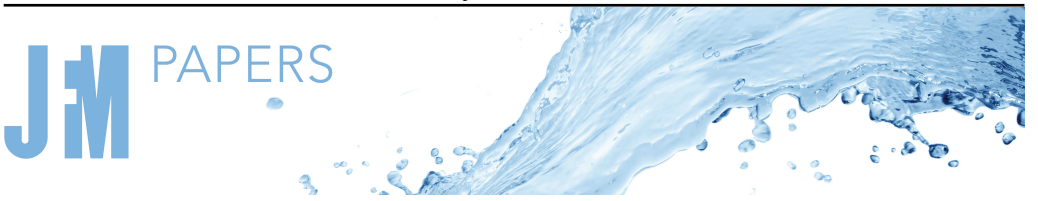
J. Fluid Mech. (2026), *vol.* 1041, A37, doi:10.1017/jfm.2026.11923



Adaptive shape control for microswimmer navigation in turbulence

Jingran Qiu[1], **Lorenzo Piro**[2], **Luca Biferale**[2], **Massimo Cencini**[3,4], **Bernhard Mehlig**[1] **and Kristian Gustavsson**[1]

[1]Department of Physics, Gothenburg University, Origovägen 6 B, Gothenburg 41296, Sweden

[2]Department of Physics & INFN, University of Rome 'Tor Vergata', Via della Ricerca Scientifica 1, Rome 00133, Italy

[3]Istituto dei Sistemi Complessi, CNR, Via dei Taurini 19, Rome 00185, Italy

[4]INFN 'Tor Vergata', Via della Ricerca Scientifica 1, Rome 00133, Italy

Corresponding author: Lorenzo Piro, lorenzo.piro@roma2.infn.it

(Received 6 March 2026 UTC; revised 25 June 2026 UTC; accepted 13 July 2026 UTC)

Navigation in turbulent environments is a fundamental challenge for biological and artificial microswimmers. While most existing studies focus on adapting motility or steering, the role of active morphological changes in navigation remains poorly explored. Here, we investigate a shape-changing spheroidal microswimmer tasked with maximising its displacement from an initial position in two-dimensional turbulent flows. To uncover the underlying mechanisms in a controlled setting, we first train the swimmer in stochastic flows with tuneable temporal correlations and then test the resulting strategies in fully resolved Navier–Stokes turbulence. Using reinforcement learning, the microswimmer learns to adapt its aspect ratio based on its orientation and local velocity-gradient signals. The learned strategies outperform fixed-shape and short-time-optimal baselines across different flow regimes and remain effective when transferred from stochastic flows to fully resolved turbulence. Guided by the learned policies, we propose a minimal analytical model that captures the essential navigation mechanisms and reproduces the performance across flow regimes. These results show that adaptive morphology provides a robust and physically interpretable control paradigm for microswimmer navigation in complex flows.

Key words: machine learning, micro-organism dynamics, turbulent flows

1. Introduction

Navigation and exploration of microswimmers in turbulent and complex flow environments are fundamental tasks in both natural and artificial systems. In nature,

© The Author(s), 2026. Published by Cambridge University Press.

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

for example, phytoplankton and zooplankton live in a turbulent environment, spending significant effort in vertical migration and nutrient foraging (Hays 2003; Guasto *et al.* 2012). Artificial microswimmers, by contrast, have shown promise for biomedical applications such as targeted drug delivery and microsurgery (Huang *et al.* 2019; Tsang *et al.* 2020).

An intriguing scenario arises when microswimmers perceive and adapt to environmental cues. This capacity is common among many plankton species, which can sense and respond to environmental cues such as light (phototaxis) (Jékely *et al.* 2008), chemical gradients (chemotaxis) (Stocker & Seymour 2012) and hydromechanical signals (Kiørboe & Visser 1999). Beyond modulating motility, some plankton also alter their morphology in response to environmental changes. For instance, Sengupta *et al.* (2017) showed that certain algae undergo shape transformations under turbulent conditions, thereby reversing their vertical migration by altering their stable orientation. Similarly, some diatoms form long chains to adjust their effective aspect ratio, thereby modifying their dynamics in turbulence and improving vertical migration efficiency (Smayda 2010; Borgnino *et al.* 2018; Lovecchio *et al.* 2019).

In synthetic systems, artificial microswimmers must adapt to heterogeneous and time-dependent flow environments to achieve autonomous navigation. Despite substantial progress, this objective has not yet been realised. Advances in biohybrid microrobots have demonstrated adaptive and partially autonomous behaviour (Ricotti *et al.* 2017; Huang *et al.* 2019; Tsang *et al.* 2020). For example, Huang *et al.* (2016) developed a microswimmer with deformable flagella that enable controllable swimming gaits. However, fully autonomous microswimmers that operate without external control remain a major challenge, since adaptive capabilities are typically hard coded during fabrication, which limits flexibility and practical applications.

The study of adaptive navigation strategies is therefore important not only for gaining insight into plankton ecology, but also for guiding the design of artificial microswimmers. In both settings, microswimmers must account for the local flow field to navigate efficiently. In recent years, the motion of microswimmers in turbulent environments has been investigated extensively. Modelling them as Lagrangian spheroidal particles swimming in their instantaneous direction explains phenomena such as small-scale clustering (Durham *et al.* 2013; Gustavsson *et al.* 2016) and preferential alignment with flow structures (Zhan *et al.* 2014; Borgnino *et al.* 2019). Particle shape plays a central role in this context: it alters the hydrodynamic torque on the swimmer, modifying preferential orientation (Borgnino *et al.* 2019; Ma *et al.* 2022) and location in the flow (Torney & Neufeld 2007; Gustavsson *et al.* 2016; Borgnino *et al.* 2018). Because the rotational dynamics directly influences the swimming direction and stability, adaptive morphological changes may provide significant advantages for navigation in complex environments.

In this context, reinforcement learning has recently emerged as a powerful tool for discovering optimal strategies in complex environments (Sutton & Barto 2018; Mehlig 2021). Since the pioneering work of Colabrese *et al.* (2017), reinforcement learning has been widely applied to the study of microswimmer navigation strategies (Gustavsson *et al.* 2017; Alageshan *et al.* 2020; Gunnarson *et al.* 2021; Borra *et al.* 2022; Xu *et al.* 2022; Zou *et al.* 2022; Qiu *et al.* 2022*a*, *b*; Mousavi *et al.* 2024, 2025). However, except for Colabrese *et al.* (2018), which considered density-changing microswimmers, most existing studies have focused on adaptive motility in response to environmental cues, such as swimming speed, steering direction or locomotory gaits. In contrast, navigation strategies based on active deformation or morphological adaptation remain largely unexplored.

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

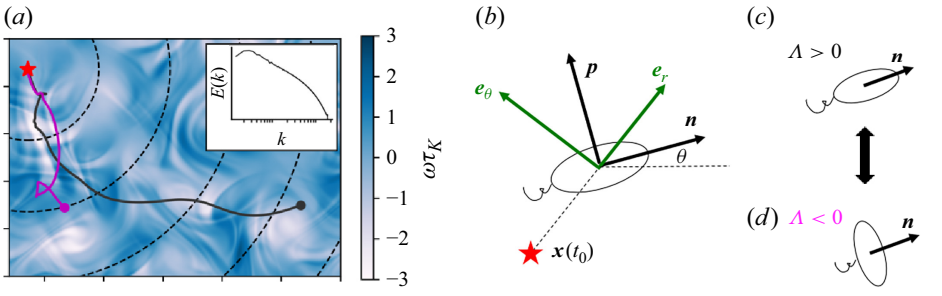


Figure 1. (*a*) Microswimmer navigation maximising distance from the initial position $\boldsymbol{x}(t_0)$ (red star). Equidistant black dashed circles denote radial distances. Two example trajectories of microswimmers following (2.1) with constant shape factors $\Lambda = -0.98$ (magenta) and $\Lambda = 0.98$ (black), evolving in a two-dimensional turbulent flow obtained from DNS. Inset: energy spectrum of the two-dimensional turbulent flow. (*b*) Schematic of a spheroidal microswimmer and its reference frame. (*c, d*) Microswimmers can adjust their aspect ratio, becoming (*c*) prolate, $\Lambda > 0$, or (*d*) oblate, $\Lambda < 0$.

In this study, we use reinforcement learning to investigate the navigation strategy of shape-changing spheroidal microswimmers in two-dimensional stochastic and turbulent flows. We consider the task in which a microswimmer needs to maximise its distance from its initial position over a fixed time horizon (figure 1*a*). This serves as a minimal model for ecologically relevant behaviours such as evasion from harmful regions (Incze *et al.* 2001), and efficient exploration of nutrition-rich environments (Bartumeus *et al.* 2003). The microswimmer is assumed to sense its instantaneous orientation and local flow-velocity gradients, and to adapt its aspect ratio accordingly. While our primary interest is navigation in realistic turbulent flows, we first use a stochastic flow model with adjustable temporal correlations to identify the mechanisms underlying learned strategies, and then assess their robustness and relevance using direct numerical simulations of two-dimensional turbulence.

Our main findings can be summarised as follows: (i) adaptive shape control enhances navigation. The learned strategy consistently outperforms heuristic baselines, with the largest gains in slowly evolving flows. (ii) A minimal analytical model captures the learned behaviour. A reduced interpretable strategy distilled from the reinforcement-learning policy reproduces performance across flow regimes. (iii) The learned adaptation mechanisms are robust and transferable. Policies trained in stochastic flows transfer successfully to turbulent flows and retain high performance under moderate variations of the swimming parameters, indicating that the extracted mechanisms are universal rather than artefacts of a specific flow model.

Taken together, these results show that active shape modulation provides a physically interpretable and efficient control paradigm for microswimmer navigation in complex flows, revealing how morphological adaptation can be leveraged to exploit environmental structure over multiple time scales.

The paper is organised as follows. Section 2 introduces the microswimmer model, stochastic flow model and reinforcement-learning framework. Section 3 defines the benchmark strategies. Section 4 analyses the strategies learned in different flow regimes, compares their performances with the benchmarks, explains the underlying physical mechanisms and examines their transferability to direct numerical simulations (DNS) of turbulence. Finally, § 5 summarises the main findings and outlines future directions.

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

2. Methods

2.1. *Microswimmer model*

We model microswimmers as self-propelled spheroidal particles advected by a fluid flow (figure 1*b*). Each swimmer is characterised by its position $\boldsymbol{x}(t)$ and by a unit orientation vector $\boldsymbol{n}(t)$ aligned with its symmetry axis. The shape is parametrised by the factor $\Lambda = (\lambda^2 - 1)/(\lambda^2 + 1)$, where λ is the spheroid aspect ratio. We assume that microswimmers can actively deform and control Λ, as illustrated in figures 1(*c*) and 1(*d*).

Assuming small microswimmers and negligible inertia, their dynamics is governed by the standard model for elongated particles in viscous flows (Pedley & Kessler 1987; Durham *et al.* 2009; Qiu *et al.* 2022*a*; Mousavi *et al.* 2024)

$$\dot{\boldsymbol{x}} = \boldsymbol{u} + v_s \boldsymbol{n}, \tag{2.1a}$$

$$\dot{\boldsymbol{n}} = \frac{1}{2}\boldsymbol{\omega} \times \boldsymbol{n} + \Lambda\, (\mathbb{I} - \boldsymbol{nn}) \cdot \mathbb{S} \cdot \boldsymbol{n}. \tag{2.1b}$$

Here, $\boldsymbol{u}(\boldsymbol{x}, t)$ denotes the fluid velocity evaluated at the microswimmer position, v_s denotes swimming speed, $\boldsymbol{\omega} = \boldsymbol{\nabla} \times \boldsymbol{u}$ is the vorticity, $\mathbb{S} = (1/2)(\boldsymbol{\nabla u} + \boldsymbol{\nabla u}^T)$ is the strain-rate tensor and $\mathbb{I}$ is the identity matrix.

