

Reinforcement learning of a multi-link swimmer at low Reynolds numbers

Cite as: Phys. Fluids **35**, 032003 (2023); doi: [10.1063/5.0140662](https://doi.org/10.1063/5.0140662)

Submitted: 29 December 2022 · Accepted: 11 February 2023 ·

Published Online: 1 March 2023



View Online



Export Citation



CrossMark

Ke Qin,^{1,2}  Zonghao Zou,^{1,3}  Lailai Zhu,⁴  and On Shun Pak^{1,a)} 

AFFILIATIONS

¹Department of Mechanical Engineering, Santa Clara University, Santa Clara, California 95053, USA

²Department of Energy Science and Engineering, Stanford University, Stanford, California 94305, USA

³Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, New York 14850, USA

⁴Department of Mechanical Engineering, National University of Singapore, 117575, Singapore

^{a)} Author to whom correspondence should be addressed: opak@scu.edu

ABSTRACT

The use of machine learning techniques in the development of microscopic swimmers has drawn considerable attention in recent years. In particular, reinforcement learning has been shown useful in enabling swimmers to learn effective propulsion strategies through its interactions with the surroundings. In this work, we apply a reinforcement learning approach to identify swimming gaits of a multi-link model swimmer. The swimmer consists of multiple rigid links connected serially with hinges, which can rotate freely to change the relative angles between neighboring links. Purcell [“Life at low Reynolds number,” *Am. J. Phys.* **45**, 3 (1977)] demonstrated how the particular case of a three-link swimmer (now known as Purcell’s swimmer) can perform a prescribed sequence of hinge rotation to generate self-propulsion in the absence of inertia. Here, without relying on any prior knowledge of low-Reynolds-number locomotion, we first demonstrate the use of reinforcement learning in identifying the classical swimming gaits of Purcell’s swimmer for case of three links. We next examine the new swimming gaits acquired by the learning process as the number of links increases. We also consider the scenarios when only a single hinge is allowed to rotate at a time and when simultaneous rotation of multiple hinges is allowed. We contrast the difference in the locomotory gaits learned by the swimmers in these scenarios and discuss their propulsion performance. Taken together, our results demonstrate how a simple reinforcement learning technique can be applied to identify both classical and new swimming gaits at low Reynolds numbers.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0140662>

I. INTRODUCTION

The hydrodynamics of swimming microorganisms has attracted considerable attention in the past decades.^{1–5} Microorganisms typically swim at low Reynolds numbers, where the viscous forces dominate the inertial forces. In such viscosity-dominated flows, common swimming strategies at high Reynolds numbers can become largely ineffective.^{4,6,7} In particular, reciprocal swimming gaits, such as the opening and closing of the shell of a scallop, would produce no net locomotion in the absence of inertia, which is now known as Purcell’s scallop theorem.⁸ Nature has evolved different strategies for microorganisms to swim effectively at low Reynolds numbers: A swimming spermatozoon undulates its flagellum that propagates a bending wave;² some bacterial cells gain thrust by rotating a bundle of flagella that resemble a helical wave;⁹ some euglenids swim by continuous body deformations.¹⁰

There has been a growing interest in developing artificial microswimmers that can self-propel like microorganisms for biomedical

applications.^{11–14} The design of locomotory gaits has been a fundamental challenge due to the stringent constraints imposed by the physics of swimming at the microscale. Purcell pioneered the design of microswimmers by his three-link swimmer,⁸ which consists of three rigid links connected by two hinges. The swimmer is free to rotate its hinges to adjust the angles made between two neighboring links. Purcell demonstrated how the swimmer can undergo a sequence of configurational changes to escape from constraints due to kinematic reversibility and generate self-propulsion in the absence of inertia. Since then, the model has been commonly used to explore fundamental aspects of self-propulsion at low Reynolds numbers¹⁵ including optimality and controllability^{16–19} and the generalization to multi-link models.^{20–23} More recently a similar three-link swimmer has also been employed as a model for fish swimming in a potential flow environment.²⁴

The design of microswimmers typically relies on both ingenuity and understanding of the physics of locomotion. Recent works have

relative importance of the immediate reward r_n and the maximum future reward in the next state $\max_{a_{n+1}} Q(s_{n+1}, a_{n+1})$. For a small (large) value of γ , the swimmer is shortsighted (farsighted) and tends to focus more on the immediate (future) reward. We also incorporated an ε -greedy selection scheme⁴⁵ to allow a small probability ε to take a random action against the advice by the Q matrix, allowing the swimmer to explore new solutions without being trapped in locally optimal solutions.

