, Cargo-towing fuel-free magnetic nanoswimmers for targeted drug delivery, Small **8**, 460 (2012).

[27] T. Debnath and P. K. Ghosh, Activated barrier crossing dynamics of a janus particle carrying cargo, Phys. Chem. Chem. Phys. **20**, 25069 (2018).

[28] A. Daddi-Moussa-Ider, M. Lisicki, and A. J. T. M. Mathijssen, Tuning the upstream swimming of microrobots by shape and cargo size, Phys. Rev. Appl. **14**, 024071 (2020).

[29] R. Yamamoto, J. J. Molina, and Y. Nakayama, Smoothed profile method for direct numerical simulations of hydrodynamically interacting particles, Soft Matter **17**, 4226 (2021).

[30] M. P. Allen and G. Germano, Expressions for forces and torques in molecular simulations using rigid bodies, Mol. Phys. **104**, 3225 (2006).

[31] M. J. Lighthill, On the squirming motion of nearly spherical deformable bodies through liquids at very small reynolds numbers, Commun. Pure Appl. Math. **5**, 109 (1952).

[32] J. R. Blake, A spherical envelope approach to ciliary propulsion, J. Fluid Mech. **46**, 199 (1971).

[33] J. Happel and H. Brenner, *Low Reynolds Number Hydrodynamics* (Prentice-Hall, Englewood Cliffs, NJ, 1983).

[34] T. J. Pedley, D. R. Brumley, and R. E. Goldstein, Squirmers with swirl: A model for volvox swimming, J. Fluid Mech. **798**, 165 (2016).

[35] T. Iwashita and R. Yamamoto, Short-time motion of brownian particles in a shear flow, Phys. Rev. E **79**, 031401 (2009).

[36] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (The MIT Press, Cambridge, MA, 2018).

[37] H. van Hasselt, A. Guez, and D. Silver, Deep reinforcement learning with double q-learning, arXiv:1509.06461.

[38] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, Prioritized experience replay, arXiv:1511.05952.

[39] R. S. Sutton, Learning to predict by the methods of temporal differences, Mach. Learn. **3**, 9 (1988).

[40] H. Kobayashi and R. Yamamoto, Tumbling motion of a single chain in shear flow: A crossover from Brownian to non-Brownian behavior, Phys. Rev. E **81**, 041807 (2010).

[41] J. Einarsson, F. Candelier, F. Lundell, J. R. Angilella, and B. Mehlig, Effect of weak fluid inertia upon Jeffery orbits, Phys. Rev. E **91**, 041002(R) (2015).

[42] G. Dulac-Arnold, R. Evans, H. van Hasselt, P. Sunehag, T. Lillicrap, J. Hunt, T. Mann, T. Weber, T. Degris, and B. Coppin, Deep reinforcement learning in large discrete action spaces, arXiv:1512.07679.

[43] L. van der Maaten and G. Hinton, Viualizing data using t-SNE, J. Mach. Learn. Res. **9**, 2579 (2008).

[44] Z. Ouyang, Z. Lin, J. Lin, Z. Yu, and N. Phan-Thien, Cargo carrying with an inertial squirmer in a newtonian fluid, J. Fluid Mech. **959**, A25 (2023).