The translational dynamics in (2.1*a*) consists of passive advection by the ambient flow and active propulsion along the swimmer orientation $\boldsymbol{n}$. The rotational dynamics in (2.1*b*) describes the rotation of a spheroid by velocity gradients in viscous flow (Jeffery 1922). The contribution of the strain-rate tensor $\mathbb{S}$ is proportional to the shape factor Λ. Consequently, particle shape strongly impacts microswimmer dispersion and transport (Khurana & Ouellette 2012; Gustavsson *et al.* 2016; Piro *et al.* 2024). In particular, swimmers with different Λ can exhibit qualitatively distinct trajectories within the same flow field (figure 1*a*), indicating that active modulation of Λ may serve as an effective navigation mechanism.

We focus on a two-dimensional flow in which the microswimmer orientation is confined to the flow plane. In this case, the orientation dynamics in (2.1*b*) simplifies to

$$\dot{\theta} = \frac{1}{2}\omega + \Lambda S_{np}, \tag{2.2}$$

where θ is the angle of $\boldsymbol{n}$, ω is the out-of-plane vorticity, $S_{np} = \boldsymbol{n} \cdot \mathbb{S} \cdot \boldsymbol{p}$ and $\boldsymbol{p}$ is the in-plane unit vector perpendicular to $\boldsymbol{n}$ (figure 1*b*).

2.2. *Stochastic flow model*

To model the turbulent fluctuations, we use a two-dimensional, time-dependent, isotropic stochastic flow model (Bec *et al.* 2024). The velocity field of the flow is determined by $\boldsymbol{u}(\boldsymbol{x}, t) = (\partial\psi/\partial y, -\partial\psi/\partial x)$. Following Mousavi *et al.* (2024), the streamfunction $\psi(\boldsymbol{x}, t)$ is the superposition of M time-independent components $\psi_i(\boldsymbol{x})$, with time-dependent coefficients $c_i(t)$. The components follow the spatial correlation

$$\langle \psi_i(\boldsymbol{x})\psi_j(\boldsymbol{x}')\rangle = \frac{1}{2}\delta_{ij} u_{rms}^2 \ell_f^2 \exp\left[-\frac{(\boldsymbol{x} - \boldsymbol{x}')^2}{2\ell_f^2}\right], \tag{2.3}$$

where $u_{rms} = \sqrt{\langle \boldsymbol{u} \cdot \boldsymbol{u}\rangle}$ is the root-mean-square flow velocity, and ℓ_f is the flow correlation length scale. The time-dependent coefficients $c_i(t)$ follow an Ornstein–Uhlenbeck process with zero mean and time correlation

$$\langle c_i(t)c_j(t')\rangle = \frac{\delta_{ij}}{M} \exp\left(-\frac{|t - t'|}{\tau_f}\right), \tag{2.4}$$

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

where τ_f is the flow correlation time.

The stochastic flow is characterised by a single dimensionless parameter, the Kubo number $Ku = \tau_f u_{rms}/\ell_f$, which compares the flow correlation time τ_f with the advective time scale ℓ_f/u_{rms}. Large Ku corresponds to slowly evolving flows, approaching a time-independent limit as $Ku \to \infty$, while $Ku \to 0$ yields temporally uncorrelated flows with rapid fluctuations. This stochastic flow model reproduces essential statistical properties of small-scale velocity-gradient fluctuations in turbulent flows, showing good agreement with DNS for both inertial particles (Bec *et al.* 2024) and microswimmers (Gustavsson *et al.* 2016; Borgnino *et al.* 2019; Qiu *et al.* 2022*b*; Mousavi *et al.* 2024, 2025), especially in the regime of $Ku \sim 10$.

In this work, we vary the Kubo number by adjusting the flow correlation time τ_f, while keeping the root-mean-square flow velocity u_{rms} and length scales ℓ_f fixed. To match the stochastic flow to turbulence, we take u_{rms} as the characteristic velocity scale and the Kolmogorov time scale, $\tau_K = \langle(\partial u_i/\partial x_j)(\partial u_i/\partial x_j)\rangle^{-1/2}$, as the characteristic flow time scale.

2.3. *Reinforcement learning*

The goal is to maximise the swimmer's displacement from its initial position over a prescribed time horizon. To identify an effective control strategy, we employ proximal policy optimisation (PPO) (Schulman *et al.* 2017), a deep reinforcement-learning algorithm designed for continuous state and action spaces. The policy network maps the swimmer state to a continuous control value corresponding to the shape factor Λ. We assume that the microswimmer senses its state at discrete times separated by an update interval T_u, and updates its shape Λ accordingly. The state is composed of local signals: swimmer orientation, strain-rate $\mathbb{S}$ and vorticity ω.

Due to the radial symmetry about the initial position, we decompose the orientation $\boldsymbol{n}$ in the polar basis spanned by the radial unit vector $\boldsymbol{e}_r = \boldsymbol{x}/|\boldsymbol{x}|$ and the azimuthal unit vector $\boldsymbol{e}_\theta$ (figure 1*b*). The projected components $n_r = \boldsymbol{n}\cdot\boldsymbol{e}_r$ and $n_\theta = \boldsymbol{n}\cdot\boldsymbol{e}_\theta$ are included in the state. The former indicates whether the swimmer is heading away from or toward the origin, while the latter provides information on which direction to rotate.

Consistent with previous studies (Kiørboe & Visser 1999; Qiu *et al.* 2022*a*,*b*; Borra *et al.* 2022), we assume that velocity gradients are measured in the body frame $\boldsymbol{n}$, $\boldsymbol{p}$ (figure 1*b*). In two-dimensional incompressible flow, the strain-rate tensor has two independent components, $S_{nn} = \boldsymbol{n}\cdot\mathbb{S}\cdot\boldsymbol{n}$ and $S_{np} = \boldsymbol{n}\cdot\mathbb{S}\cdot\boldsymbol{p}$, while the vorticity reduces to the scalar ω. These quantities are included in the state because they directly govern the rotational dynamics (2.2), encode the local flow topology and have been shown to be central to efficient microswimmer navigation in diverse settings (Jeffery 1922; Gunnarson *et al.* 2021; Monthiller *et al.* 2022; Qiu *et al.* 2022*b*; Mousavi *et al.* 2024, 2025).

To optimise swimmer displacement from the initial location $\boldsymbol{x}(t_0)$, we introduce the step-wise reward

$$r_i = |\boldsymbol{x}(t_i) - \boldsymbol{x}(t_0)| - |\boldsymbol{x}(t_{i-1}) - \boldsymbol{x}(t_0)|, \tag{2.5}$$

where $\boldsymbol{x}(t_i)$ denotes the microswimmer position at the discrete update time $t_i = i\,T_u$. Since rewards are evaluated at fixed time intervals, maximising the cumulative reward is equivalent to maximising the radial velocity away from $\boldsymbol{x}(t_0)$. Noting that $\sum_{i=1}^{N} r_i = |\boldsymbol{x}(t_N) - \boldsymbol{x}(t_0)|$, we define the performance index as

$$\hat{R} = |\boldsymbol{x}(t_N) - \boldsymbol{x}(t_0)|/R_{max}^{(0)}\,, \tag{2.6}$$

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

Description	Parameter	Value
Swimming speed	v_s	$0.707u_{rms}$
Shape factor	Λ	$-0.98 \leq \Lambda \leq 0.98$
Kubo number	Ku	0.1, 1, 10, 10^7
Signals		n_r, n_θ, S_{nn}, S_{np}, ω
Update interval	T_u	$0.1\tau_K$
Episode duration	T	$200\tau_K$

Table 1. Summary of the simulation parameters and of the states (signals) available to the swimmer.

where $R_{max}^{(0)} \equiv v_s T$ is the distance travelled over the duration of an episode $T = NT_u$ by a swimmer moving in a straight line from $\boldsymbol{x}(t_0)$ at speed v_s in quiescent flow. In our simulations in § 4 we use $v_s = 0.707u_{rms}$ and $T = 200\tau_K$.

Training is performed over a large number of episodes. At the start of each episode, the flow coefficients c_i and the swimmer's initial position and orientation are randomly sampled. The swimming dynamics is then numerically integrated over the episode duration T (see Appendix A) following a policy fixed for one episode. The trajectories collected during each episode are used to update the policy parameters using the PPO algorithm (see Appendix B). Convergence is assumed when the performance index $\hat{R}$ plateaus as the number of episode increases.

The parameters of numerical simulations and training are summarised in table 1. The swimming speed is set to $v_s = 0.707u_{rms}$, corresponding to microswimmers with a moderate propulsion speed compared with the fluid flow. The shape factor Λ is constrained between -0.98 and 0.98, corresponding to aspect ratios between 0.1 and 10. We consider four Kubo numbers, $Ku = 0.1$, 1, 10 and 10^7, covering regimes from rapidly decorrelating to effectively steady flows. The episode duration is fixed to $T = 200\tau_K$, ensuring that initial transients are negligible. This is guaranteed by the fact that the swimmer can travel across a distance much larger than the flow correlation length, $R_{max}^{(0)}/\ell_f = 70.7$ for all Ku, and that the episode is at least ten times longer than the flow correlation time τ_f for time-dependent flows ($Ku \leq 10$). The control-update interval is set to $T_u = 0.1\tau_K$, half of the correlation time of the most rapidly decorrelating flow at $Ku = 0.1$, allowing the microswimmer to respond to temporal variations in the local flow. The corresponding travelling distance over T_u is approximately $0.035\ell_f$.

For each value of the Kubo number Ku, the training procedure is repeated three times using different random seeds. In all cases, the performance of the learned smart strategy increases with the number of episodes and saturates after approximately 5000 episodes, indicating convergence (figure 8). These three independent runs are sufficient to verify that the learning dynamics is insensitive to the specific random seed: the training curves collapse and the converged performances are quantitatively similar. Consequently, we report results from a single representative training for each Ku. The robustness of the smart strategies to stochastic perturbations to the microswimmer dynamics and to variations in training parameters (such as the update interval T_u, the swimming speed v_s and the episode duration T) is documented in Appendix C.

3. Baseline strategies

To assess the quality of the learned policies, we compare their performances with two baseline strategies: one based on optimal control and one based on short-time optimisation.

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

3.1. *Naive strategy*

An optimal-control formulation of the navigation problem is obtained by the cost functional

$$\mathcal{C} = -|\boldsymbol{x}(t) - \boldsymbol{x}(t_0)|^2, \tag{3.1}$$

so that minimising $\mathcal{C}$ is equivalent to maximising the distance from the initial position in the shortest time. This choice is consistent with the problem definition and with the reward in (2.5). Applying Pontryagin's minimum principle (Bryson & Ho 2018) to the associated Hamiltonian shows that the optimal strategy is to maintain a constant shape factor $\Lambda = -1$ (see Appendix D). The control problem can, in fact, be mapped to Zermelo's minimum-time navigation problem (Zermelo 1931) with an infinitely distant target. In this framework, the angular dynamics for $\Lambda = -1$ in (2.1b) coincides with that obtained from Zermelo's formulation (Piro *et al.* 2024). Consequently, provided the initial orientation is appropriately chosen, swimmers with $\Lambda = -1$ reach any admissible target point, including the point farthest from the origin, faster than microswimmers with other values of Λ.