In Sec. III, we report the propulsion strategies of the multi-link swimmer identified by Q -learning. In this work, we non-dimensionalize lengths by L , time by $1/\omega$, and forces by $L^2 \xi_\perp \omega$. Hereafter, we will only refer to dimensionless variables unless otherwise stated. In the implementation of Q -learning, we set $\alpha = 1$ to maximize the learning speed in a deterministic system, $\gamma = 0.95$, to emphasize the future rewards, and $\varepsilon = 0.05$ for exploration.

III. RESULTS AND DISCUSSION

A. Three-link swimmer

We first consider the simplest setup, a three-link swimmer ($N = 3$), which has the minimal degrees of freedom for self-propulsion at a vanishing Reynolds number. Similar to Purcell's swimmer,⁸ we allow only one of the hinges to rotate (i.e., increasing or decreasing one of the angles Φ_i) in each action step. For a three-link swimmer, the amplitude of each hinge rotation is set to be $\hat{\Phi} = \pi/3$. However, unlike Purcell's swimmer, we do not prescribe any sequence of actuation. Instead, the three-link model progressively learns how to swim based on its interaction with the surrounding viscous fluid environment via reinforcement learning. Figure 2(b) displays a typical learning process: The swimmer initially struggles to find effective swimming gaits as illustrated in Fig. 2(c). Nevertheless, after gaining sufficient experience interacting with the environment, the swimmer learns an effective swimming policy and consistently performs the same sequence of action shown in Fig. 2(d) to generate a net translation. It is noteworthy that without any prior knowledge of low-Reynolds-number locomotion, the policy identified by the swimmer

via reinforcement learning recovers the classical gaits of a Purcell's three-link swimmer.⁸

B. N -link swimmer

Next, we extend the studies by applying the same reinforcement learning approach to identify swimming gaits for swimmers with an increased number of links ($N > 3$). To keep the total length of the swimmer constant, the length of each link is given by L/N . In order for swimmers with different N to have comparable geometric configurations (see first row of Fig. 3), the amplitude of hinge rotation is also adjusted accordingly as $\hat{\Phi} = 2\pi/[3(N-1)]$.

As the number of links increases, the swimmers display more complex dynamics due to the increased degrees of freedom. While a three-link swimmer can only generate net displacements horizontally, swimmers with more links translate both vertically and horizontally in the learning process, although the reward is only based on the horizontal displacement. This feature of the system introduces variability into the learning process when the number of links is greater than three ($N > 3$). Unlike the case of a three-link swimmer, where only one effective swimming policy is possible, multiple swimming policies emerge when $N > 3$. For each value of N , we performed 100 trials and display in Fig. 3 the swimming policies most frequently identified by reinforcement learning. It is noteworthy that these swimming gaits share a common feature: They all propagate a traveling wave of hinge actuation in a direction opposite to the net translation. This feature is illustrated by the red dots in Fig. 3, which indicates the particular hinge that has been rotated relative to the previous action step. We can see in Fig. 3 that for each value of N , the red dot first travels from the left to the right, undergoing a sequential hinge rotation with a reduction in the relative angles Φ_i , followed by another sequential hinge rotation but with an increase in the relative angles Φ_i (see also integral multi-media of Fig. 3 for animations).