In turbulent and stochastic flows, however, trajectories with $\Lambda = -1$ are extremely sensitive to initial conditions and most become trapped in high-vorticity regions (Biferale *et al.* 2019; Piro *et al.* 2022). Realising the formal optimal-control solution therefore requires fine tuning of the initial orientation for each initial position in a given flow realisation, and the resulting performance is not representative of a generic swimmer. For this reason, our naive baseline is not simply the theoretical optimum $\Lambda = -1$, but the constant-Λ swimmer that performs best on average over many initial conditions and flow realisations, within a physically reasonable range of shapes. To this end, we restrict the shape factor to $\Lambda \in [-0.98, 0.98]$, corresponding to aspect ratios between 0.1 and 10, which already cover very elongated oblate and prolate shapes while remaining compatible with experimentally plausible morphologies. We then perform a parameter sweep over Λ in this interval, both in the stochastic flow and in DNS turbulence, and evaluate the mean distance travelled over ensembles of random initial positions and orientations. In all the considered scenarios, we find that the most elongated microswimmers ($\Lambda = 0.98$) achieve the largest mean distance and the highest mean performance index $\hat{R}$. For this reason, we adopt $\Lambda = 0.98$ as the naive constant-shape baseline in all subsequent analyses.

3.2. *Short-time optimisation (STO) strategy*

A more refined baseline is based on the idea of STO, which has demonstrated its capacity to solve the optimal steering strategy of a microswimmer performing vertical migration (Monthiller *et al.* 2022) or escaping high-strain regions (Mousavi *et al.* 2025). In our set-up, we derive the STO by maximising the reward in (2.5) over a short time interval δt. Consider a microswimmer at position $\boldsymbol{x}_t$ at time t and choose Λ to maximise the incremental reward until $t + \delta t$, i.e. maximise $|\boldsymbol{x}(t + \delta t) - \boldsymbol{x}(0)|$. Shifting the coordinate system so that $\boldsymbol{x}(0) = 0$, we have

$$|\boldsymbol{x}(t + \delta t)| = |\boldsymbol{x}| + \delta t\, \dot{\boldsymbol{x}} \cdot \boldsymbol{e}_r + \frac{1}{2}\delta t^2 \Big[\ddot{\boldsymbol{x}} \cdot \boldsymbol{e}_r + \frac{1}{|\boldsymbol{x}|} \left(\dot{\boldsymbol{x}}^2 - (\dot{\boldsymbol{x}} \cdot \boldsymbol{e}_r)^2 \right) \Big] + O(\delta t^3), \tag{3.2}$$

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

where the right-hand side is evaluated at time t. Using the dynamics in (2.1), the only shape-dependent contribution to (3.2) is contained in $(1/2)\delta t^2 \ddot{\boldsymbol{x}} \cdot \boldsymbol{e}_r$

$$|\boldsymbol{x}(t+\delta t)| = \frac{v_s \Lambda \delta t^2}{2} \boldsymbol{e}_r \cdot (\mathbb{I} - \boldsymbol{nn}) \cdot \mathbb{S}(\boldsymbol{x}(t), t) \cdot \boldsymbol{n} + \Lambda\text{-independent terms} + O(\delta t^3). \tag{3.3}$$

Here, $\boldsymbol{e}_r \cdot (\mathbb{I} - \boldsymbol{nn})$ is the projection of $\boldsymbol{e}_r$ on the subspace perpendicular to $\boldsymbol{n}$. In two spatial dimensions, this projection becomes $(\boldsymbol{e}_r \cdot \boldsymbol{p})\boldsymbol{p} = -n_\theta \boldsymbol{p}$, implying that $|\boldsymbol{x}(t+\delta t)|$ in this short-time expansion is maximised by choosing

$$\Lambda = -\Lambda_{max} \mathrm{sign}(n_\theta(t))\ \mathrm{sign}(S_{np}(\boldsymbol{x}(t), t)). \tag{3.4}$$

The resulting STO strategy is therefore a bang–bang control: the shape factor Λ always takes one of the two extremal admissible values $\pm\Lambda_{max}$. This is consistent with Pontryagin's maximum principle for a control that enters linearly in the cost functional and is bounded, which predicts that optimal controls saturate the bounds whenever possible (Bryson & Ho 2018); here, we obtain this behaviour directly from the short-time perturbative expansion.

In conclusion, STO is achieved by taking $|\Lambda|$ constant over δt and as large as allowed (Λ_{max}), with the sign set by the local strain and microswimmer orientation. If this strategy is followed and updated continuously, the angular dynamics (2.2) becomes

$$\dot{\theta} = \frac{1}{2}\omega(\boldsymbol{x}(t), t) - \Lambda_{max} \mathrm{sign}(n_\theta(t)) |S_{np}(\boldsymbol{x}(t), t)|. \tag{3.5}$$

The mechanism can be understood as follows. Neglecting $\omega(\boldsymbol{x}_t, t)$, the angular dynamics has fixed points at $n_\theta^* = 0$, which corresponds to $\boldsymbol{n}^* = \pm\boldsymbol{e}_r(t)$. The sign chosen in (3.4) is such that $\boldsymbol{n}^* = +\boldsymbol{e}_r(t)$ is stable. This deformation causes the strain to rotate the microswimmer away from its initial position, thereby maximising its radial swimming velocity. In simulations, we set $\Lambda_{max} = 0.98$ and update Λ using the same update interval T_u as in the smart strategy, ensuring a fair comparison of performance.

4. Results

We first compare smart navigation strategies learned via reinforcement learning with naive and STO baselines across flow regimes. We then analyse the smart strategies, identifying the physical mechanisms that govern their performances in rapidly fluctuating ($Ku \to 0$) and steady ($Ku \to \infty$) flows, and how these mechanisms interpolate at intermediate Kubo numbers. From this, we distil a minimal analytical model that captures the essential ingredients of efficient navigation. Finally, we evaluate the smart strategy and the minimal model in two-dimensional DNS of turbulence in the direct enstrophy-cascade regime.

4.1. *Performance comparison across flow regimes*

Figure 2(*a*) shows the maximal normalised performance index $\hat{R}$ versus Kubo number for different control strategies. The STO baseline achieves its highest performance at small Ku values, and $\hat{R}$ decreases steadily as Ku grows. This behaviour arises because the STO strategy focuses on optimising short-term travel distance (see § 3.2). For small Ku, where flow gradients decorrelate rapidly, the STO strategy already captures the essential physics, and we show in § 4.2 that the smart strategy effectively reduces to STO in this limit. As flow persistence increases, however, STO becomes progressively less effective, and in

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

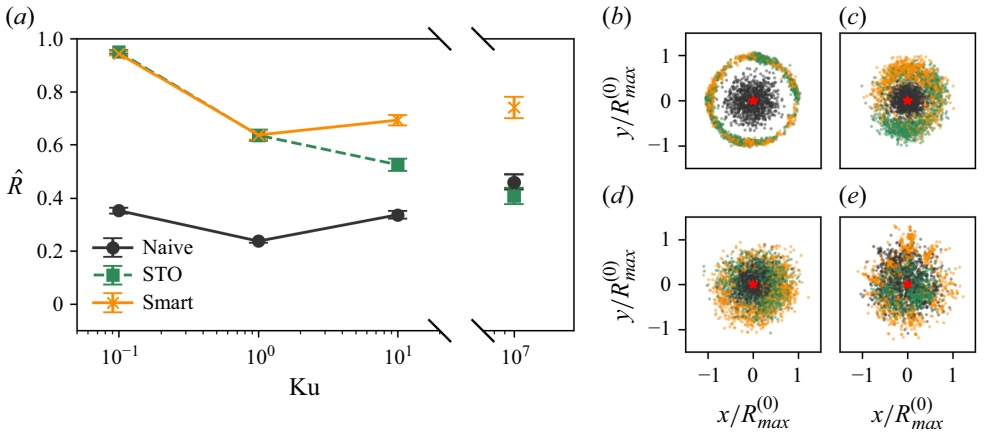


Figure 2. Comparison across flow regimes of reinforcement-learned smart strategy (orange), naive strategy ($\Lambda = 0.98$) (black) and the STO strategy (3.4) (green). (*a*) Normalised performance index $\hat{R}$ (2.6) against Kubo number Ku. Error bars show the standard deviation of $\hat{R}$ over 100 flow realisations with 10^3 trajectories each. (*b–e*) Final positions of 10^3 microswimmers starting from the origin (red star) for (*b*) $Ku = 0.1$, (*c*) $Ku = 1$, (*d*) $Ku = 10$, (*e*) $Ku = 10^7$.

§ 4.3 we demonstrate that the smart strategy in the large-Ku regime relies on qualitatively different mechanisms that exploit long-time correlations in the flow.

The naive (constant shape) baseline shows a non-monotonic dependence on Ku and performs worse than the STO strategy except at very large Ku ($Ku = 10^7$), where the flow is effectively steady, causing the STO strategy to fail, as explained in the previous paragraph. In contrast, the smart strategy obtained via reinforcement leaning based on all signals (n_r, n_θ, S_{nn}, S_{np}, ω) maintains high performance across all values of Ku. For $Ku \leq 1$, it matches the STO performance, indicating little room for improvement in rapidly fluctuating flows where STO is essentially optimal. At larger Ku, however, it significantly outperforms both baselines: at $Ku = 10$, the displacement is approximately 30 % larger than STO, and in the quasi-steady limit ($Ku = 10^7$) the gain exceeds 50 % relative to both STO and the naive strategies. These improvements reflect not only the degradation of STO as Ku increases, but also the fact that the performance of the smart strategy itself grows markedly with Ku. This trend shows that the smart policy actively exploits long-term flow correlations that become stronger at larger Ku.

Figure 2(*b–e*) shows strategy performance via the final positions reached at different Kubo numbers. The trends mirror those of $\hat{R}$: for large Ku, smart microswimmers (orange) travel substantially farther than STO and naive swimmers, whereas for small Ku, STO performs comparably to the smart strategy. Representative trajectories are provided in supplementary movies are available at https://doi.org/10.1017/jfm.2026.11923.

Taken together, these results show that the reinforcement-learning strategy enables faster escape from the initial position, particularly in slowly evolving flows. The non-monotonic dependence of the normalised performance index on Ku, with a minimum around $Ku = 1$, reflects that navigation is most challenging in this intermediate regime, where the flow is neither rapidly decorrelating nor quasi-steady and neither STO nor quasi-steady strategies are fully effective. This observation further suggests that distinct mechanisms underlie efficient navigation in rapidly fluctuating and steady flows. In the following sections, we analyse these mechanisms and their dependence on the flow regime.

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

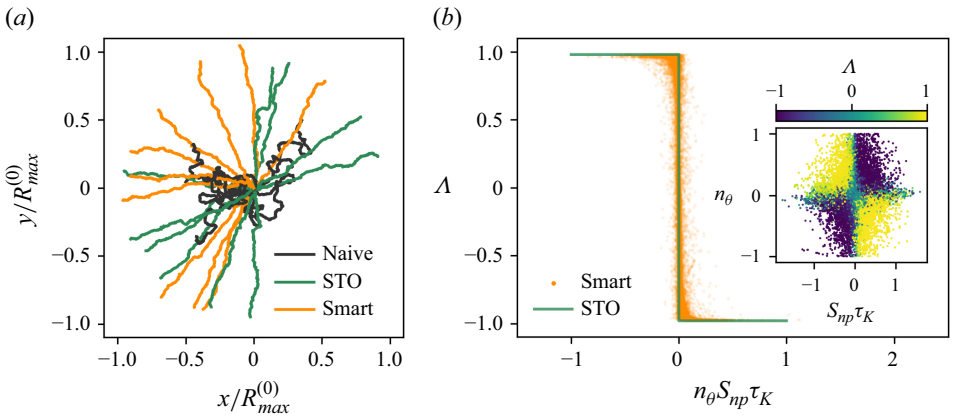


Figure 3. (*a*) Trajectories of microswimmers using different strategies at $Ku = 0.1$: smart (orange), naive ($\Lambda = 0.98$) (black) and STO (3.4) (green). (*b*) Scatter plot of Λ against $n_\theta S_{np}$ for the smart and STO strategies. Inset: Λ against n_θ and S_{np} for the smart strategy, based on 10^3 trajectories.

4.2. *Smart strategy in rapidly fluctuating flow $Ku \to 0$: STO recovery*

In the rapidly fluctuating regime ($Ku = 0.1$), the smart strategy effectively coincides with STO. Figure 2(*a*) shows identical performance, and the trajectories are qualitatively the same: nearly straight paths directed away from the initial position (figure 3*a*). Moreover, the learned control Λ for all signals ($n_r, n_\theta, S_{nn}, S_{np}, \omega$) collapses onto the STO prediction (3.4), when plotted against $n_\theta S_{np}$ (figure 3*b*). Training using only n_θ and S_{np} yields essentially the same behaviour as training with the full state vector, confirming that reinforcement learning recovers STO as the optimal strategy at small Ku (data not shown).

The effectiveness of STO at small Ku follows from the rapid temporal decorrelation of the flow. As velocity gradients fluctuate quickly, long-term planning provides little advantage, and STO approximates the globally optimal strategy. Consequently, the smart strategy does not outperform STO.

In the rapidly fluctuating regime, STO trajectories are nearly straight. The effective STO dynamics in (3.5) breaks symmetry, making swimmers on average align with the radial direction for all values of Ku. However, vorticity causes fluctuations around this direction, and for small Ku it acts as white noise on the otherwise stable orientation. The orientation change over a time interval τ due to vorticity is

$$\Delta\theta = \int_0^\tau \omega \mathrm{d}t. \tag{4.1}$$

In the limit $\tau \gg \tau_f$, this yields the diffusive scaling $(\Delta\theta)^2 \approx 2\omega_{rms}^2 \tau_f \tau$. A shorter correlation time τ_f therefore reduces orientation perturbations over a fixed time interval, helping the microswimmer remain close to the optimal direction. In this regime, STO is sufficient: it captures the essential physics of efficient microswimmer navigation.

4.3. *Smart strategy in steady flow $Ku \to \infty$: correlation among n_r, S_{np} and ω*

In the limit of large Kubo numbers, the smart strategy differs qualitatively from STO and performs substantially better. To clarify the mechanism, we analysed correlations between the learned control Λ and all three-signal combinations, identifying n_r, S_{np} and ω as the dominant contributors to the policy (see also § 4.5 for the corresponding ablation study, which provides quantitative evidence in support of this conclusion).

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

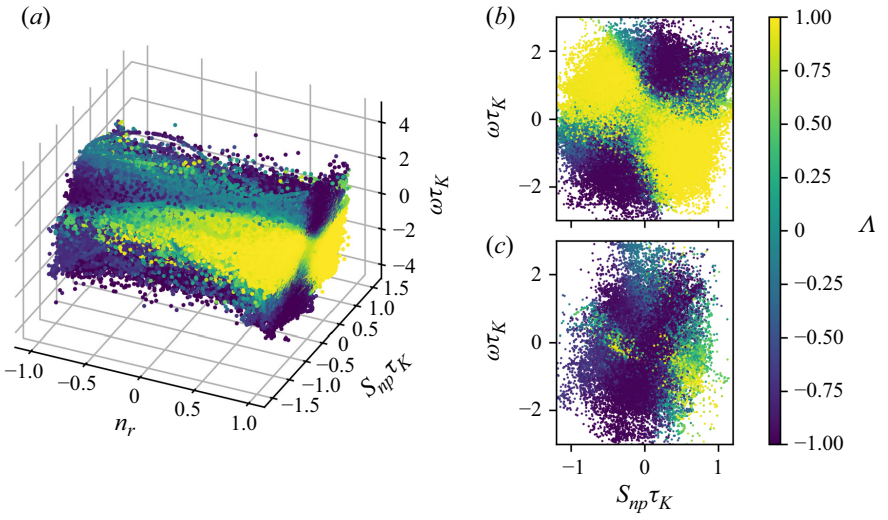


Figure 4. Shape factor Λ (colour) for the smart strategy at $Ku = 10^7$ as a function of (*a*) n_r, S_{np} and ω; (*b*) S_{np} and ω conditioned on $n_r \geq 0.8$; (*c*) S_{np} and ω with $n_r \leq -0.8$. Data are based on 4×10^4 trajectories.

Figure 4(*a*) shows Λ as a function of these variables, revealing non-trivial correlations between Λ and strain/vorticity that depend sensitively on swimmer orientation. For orientations predominantly away from the initial position ($n_r > 0$), the policy organises clearly in the (S_{np}, ω) plane (figure 4*b*): $\Lambda > 0$ when S_{np} and ω have opposite signs or are both small, and $\Lambda < 0$ when they share the same sign. Thus, the smart strategy selects Λ so that the strain-induced rotation term $\Lambda \boldsymbol{p} \cdot \mathbb{S} \cdot \boldsymbol{n}$ counteracts the vorticity contribution in the rotational dynamics (2.2), reducing the net angular velocity. The swimmer thereby stabilises its orientation as it moves away from the initial position, thereby weakening the effect of persistent velocity gradients. This mechanism is absent in short-time-optimal strategies and becomes effective only when the flow has long enough memory.

When the microswimmer is oriented towards the initial location ($n_r < 0$), a different behaviour emerges: the smart policy strongly favours oblate shapes ($\Lambda < 0$), largely independent of the instantaneous values of S_{np} and ω (figure 4*c*). Oblate microswimmers preferentially align against the local flow direction and exhibit self-trapping (Jeffery 1922; Borgnino *et al.* 2019). Here, this mechanism acts as a protective response: by reducing its mobility, the microswimmer suppresses backward motion and avoids being advected back towards the initial position.

In addition to the two dominant mechanisms at large Ku identified in figures 4(*b*) and 4(*c*), we identify a subdominant mechanism involving the second strain component, S_{nn}. In the smart strategy, Λ is weakly negatively correlated with S_{nn}: extreme positive (negative) values of S_{nn} are associated with negative (positive) Λ. For example, among the 1 % most negative values of S_{nn}, the probability of $\Lambda > 0$ is 82 %. Because such events are rare, their contribution to the overall performance is limited. Still, the trend is consistent with prolate and oblate particles known to align with stretching and compressive directions in turbulence (Pumir & Wilkinson 2011; Parsa *et al.* 2012; Gustavsson *et al.* 2014; Zhan *et al.* 2014; Ni *et al.* 2014; Zhao & Andersson 2016). By actively modulating its shape, the smart swimmer disrupts this passive alignment, helping it avoid strongly stretching or compressive regions that would otherwise destabilise its orientation during navigation.

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

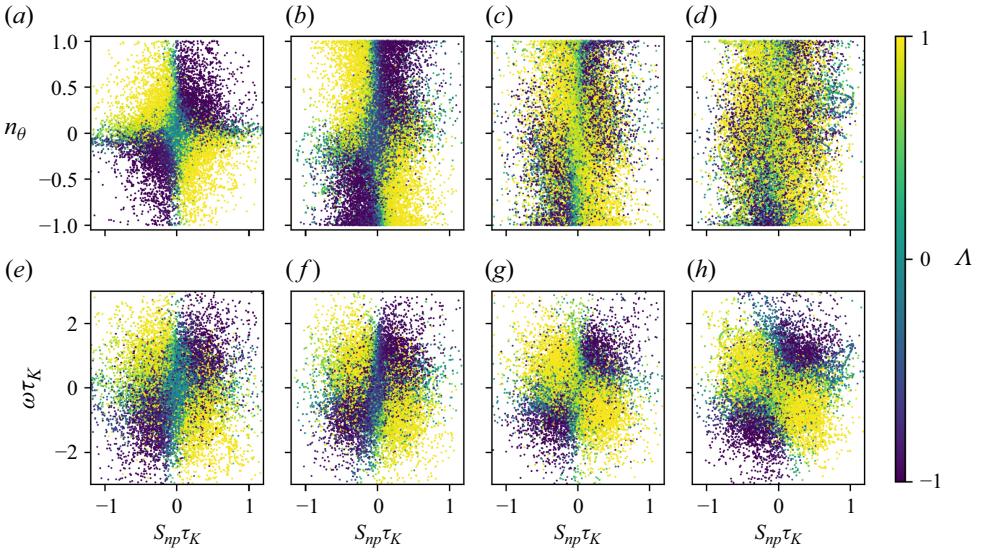


Figure 5. Shape factor Λ (colour) for the smart strategy as a function of S_{np} and n_θ (top row), and S_{np} and Ω (bottom row). Panels show: (*a*, *e*) $Ku = 0.1$; (*b*, *f*) $Ku = 1$; (*c*, *g*) $Ku = 10$; (*d*, *h*) $Ku = 10^7$. The data are based on 10^3 trajectories.

4.4. *Intermediate Ku transition*

At intermediate values of the Kubo number, $Ku = 1$ and 10, the smart strategies smoothly interpolate between the small-Ku and large-Ku limits. This gradual transition is evident in figure 5, which shows the smart control action Λ as a function of the relevant state variables for different values of Ku.

At $Ku = 0.1$, the control landscape of S_{np} and n_θ shows the characteristic dependence associated with the STO strategy (3.4) (figures 3*b* and 5*a*). As Ku increases, this structure progressively weakens and eventually disappears (figure 5*b–d*), reflecting that STO becomes less effective in temporally correlated flows. Accordingly, the smart strategy relies less on this mechanism when the flow decorrelates more slowly.

A complementary trend is observed in the stabilisation mechanism of the quasi-steady regime. The S_{np}–ω correlation, which dominates the smart strategy at $Ku = 10^7$ (figures 4*b* and 5*h*), transitions gradually as Ku decreases (figure 5*e–h*). While this correlation persists even at small Kubo numbers, its physical origin changes. At $Ku = 0.1$, the correlation emerges because the vorticity acts as a perturbation to the orientation n_θ (see § 4.2), which correlates ω with n_θ. Hence, ω acquires a correlation with S_{np} via the n_θ–S_{np} correlation in the STO strategy (3.4). As Ku increases, long-time orientation stabilisation becomes more important as flow correlation time increases.

Remarkably, because the strategies are trained independently at each value of Ku, the observed smooth evolution of control patterns is not imposed by construction but arises naturally from the interaction between the microswimmer and the flow. This continuity suggests that efficient microswimmer navigation is governed by a small set of underlying physical mechanisms whose relative importance is controlled by the flow regime. The smart microswimmer adapts by continuously reweighting these mechanisms in response to the temporal and spatial structure of the flow, rather than switching abruptly between qualitatively distinct strategies.

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

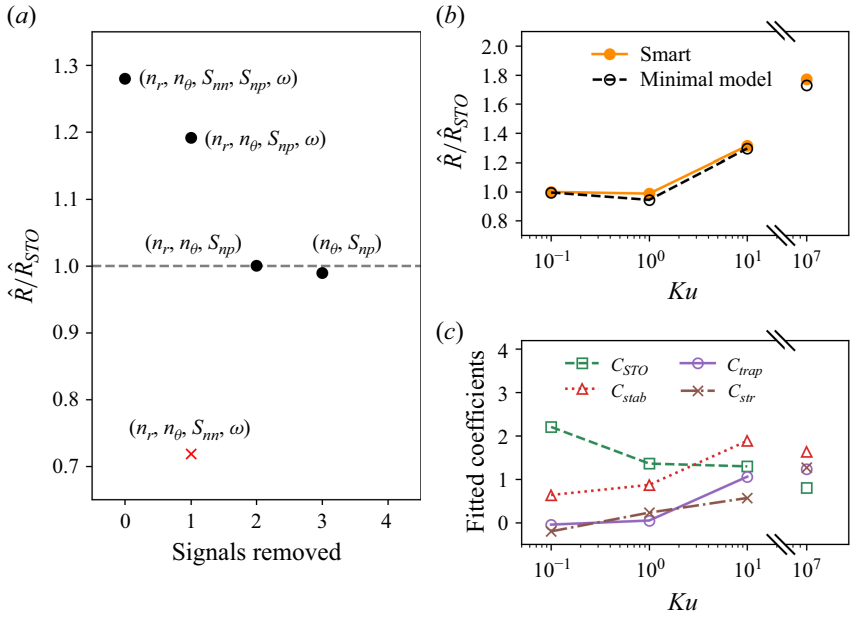


Figure 6. (*a*) Relative performance of smart strategies compared with the STO baseline at $Ku = 10$, trained with different sets of input signals. Brackets denote the signals available to each strategy. (*b*) Performance as a function of *Ku* for the smart strategy and the minimal model (4.2). (*c*) Fitted coefficients of the minimal model as functions of *Ku*.