To summarize, the strategies identified by reinforcement learning here generalize the special case of $N = 3$ (Purcell's swimmer⁸) to a class of swimming gaits that propagate a traveling wave of actuation

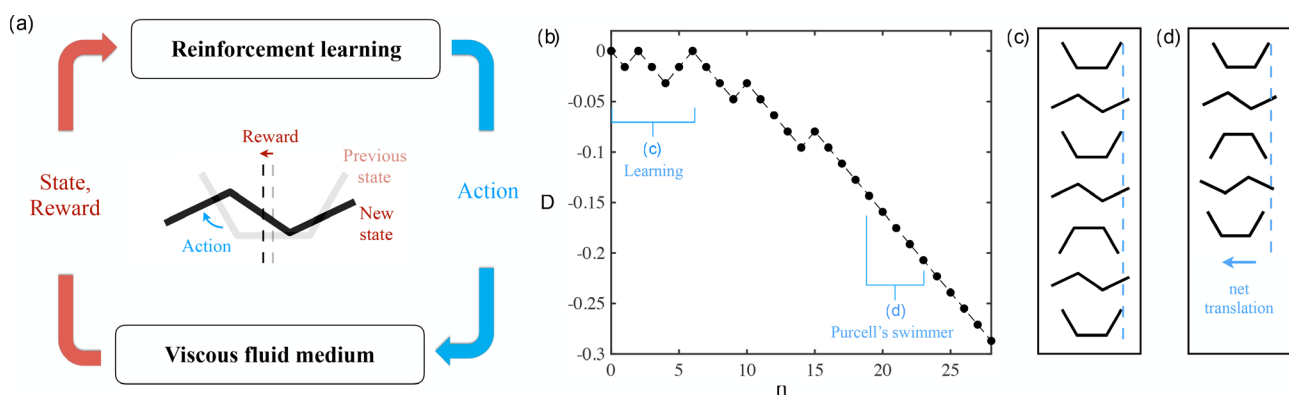


FIG. 2. (a) Reinforcement learning of a multi-link swimmer at low Reynolds numbers. With a goal of self-propulsion, a multi-link swimmer performs an action to rotate one or more than one of its hinges to deform, which leads to the change of its centroid. A reward defined based on the change of the swimmer's centroid measures the immediate success of the action, which serves as an input to the reinforcement learning algorithm to advise the next action. (b) The cumulative displacement of a three-link swimmer's centroid D as a function of learning steps n . (c) In the initial phase of the learning process, the swimmer gains experience by performing different actions to interact with the viscous fluid medium. (d) Through reinforcement learning, the swimmer eventually repeats a sequence of cyclic motions that results in a net translation in the negative x -direction. The swimming gaits identified via reinforcement learning recover those of the classical Purcell's swimmer.⁸

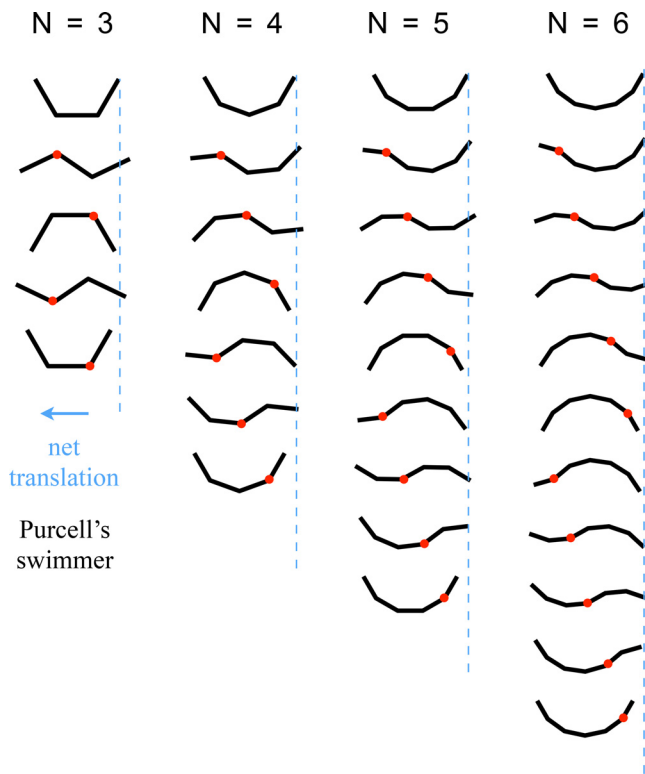


FIG. 3. Swimming gaits identified by reinforcement learning of an N -link swimmer with single hinge rotation, where only one hinge is allowed to rotate in each action step. The red dot indicates the particular hinge that has been rotated relative to the previous action step. For all values of N studied here, the red dot moves from the left to the right, propagating a traveling wave of actuation (hinge rotation) along the multi-link swimmer. Multimedia view: <https://doi.org/10.1063/5.0140662.1>