4.5. *From learned smart strategies to a minimal analytical model of navigation*

In §§ 4.2 and 4.3, we identified four distinct physical mechanisms that underpin efficient navigation in stochastic flows: (i) the STO mechanism, which relies on S_{np} and n_θ and is effective at small *Ku*; (ii) an orientation-stabilisation mechanism, controlled by S_{np} and ω, which dominates at large *Ku*; (iii) a self-trapping mechanism, associated with negative radial orientation $n_r < 0$, also relevant at large *Ku*; and (iv) an extreme strain-avoidance mechanism, governed by S_{nn}, which contributes primarily in the large-*Ku* regime.

To quantify the relative importance of these mechanisms in the smart strategy, we conduct a 'signal-ablation' experiment at $Ku = 10$. In this procedure, the microswimmer is trained multiple times using progressively reduced sets of input signals. By selectively removing the signals associated with specific mechanisms, the corresponding behaviours are suppressed, allowing a direct assessment of their relevance to strategy performance.

Figure 6(*a*) reports the performance of the trained strategies relative to the STO baseline. When all signals are available, the smart strategy outperforms the STO baseline by nearly 30 %. Removing S_{nn} suppresses the extreme strain-avoidance mechanism and leads only to a marginal decrease in performance, consistent with its secondary role (see § 4.3). Further removing ω, such that only n_r, n_θ and S_{np} remain, eliminates the orientation-stabilisation mechanism and causes an abrupt drop in performance to the level of the STO strategy. Excluding n_r as well results in just a mild decrease, since the remaining signals n_θ and S_{np} are sufficient to recover the STO mechanism. Beyond figure 6(*a*), any policy trained on three or fewer signals is outperformed by STO (not shown).

To further assess whether each signal is actively used by the smart swimmer, we evaluated, without retraining, the policy trained at $Ku = 10$ after randomly permutating the values in selected input signals across times and trajectories, thereby disrupting their

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

correlations with the remaining inputs. The resulting performance degradation follows the same hierarchy as in the ablation study, showing that the policy relies on these correlations during navigation. This confirms that the dominant signals (i.e. n_r, n_θ, S_{np} and ω) are not only important for learning but also for execution.

Amongst all inputs, S_{np} emerges as the most critical signal. This is expected because it underlies both the STO and orientation-stabilisation mechanisms, and this is confirmed by the red cross in figure 6(*a*): removing S_{np} while retaining all other signals leads to a failure of the smart strategy. For completeness, we also trained policies using S_{np} as the only input signal and found that their performance remains comparable to the naive baseline across Kubo numbers, confirming that none of the mechanisms identified can operate effectively on S_{np} alone.

Guided by this mechanistic understanding, we propose a minimal analytical model for the smart strategy

$$\Lambda = \Lambda_{max} \tanh\Big[- c_{STO} n_\theta \operatorname{sign}(S_{np}) - c_{stab} H(n_r) \omega \tau_K \operatorname{sign}(S_{np}) + c_{trap} n_r - c_{str} \tanh(S_{nn} \tau_K) \Big], \tag{4.2}$$

where $H(x)$ denotes the Heaviside function and $\{c_{STO}, c_{stab}, c_{trap}, c_{str}\}$ are Ku-dependent coefficients obtained by fitting the model to the learned smart strategies for different Ku using the curve_fit function from the SciPy library (Virtanen *et al.* 2020). The model combines the four identified mechanisms through a weighted sum, while the outer hyperbolic tangent constrains the control output to the admissible range of Λ.

Each term in (4.2) has a clear physical interpretation. The first term represents the STO strategy. The second term corresponds to the orientation-stabilisation mechanism and is activated only when $n_r > 0$ via the Heaviside function. The third term captures the positive correlation between Λ and n_r responsible for the self-trapping mechanism (see figure 4), while the final term models the strain-avoidance response associated with S_{nn}.

The performance of the minimal model is shown in figure 6(*b*). Its efficiency is nearly indistinguishable from that of the full smart strategy across the range of considered Ku values, demonstrating that the model captures the essential physics underlying efficient navigation.

Beyond reproducing performance, the minimal model provides direct insight into how navigation strategies evolve across flow regimes, as the fitted coefficients quantify the relative weight of each mechanism in the composite strategy. Figure 6(*c*) shows the coefficients as functions of Ku. Consistent with the STO mechanism being effective only at small Ku, the magnitude of c_{STO} decreases monotonically with increasing Ku. In contrast, the magnitudes of the coefficients associated with large-Ku mechanisms (c_{stab}, c_{trap} and c_{str}) tend to increase with Ku.

4.6. *Transfer learning in DNS turbulent flows*

We test the robustness of the smart strategies by transferring them to two-dimensional turbulence generated by DNS. The flow consists of an incompressible velocity field, $\boldsymbol{\nabla} \cdot \boldsymbol{u} = 0$, governed by the Navier–Stokes equations

$$\partial_t \boldsymbol{u} + \boldsymbol{u} \cdot \boldsymbol{\nabla} \boldsymbol{u} = -\boldsymbol{\nabla} p + \nu \Delta \boldsymbol{u} - \alpha \boldsymbol{u} + \boldsymbol{F}, \tag{4.3}$$

where p is pressure, ν the viscosity, $\boldsymbol{F}$ a large-scale external forcing that drives an enstrophy cascade at smaller scales (Boffetta & Ecke 2012) and the term $-\alpha \boldsymbol{u}$ represents friction preventing energy accumulation at large scales. Equation (4.3) is integrated using a standard pseudospectral method (see Appendix A). Representative snapshots of the flow

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

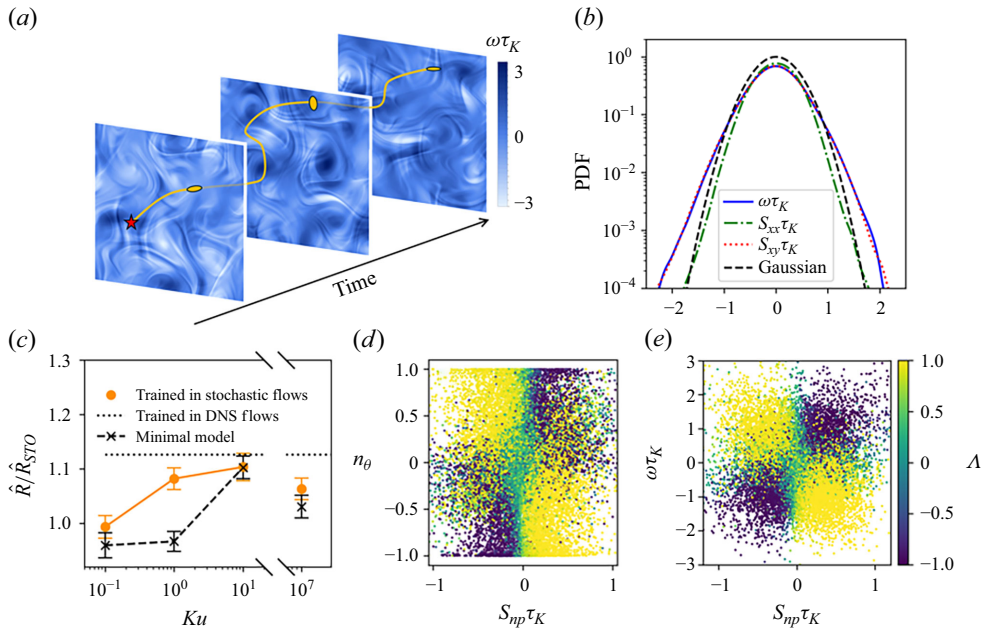


Figure 7. (*a*) Temporal snapshots of the vorticity field of the turbulent flow obtained from DNS. The orange curve is a qualitative illustration of a microswimmer trajectory. The red star indicates its initial position, and the orange spheroids represent its shape and position in the flow field at successive time snapshots. (*b*) Probability distribution function (PDF) of vorticity ω (blue), normal strain S_{xx} (green) and shear strain S_{xy} (red) of the turbulent flow, compared with a Gaussian distribution (black). (*c*) Performance of the strategies relative to the STO baseline (3.4) in two-dimensional turbulence. The orange curve represents the strategy learned in stochastic flows at different Ku. The black dashed curve corresponds to the minimal model (4.2) using the coefficients from figure 6(*c*). The black dotted line shows the performance of the strategy trained in the turbulent flow. Error bars indicate standard deviations over random initial positions and independent flow realisations. (*d*, *e*) Shape factor Λ projected onto (*d*) S_{np} and n_θ; (*e*) S_{np} and ω for the DNS-trained strategy.

vorticity field and the statistical properties of flow-velocity gradients are shown in figures 7(*a*) and 7(*b*).

The orange curve in figure 7(*c*) shows the DNS performance of the smart strategies trained in stochastic flows at different Ku. Performance depends on Ku: the strategy trained at $Ku = 10$ performs best, exceeding the STO baseline and nearly matching a strategy trained in DNS turbulence (black dotted line). This indicates that stochastic flows at $Ku = 10$ capture key dynamical features of DNS turbulence, consistent with earlier observations that large but finite-Ku models reproduce essential Lagrangian statistics of turbulent transport of passive and active particles (Gustavsson *et al.* 2016; Borgnino *et al.* 2019; Bec *et al.* 2024).

Figures 7(*d*) and 7(*e*) show that the (S_{np}, n_θ) and (S_{np}, ω) correlations underlying the STO and orientation-stabilisation mechanisms in stochastic flows, persist in the DNS-trained strategy. Moreover, the close agreement between the minimal model (4.2) (black dashed in figure 7*c*) and the transferred strategy at $Ku = 10$ further confirms that these mechanisms capture the essential physics of efficient navigation and remain robust in realistic turbulence.

Residual differences between the DNS-trained and transferred strategies likely reflect fine-scale structural discrepancies between stochastic and Navier–Stokes turbulence. Away from $Ku = 10$, these mismatches grow and transfer performance degrades. An exception is $Ku = 1$, where the transferred strategy performs unexpectedly well, outperforming

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

both STO and the minimal model. This is because, at intermediate Ku, the smart strategy learns to combine STO with long-time stabilisation, enhancing robustness to the broader range of scales present in DNS turbulence. By contrast, the minimal model retains only the dominant mechanisms identified in the small- and large-Ku limits, and therefore reproduces the smart strategy very well in those regimes and at $Ku = 10$. Around $Ku = 1$, where navigation is intrinsically more difficult, the smart policy relies on more nuanced combinations of the available signals than those captured by the minimal model. Consequently, the minimal model underestimates the performance of the smart strategy and cannot fully reproduce the high transferability observed in two-dimensional turbulence.

5. Conclusions

We have investigated the navigation of shape-changing spheroidal microswimmers tasked with maximising displacement from their initial position. Using reinforcement learning, we identified smart strategies in which the microswimmer adaptively adjusts its shape factor, Λ, in response to local orientation and fluid-gradient signals.

Across flows with different temporal correlations, characterised by the Kubo number Ku, defined as the ratio of flow correlation to advection time scales, the learned smart strategies consistently outperform both a naive fixed-shape baseline and the STO baseline (3.4). Efficient navigation relies upon multiple complementary mechanisms: the STO mechanism, based on n_θ and S_{np}, maximises short-time displacement; orientation stabilisation, driven by S_{np} and vorticity ω, reduces rotation induced by fluid gradients and stabilises the orientation when pointing away from the initial position ($n_r > 0$); adopting an oblate shape when pointing toward the initial location ($n_r < 0$) prevents backward motion; and avoidance of strong normal strain (S_{nn}) prevents destabilising alignment. The relative contribution of these mechanisms varies with the flow's correlation time, as quantified by the minimal model (4.2).

The universality of these mechanisms is confirmed by transferring strategies trained in stochastic flows to fully resolved DNS turbulence. Strategies learned at $Ku = 10$ perform nearly as well in turbulence as those trained directly in the DNS environment, showing that the stochastic flow model effectively captures the essential Lagrangian dynamics and provides a cost-effective framework for exploring navigation strategies and their underlying physics. Beyond quantitative performance, these results highlight a novel navigation paradigm in which adaptive morphology, rather than direct actuation, enables microswimmers to exploit local mechanical cues.

Furthermore, although we allowed the shape factor Λ to span a wide interval in our main analysis, additional tests show that the smart strategy still outperforms naive and STO baselines when Λ is restricted to more moderate, biologically and experimentally plausible ranges, including the purely prolate case $\Lambda \in [0, 0.98]$ (see Appendix C), indicating that the navigation mechanisms identified here do not rely on access to extreme disc- or rod-like shapes.

Future work could explore more complex tasks, such as point-to-point navigation, incorporate energetic costs of active shape control, which may reveal qualitatively different strategies (Piro *et al.* 2024), and investigate hydrodynamic or mechanical interactions among multiple microswimmers, potentially enabling cooperative or adversarial strategies in complex flows (Borra *et al.* 2022). Another future direction is to extend the navigation strategy to three-dimensional flows, where the additional degrees of freedom increase the complexity of the state space and dynamics, posing new challenges for control. It would also be interesting to characterise in which other classes of navigation

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

tasks – such as exploration, or pursuit of distant targets – the STO and long-time stabilisation mechanisms identified here remain effective, thereby clarifying the generality and limits of the proposed control paradigm.

Moreover, the present study focuses on reactive strategies in which the microswimmer uses its instantaneous orientation and local velocity gradients without memory of past states. While such memoryless control is natural in the limits of very small and very large Kubo number, where the flow decorrelates rapidly or is effectively frozen, memory-based strategies (El Khiyati *et al.* 2023; Sankaewtong *et al.* 2024) are expected to become advantageous at intermediate Kubo numbers, where coherent structures exhibit temporal persistence and could be exploited by, e.g. enriching the swimmer's state with temporal derivatives of strain and vorticity. A systematic exploration of such memory-augmented strategies is therefore a promising future direction. Furthermore, while this study employs an idealised spheroidal model, future research may explore alternative adaptive morphologies, such as swimmers with flexible (Tarama *et al.* 2014) or composite (Zou *et al.* 2022) body structures.

Supplementary movies. Supplementary movies are available at https://doi.org/10.1017/jfm.2026.11923.

Acknowledgements. The authors thank A. Celani and N. Mousavi for helpful discussions.

Funding. We acknowledge financial support from Vetenskapsrådet, Grants No. 2018-03974, No. 2023-03617 (J.Q. and K.G.), and No. 2021-4452 (B.M.). K.G., B.M. and J.Q. acknowledge support from the Knut and Alice Wallenberg Foundation, Grant No. 2019.0079. This work was also supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 882340).

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

Appendix A. Numerical simulation

We compute microswimmer trajectories using the Eulerian–Lagrangian scheme of Qiu *et al.* (2022*a*, *b*) and Mousavi *et al.* (2024, 2025). In this scheme, (2.1) is integrated with a second-order method with time step δt. At each step, the fluid velocity and its gradients at the swimmer position are obtained by second-order interpolation from the Eulerian flow field.

The two-dimensional stochastic flow is stored on a uniform 400×400 Eulerian grid with spacing $0.025\ell_f$. As explained in § 2.2, velocity field in (2.3) is generated from $M = 5$ streamfunction components with coefficients evolving via an Ornstein–Uhlenbeck process using the same time step δt. To allow unbounded swimmer motion, the domain is periodically extended in both spatial directions. Our results are insensitive to the domain size, because the flow and its gradients decorrelate before a swimmer traverses the domain. This was confirmed using a ten times larger domain.

Two-dimensional turbulence is obtained by DNS of (4.3) by means of a standard 2/3 de-aliased pseudospectral method on a bi-periodic square domain of size 2π using 512×512 grid points. The flow is forced at large scales by a Gaussian, temporally uncorrelated forcing $\boldsymbol{F}$ peaked at wavenumber $k_f = 3$, in order to induce a direct enstrophy cascade. Friction $-\alpha \boldsymbol{v}$, preventing energy buildup at large scales, is numerically implemented via a hypo-friction term (Boffetta & Ecke 2012). We stored the velocity and gradient fields at approximately every tenth of the Kolmogorov characteristic time. During the simulation of microswimmers, linear interpolation is used to obtain snapshots between successive stored flow fields. The flow velocity and its gradients are then interpolated to the position of microswimmers using the same second-order Lagrangian scheme as in the stochastic

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

flow model. Similar to the stochastic flow, the turbulent flow is also periodically extended, and the domain size is sufficiently large.

Velocity and gradient fields are stored at intervals of approximately one tenth of the Kolmogorov time. During swimmer simulations, intermediate flow snapshots are obtained by linear interpolation in time, and spatial interpolation to swimmer positions uses the same second-order Lagrangian scheme as in the stochastic model. The turbulent flow is also periodically extended, and the domain is sufficiently large to avoid finite-size effects.

In both the stochastic model and DNS turbulence, the microswimmer dynamics is integrated with the time step $\delta t = 0.01\tau_K$. This choice resolves both flow fluctuations and the swimmer dynamics, even for the most rapidly evolving case ($Ku = 0.1$), where the flow correlation time τ_f is still an order of magnitude larger than δt.

Appendix B. Reinforcement-learning algorithm

We use PPO (Schulman *et al.* 2017) to identify smart deformation strategies for a microswimmer tasked with maximising its displacement from its initial position. The PPO algorithm is particularly well suited to the present problem because both the state variables (orientation and local velocity-gradient signals) and the control variable (shape factor Λ) are continuous, whereas many value-based algorithms such as Q-learning are designed for discrete action spaces. Moreover, PPO is widely recognised as a robust on-policy actor–critic method that performs reliably in noisy environments and can be applied to a range of continuous-control tasks without extensive hyperparameter tuning (Sutton & Barto 2018). It optimises a stochastic policy $\pi(\boldsymbol{a}_i|\boldsymbol{s}_i)$, defined as the probability distribution of taking action $\boldsymbol{a}_i$ given state $\boldsymbol{s}_i$ at time $t = iT_u$.

The objective is to approximate the optimal policy that maximises the expected discounted return, $R = \sum_{i=0}^{N} \gamma^i r_i$, where the discount factor $\gamma \in (0, 1)$ balances future and instantaneous rewards. The PPO algorithm achieves this by maximising the clipped surrogate objective function

$$J(\boldsymbol{w}) = \mathbb{E}\left\{\min\left[P_r(\boldsymbol{w})A_i, \mathrm{clip}(P_r(\boldsymbol{w}), 1-\epsilon, 1+\epsilon)A_i\right]\right\}, \tag{B1}$$

where $\boldsymbol{w}$ denotes the parameters of the neural-network, $\mathbb{E}$ is the expectation over the sampled trajectories and $P_r(\boldsymbol{w}) = \pi_{\boldsymbol{w}}(\boldsymbol{a}_i|\boldsymbol{s}_i)/\pi_{\boldsymbol{w}_{old}}(\boldsymbol{a}_i|\boldsymbol{s}_i)$ is the probability ratio between the updated and previous policies. The advantage A_i measures how much the selected action outperforms the expected action under the current policy, which is computed using generalised advantage estimation following the standard implementation (Schulman *et al.* 2017). Maximising $P_r(\boldsymbol{w})A_i$ increases the likelihood of favourable actions, while the clipping and minimising operations restrict excessively large policy updates, thus improving training stability (Schulman *et al.* 2017). For instance, when an action is favourable ($A_i > 0$), the optimisation tends to increase P_r. With the clipping and minimisation, however, $J(\boldsymbol{w})$ is capped at $(1+\epsilon)A_i$, preventing P_r from growing arbitrarily large. Conversely, when an action is unfavoured ($A_i < 0$), the optimisation decreases P_r. If $P_r < 1-\epsilon$, $J(\boldsymbol{w})$ is limited to $(1-\epsilon)A_i$, preventing an excessive reduction of P_r.

The PPO algorithm uses two neural networks: an actor network and a critic network. The actor network takes the state $\boldsymbol{s}_i$ as input and outputs the policy $\pi(\boldsymbol{a}_i|\boldsymbol{s}_i)$. We assume a Gaussian action distribution, with the network outputting the mean and the standard deviation represented by a trainable scalar parameter. During training, actions are sampled from this distribution and then passed through a tanh function to bound the action values. Since this transformation changes the action distribution, the corresponding probability ratio and entropy terms are modified accordingly in the objective function $J(\boldsymbol{w})$. During

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

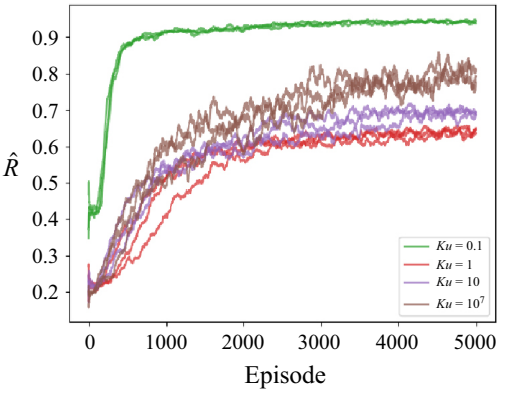


Figure 8. Training curves of the performance index $\hat{R}$ against training episode. For each Ku, three independent runs with different random seeds are shown.

evaluation, actions are sampled in the same way except that the standard deviation is set to zero, such that the actions are deterministically equal to the mean of the distribution. The critic network also takes $\boldsymbol{s}_i$ as input and outputs a scalar state-value estimate $V(\boldsymbol{s}_i)$, which is used in the calculation of the advantage function

$$\begin{aligned} &A_i = \delta_i + (\gamma\lambda_{GAE})\delta_{i+1} + \cdots + (\gamma\lambda_{GAE})^{N-i+1}\delta_{N-1}, \\ &\text{where } \delta_i = r_i + \gamma V(\boldsymbol{s}_{i+1}) - V(\boldsymbol{s}_i), \end{aligned} \tag{B2}$$

and λ_{GAE} is the generalised advantage estimation parameter (see Schulman *et al.* (2017) for details).

Both the actor and critic networks are fully connected feed-forward neural networks with two hidden layers of 64 neurons each, using tanh activation functions. For the actor network, a linear output layer is used for the mean of the Gaussian action distribution. The critic network employs a linear output layer for the value estimate. The network parameters are updated via gradient ascent on $\nabla_{\boldsymbol{w}} J(\boldsymbol{w})$.

The main hyperparameters used in PPO training are selected based on standard choices in the PPO literature (Schulman *et al.* 2017) and simple empirical testing to ensure stable, reproducible convergence. They are summarised as follows. The discount rate is set to $\gamma = 0.99$, the PPO clipping parameter is set to $\epsilon = 0.1$ and the generalised advantage estimation parameter is set to $\lambda_{GAE} = 0.99$. An entropy regularisation term with coefficient 0.005 is included to promote exploration and prevent early convergence of policy. The actor and critic networks are optimised using the Adam optimiser as implemented in PyTorch (Paszke *et al.* 2019). The learning rates decrease linearly from 2×10^{-4} for the actor network and 4×10^{-5} for the critic network to zero over 5000 training episodes. Each policy update comprises four epochs, each over 200 minibatches with a batch size of 25.

The reinforcement-learning algorithm is coupled to the numerical simulation of microswimmer dynamics. In each episode, ten microswimmers are initialised with random initial positions and orientations, as described in § 2.3, and simulated in parallel to collect trajectory data efficiently. At each decision time $t = i\,T_u$, the state variables are computed, and the PPO agent outputs actions that set the shape factors of the microswimmers for the subsequent time interval. The microswimmer trajectories are then advanced in time using the numerical scheme described in Appendix A. This online coupling allows the agent to learn directly from the physically resolved Lagrangian dynamics.

Training is performed for 5000 episodes, and each run uses three independent random seeds to ensure statistical robustness. Figure 8 shows the evolution of the navigation

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

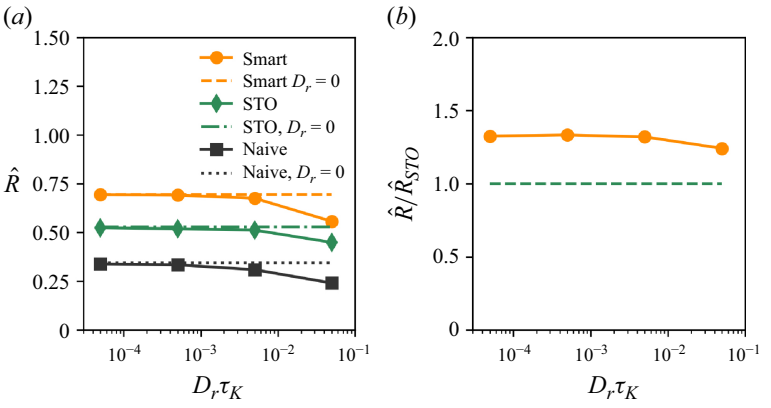


Figure 9. Robustness against rotational noise. (*a*) Performances of smart (orange), naive (black) and STO (green) strategies at different noise levels. The smart strategy is trained at $Ku = 10$, $D_r = 0$. (*b*) Performance of the smart strategy relative to the STO baseline across varying noise levels.

performance as a function of training episode. The performance saturates after a sufficient number of episodes, indicating convergence of the learned strategy. The resulting strategies after training are evaluated on flow realisations that differ from the training data to test their generalisability. All results shown in the paper are obtained from this evaluation process.

Appendix C. Robustness against rotational noise and other parameters

Here, we test the robustness of the smart strategy learned at $Ku = 10$.

First, we consider the robustness against rotational noise. The two-dimensional Brownian noise is implemented as a displacement in $\boldsymbol{n}$ with a variance of $\sqrt{2D_r\delta t}$, where D_r is the rotational diffusion coefficient, and δt is the time step of the integration of the dynamics. Figure 9 shows the performance of the smart, naive, and STO strategies at different noise levels. The performance of the smart strategy begins to decrease when $D_r\tau_K > 5 \times 10^{-3}$, yet it still outperforms both baselines even at the largest noise level considered.

Figure 10(*a*) shows the robustness of the smart strategy against different episode durations T. The smart strategy trained at $T = 200\tau_K$ performs well also at other episode durations, yielding an almost constant normalised performance $\hat{R}$. This indicates that, with the smart strategy (orange curve), the travel distance increases almost linearly with time $|\boldsymbol{x}(t_N) - \boldsymbol{x}(t_0)| \propto T$. Such ballistic-like transport surpasses the diffusive dispersion of naive microswimmers with constant shape whose normalised performance scales as $\hat{R} \propto T^{-1/2}$ (figure 10(*a*), black solid curve). Figure 10(*b*) illustrates the sensitivity of the smart and STO strategies to the update interval T_u. The performance of both strategies decreases as T_u increases, reflecting the growing delay between state measurement and action execution. Since both strategies rely on signals from the local velocity gradients, they fail when T_u becomes comparable to or exceeds the Kolmogorov time scale τ_K, which sets the correlation time of these gradients. Nevertheless, the smart strategy continues to outperform both the STO and naive baselines, whereas the STO strategy approaches the naive one when $T_u \gtrsim \tau_K$.

Furthermore, figure 10(*c*) demonstrates the robustness of strategies with respect to the swimming speed v_s. The smart strategy outperforms both the STO and naive baselines on all the values of v_s. Interestingly, deviating from the specific v_s used during training, either

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

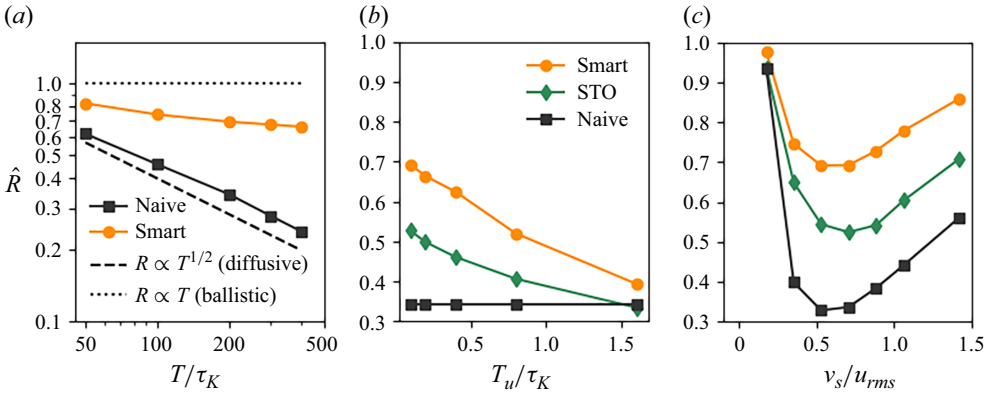


Figure 10. Performance index $\hat{R}$ (2.6) of smart (orange), naive (black) and STO (green) strategies evaluated for different (*a*) episode durations T; (*b*) update intervals T_u; (*c*) swimming speeds v_s. The smart strategy was trained at $Ku = 10$, $T/\tau_K = 200$, $T_u/\tau_K = 0.1$ and $v_s/u_{rms} = 0.707$.

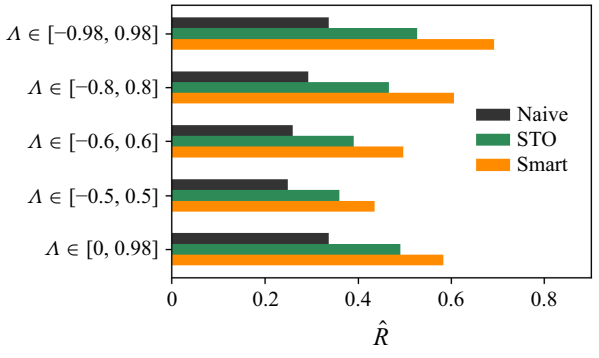


Figure 11. Robustness of smart strategy performance to restricted shape ranges at $Ku = 10$: normalised performance index $\hat{R}$ for naive (black), STO (green) and smart (orange) strategies when the shape factor is constrained to $\Lambda \in [-0.98, 0.98]$, $[-0.8, 0.8]$, $[-0.6, 0.6]$, $[-0.5, 0.5]$ and $[0, 0.98]$.

by increasing or decreasing the speed, results in increased $\hat{R}$. The dependence of $\hat{R}$ on v_s arises from the definition of normalisation $\hat{R} = R/R^{(0)}_{max}$. In the limit as $v_s \to 0$, $\hat{R}$ tends to infinity because the dimensional travel distance R is bounded by the flow advection (scaled by u_{rms}), while the reference $R^{(0)}_{max} = v_s T$ vanishes. In contrast, in the limit of large v_s, the influence of flow becomes negligible, and $\hat{R}$ is expected to converge to unity. Overall, the superior performance of the smart strategy relative to the baselines indicates that its mechanisms are robust to variations in swimming speed.

Finally, to assess the robustness of the navigation mechanisms with respect to realistic constraints on swimmer shape, we evaluated the smart policy trained with $\Lambda \in [-0.98, 0.98]$ under progressively restricted shape ranges, without retraining. For each range, the naive baseline uses a constant shape with $\Lambda = |\Lambda|_{max}$. The STO baseline is computed from (3.4) and then clipped to the admissible interval. In particular, when the action range is restricted to $[0, 0.98]$, the STO strategy sets $\Lambda = 0$ whenever (3.4) yields a negative value. The smart strategy uses the same policy as in the main text, with its output Λ clipped to the corresponding admissible interval. Figure 11 shows the normalised performance $\hat{R}$ for naive (black), STO (green) and smart (orange) strategies when $|\Lambda|_{max}$

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

is constrained to $=0.8, 0.6, 0.5$, corresponding to aspect ratios of approximately 3, 2 and 1.7 for prolate spheroids and 0.33, 0.5 and 0.73 for oblate spheroids, as well as for the purely prolate case $\Lambda \in [0, 0.98]$. In all cases, the smart strategy maintains a clear advantage over both baselines, indicating that the smart navigation mechanisms remain effective even when only moderate shape changes, compatible with biological or artificial microswimmers, are allowed. Moreover, restricting the shape factor to [0, 0, 98] has little effect on the performance of either the smart strategy or the STO strategy. This suggests that allowing the swimmer to approach a spherical shape may already be sufficient to realise the benefits of shape control.

Appendix D. Optimal-control theory

In this appendix, we briefly outline how optimal-control theory leads to the escape strategy discussed in § 3. We consider a spheroidal self–propelled microswimmer whose translational and orientational dynamics in two dimensions are governed by (2.1*a*) and (2.2), which we report here for convenience

$$\dot{\boldsymbol{x}} = \boldsymbol{u} + v_s \boldsymbol{n}, \tag{D1}$$

$$\dot{\theta} = \frac{1}{2}\omega + \Lambda S_{np}, \tag{D2}$$

where $\boldsymbol{n} = (\cos\theta, \sin\theta)$ and the shape factor for spheroids takes the simple form $\Lambda = ((\lambda^2 - 1)/(\lambda^2 + 1))$, with λ the spheroid aspect ratio.

We now aim at finding the optimal protocol for the control λ, or equivalently for the shape factor Λ, that maximises its distance from the initial position $\boldsymbol{x}(t_0)$ over a fixed episode duration T. This is formulated by minimising the cost functional (3.1) and identifying the state vector with $\boldsymbol{q} = (\boldsymbol{x}, \theta)$, where $\boldsymbol{x}$ denotes the swimmer position and θ its orientation.

To apply Pontryagin's minimum principle, we introduce the costate variables $\boldsymbol{\pi}_x$ and π_θ conjugate to $\boldsymbol{x}$ and θ, respectively. The system Hamiltonian reads (Bryson & Ho 2018)

$$\mathcal{H} = \boldsymbol{\pi}_x \cdot (\boldsymbol{u} + v_s \boldsymbol{n}) + \pi_\theta \left(\frac{1}{2}\omega + \Lambda S_{np} \right) . \tag{D3}$$

The necessary condition for optimality is $\partial_\lambda \mathcal{H} = 0$, which yields

$$\frac{\partial \mathcal{H}}{\partial \lambda} = \pi_\theta S_{np} \frac{\mathrm{d}\Lambda}{\mathrm{d}\lambda} = \pi_\theta S_{np} \frac{4\lambda}{(\lambda^2 + 1)^2} = 0 \implies \lambda = 0 \iff \Lambda = -1 . \tag{D4}$$

The optimal escape strategy, therefore, consists of maintaining a constant disk-like shape. In this limit, the flow-induced torque in Jeffery's equation has the same torque structure arising in Zermelo's minimum-time navigation problem (see discussion in § 3 and Piro *et al.* (2024)).

References

Alageshan, J.K., Verma, A.K., Bec, J. & Pandit, R. 2020 Machine learning strategies for path-planning microswimmers in turbulent flows. *Phys. Rev. E* **101** (4), 043110.

Bartumeus, F., Peters, F., Pueyo, S., Marrasé, C. & Catalan, J. 2003 Helical Lévy walks: adjusting searching statistics to resource availability in microzooplankton. *Proc. Natl Acad. Sci.* **100** (22), 12771–12775.

Bec, J., Gustavsson, K. & Mehlig, B. 2024 Statistical models for the dynamics of heavy particles in turbulence. *Annu. Rev. Fluid Mech.* **56**, 189–213.

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

BIFERALE, L., BONACCORSO, F., BUZZICOTTI, M., LEONI, P.C.D. & GUSTAVSSON, K. 2019 Zermelo's problem: optimal point-to-point navigation in 2D turbulent flows using reinforcement learning. *Chaos* **29**, 103138.

BOFFETTA, G. & ECKE, R.E. 2012 Two-dimensional turbulence. *Annu. Rev. Fluid Mech.* **44** (2012), 427–451.

BORGNINO, M., BOFFETTA, G., DE LILLO, F. & CENCINI, M. 2018 Gyrotactic swimmers in turbulence: shape effects and role of the large-scale flow. *J. Fluid Mech.* **856**, R1.

BORGNINO, M., GUSTAVSSON, K., DE LILLO, F., BOFFETTA, G., CENCINI, M. & MEHLIG, B. 2019 Alignment of nonspherical active particles in chaotic flows. *Phys. Rev. Lett.* **123** (13), 138003.

BORRA, F., BIFERALE, L., CENCINI, M. & CELANI, A. 2022 Reinforcement learning for pursuit and evasion of microswimmers at low Reynolds number. *Phys. Rev. Fluids* **7**, 023103.

BRYSON, A.E. & HO, Y.C. 2018 *Applied Optimal Control: Optimization, Estimation, and Control*. Routledge.

COLABRESE, S., GUSTAVSSON, K., CELANI, A. & BIFERALE, L. 2017 Flow navigation by smart microswimmers via reinforcement learning. *Phys. Rev. Lett.* **118** (15), 158004.

COLABRESE, S., GUSTAVSSON, K., CELANI, A. & BIFERALE, L. 2018 Smart inertial particles. *Phys. Rev. Fluids* **3** (8), 084301.

DURHAM, W.M., CLIMENT, E., BARRY, M., DE LILLO, F., BOFFETTA, G., CENCINI, M. & STOCKER, R. 2013 Turbulence drives microscale patches of motile phytoplankton. *Nat. Commun.* **4** (1), 2148.

DURHAM, W.M., KESSLER, J.O. & STOCKER, R. 2009 Disruption of vertical motility by shear triggers formation of thin phytoplankton layers. *Science* **323** (5917), 1067–1070.

EL KHIYATI, Z., CHESNEAUX, R., GIRALDI, L. & J., BEC 2023 Steering undulatory micro-swimmers in a fluid flow through reinforcement learning. *Eur. Phys. J. E* **46** (6), 43.

GUASTO, J.S., RUSCONI, R. & STOCKER, R. 2012 Fluid mechanics of planktonic microorganisms. *Annu. Rev. Fluid Mech.* **44**, 373–400.

GUNNARSON, P., MANDRALIS, I., NOVATI, G., KOUMOUTSAKOS, P. & DABIRI, J.O. 2021 Learning efficient navigation in vortical flow fields. *Nat. Commun.* **12** (1), 7143.

GUSTAVSSON, K., BERGLUND, F., JONSSON, P.R. & MEHLIG, B. 2016 Preferential sampling and small-scale clustering of gyrotactic microswimmers in turbulence. *Phys. Rev. Lett.* **116** (10), 108104.

GUSTAVSSON, K., BIFERALE, L., CELANI, A. & COLABRESE, S. 2017 Finding efficient swimming strategies in a three-dimensional chaotic flow by reinforcement learning. *Eur. Phys. J. E* **40** (12), 110.

GUSTAVSSON, K., EINARSSON, J. & MEHLIG, B. 2014 Tumbling of small axisymmetric particles in random and turbulent flows. *Phys. Rev. Lett.* **112** (1), 014501.

HAYS, G.C. 2003 A review of the adaptive significance and ecosystem consequences of zooplankton diel vertical migrations. *Hydrobiologia* **503** (1), 163–170.

HUANG, H., SAKAR, M.S., PETRUSKA, A.J., PANÉ, S. & NELSON, B.J. 2016 Soft micromachines with programmable motility and morphology. *Nat. Commun.* **7** (1), 12263.

HUANG, H., USLU, F.E., KATSAMBA, P., LAUGA, E., SAKAR, M.S. & NELSON, B.J. 2019 Adaptive locomotion of artificial microswimmers. *Sci. Adv.* **5** (1), eaau1532.

INCZE, L.S., HEBERT, D., WOLFF, N., OAKEY, N. & DYE, D. 2001 Changes in copepod distributions associated with increased turbulence from wind stress. *Mar. Ecol. Prog. Ser.* **213**, 229–240.

JEFFERY, G.B. 1922 The motion of ellipsoidal particles immersed in a viscous fluid. *Proc. R. Soc. A* **102** (715), 161–179.

JÉKELY, G., COLOMBELLI, J., HAUSEN, H., GUY, K., STELZER, E., NÉDÉLEC, F.C & ARENDT, D. 2008 Mechanism of phototaxis in marine zooplankton. *Nature* **456** (7220), 395–399.

KHURANA, N. & OUELLETTE, N.T. 2012 Interactions between active particles and dynamical structures in chaotic flow. *Phys. Fluids* **24** (9), 091902.

KIØRBOE, T. & VISSER, A.W. 1999 Predator and prey perception in copepods due to hydromechanical signals. *Mar. Ecol. Prog. Ser.* **179**, 81–95.

LOVECCHIO, S., CLIMENT, E., STOCKER, R. & DURHAM, W.M. 2019 Chain formation can enhance the vertical migration of phytoplankton through turbulence. *Sci. Adv.* **5** (10), eaaw7879.

MA, K., PUJARA, N. & THIFFEAULT, J.-L. 2022 Reaching for the surface: spheroidal microswimmers in surface gravity waves. *Phys. Rev. Fluids* **7** (1), 014310.

MEHLIG, B. 2021 *Machine Learning with Neural Networks: An Introduction for Scientists and Engineers*. Cambridge University Press.

MONTHILLER, R., LOISY, A., KOEHL, M.A.R., FAVIER, B. & ELOY, C. 2022 Surfing on turbulence: a strategy for planktonic navigation. *Phys. Rev. Lett.* **129** (6), 064502.

MOUSAVI, N., QIU, J., MEHLIG, B., ZHAO, L. & GUSTAVSSON, K. 2024 Efficient survival strategy for zooplankton in turbulence. *Phys. Rev. Res.* **6** (2), L022034.

MOUSAVI, N., QIU, J., ZHAO, L., MEHLIG, B. & GUSTAVSSON, K. 2025 Short term vs. long term: optimization of microswimmer navigation on different time horizons. *Phys. Rev. Res.* **7** (1), 013258.

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press

Ni, R., Ouellette, N.T. & Voth, G.A. 2014 Alignment of vorticity and rods with Lagrangian fluid stretching in turbulence. *J. Fluid Mech.* **743**, R3.
Parsa, S., Calzavarini, E., Toschi, F. & Voth, A.G. 2012 Rotation rate of rods in turbulent fluid flow. *Phys. Rev. Lett.* **109** (13), 134501.
Paszke, A., *et al.* 2019 PyTorch: an imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, vol. 32. Curran Associates, Inc.
Pedley, T.J. & Kessler, J.O. 1987 The orientation of spheroidal microorganisms swimming in a flow field. *Proc. R. Soc. B* **231** (1262), 47–70.
Piro, L., Golestanian, R. & Mahault, B. 2022 Efficiency of navigation strategies for active particles in rugged landscapes. *Front. Phys.* **10**, 1034267.
Piro, L., Vilfan, A., Golestanian, R. & Mahault, B.îT 2024 Energetic cost of microswimmer navigation: the role of body shape. *Phys. Rev. Res.* **6**, 013274.
Pumir, A. & Wilkinson, M. 2011 Orientation statistics of small particles in turbulence. *New J. Phys.* **13** (9), 093030.
Qiu, J., Mousavi, N., Gustavsson, K., Xu, C., Mehlig, B. & Zhao, L. 2022*a* Navigation of microswimmers in steady flow: the importance of symmetries. *J. Fluid Mech.* **932**, A10.
Qiu, J., Mousavi, N., Zhao, L. & Gustavsson, K. 2022*b* Active gyrotactic stability of microswimmers using hydromechanical signals. *Phys. Rev. Fluids* **7** (1), 014311.
Ricotti, L., Trimmer, B., Feinberg, A.W., Raman, R., Parker, K.K., Bashir, R., Sitti, M., Martel, S., Dario, P. & Menciassi, A. 2017 Biohybrid actuators for robotics: a review of devices actuated by living cells. *Sci. Robot.* **2** (12), eaaq0495.
Sankaewtong, K., Molina, J.J. & Yamamoto, R. 2024 Efficient navigation of cargo-towing microswimmer in non-uniform flow fields. *Phys. Rev. Res.* **6**, 033305.
Schulman, J., Wolski, F., Dhariwal, P., Radford, A. & Klimov, O. 2017, Proximal policy optimization algorithms. arXiv: 1707.06347.
Sengupta, A., Carrara, F. & Stocker, R. 2017 Phytoplankton can actively diversify their migration strategy in response to turbulent cues. *Nature* **543**, 555–558.
Smayda, T.J. 2010 Adaptations and selection of harmful and other dinoflagellate species in upwelling systems. 2. Motility and migratory behaviour. *Prog. Oceanogr.* **85** (1–2), 71–91.
Stocker, R. & Seymour, J.R. 2012 Ecology and physics of bacterial chemotaxis in the ocean. *Microbiol. Mol. Biol. Rev.* **76** (4), 792–812.
Sutton, R.S. & Barto, A.G. 2018 *Reinforcement Learning, Second Edition: An Introduction*. MIT Press.
Tarama, M., Menzel, A.M. & Löwen, H. 2014 Deformable microswimmer in a swirl: capturing and scattering dynamics. *Phys. Rev. E* **90** (3), 032907.
Torney, C. & Neufeld, Z. Zoltán 2007 transport and aggregation of self-propelled particles in fluid flows. *Phys. Rev. Lett.* **99** (7), 078101.
Tsang, A.C.H., Demir, E., Ding, Y. & Pak, O.S. 2020 Roads to smart artificial microswimmers. *Adv. Intell. Syst.* **2** (8), 1900137.
Virtanen, P., *et al.* 2020 SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat. Methods* **17**, 261–272.
Xu, A., Wu, H. & Xi, H. 2022 Migration of self-propelling agent in a turbulent environment with minimal energy consumption. *Phys. Fluids* **34** (3), 035117.
Zermelo, E. 1931 Über das Navigationsproblem bei ruhender oder veränderlicher Windverteilung. *ZAMM – J. Appl. Math. Mech.* **11** (2), 114–124.
Zhan, C., Sardina, G., Lushi, E. & Brandt, L. 2014 Accumulation of motile elongated micro-organisms in turbulence. *J. Fluid Mech.* **739**, 22–36.
Zhao, L. & Andersson, H.I. 2016 Why spheroids orient preferentially in near-wall turbulence. *J. Fluid Mech.* **807**, 221–234.
Zou, Z., Liu, Y., Young, Y., Pak, O.S. & Tsang, A.C.H. 2022 Gait switching and targeted navigation of microswimmers via deep reinforcement learning. *Commun. Phys.* **5** (1), 158.

https://doi.org/10.1017/jfm.2026.11923 Published online by Cambridge University Press