(hinge rotation) for self-propulsion. The direction of wave propagation is opposite to that of the resulting net translation.

C. Multiple hinge rotation

For results in Secs. III A and III B, only a single hinge is allowed to rotate in each action step. In this section, we remove this restriction to allow multiple hinge rotation in each action step and examine the swimming policies identified by reinforcement learning in these scenarios. Similar to Sec. III B, we consider the cases $3 \leq N \leq 6$ and display the swimming policies most frequently identified by reinforcement learning in Fig. 4(a). For a three-link swimmer ($N = 3$), even allowing multiple hinge rotation, the same swimming gait of Purcell's swimmer is identified, involving only a single hinge rotation at an action step. This behavior of the three-link swimmer results from the incapability of breaking the constraints due to geometrical symmetries and kinematic reversibility with such a highly limited number of degrees of freedom. However, as the number of links increases, new swimming gaits different from those displayed in Fig. 3 emerge when multiple hinge rotation is allowed.

We discuss several features of these new swimming gaits. First, the swimmer might rotate a different number of hinges at an action step. For instance, for $N = 4$, the swimmer first rotates two hinges

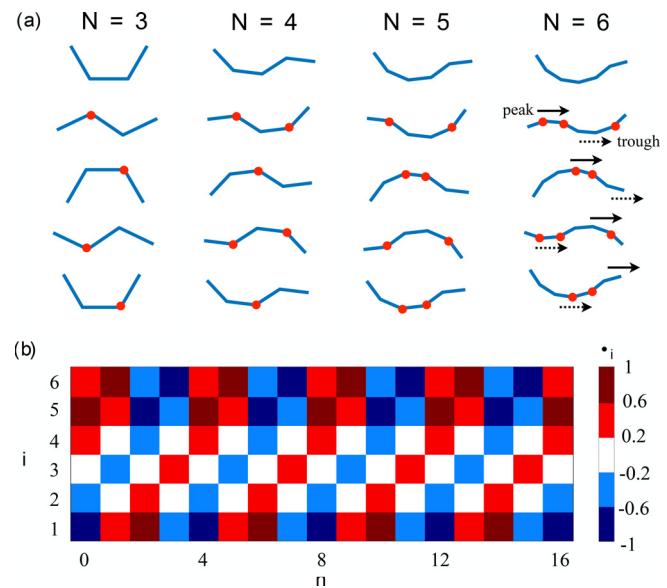


FIG. 4. (a) Swimming gaits identified by reinforcement learning of a N -link swimmer with multiple hinge rotation, where simultaneous rotation of multiple hinges is allowed in each action step. The red dots indicate the particular hinge that has been rotated relative to the previous action step. The swimmer rotates its hinges in a manner akin to the propagation of a transverse flagellar wave along the swimmer, where the peak and trough of the waveform travel from the left to the right. (b) To aid the visualization of the waveform propagation, the angles θ_i along a six-link swimmer are color-coded and displayed at different action steps n . A color pattern is observed to travel in the direction of increasing n . Multimedia view: <https://doi.org/10.1063/5.0140662.2>

simultaneously, followed by a single hinge rotation, before rotating two hinges simultaneously again. In contrast, for $N = 5$, the swimmer always actuates two hinges simultaneously at each step. Finally, for $N = 6$, the swimming gait involves the simultaneous rotation of three hinges and two hinges in different action steps. Second, we observe that the multi-link swimmer rotates their hinges in a manner that resembles the propagation of a transverse flagellar wave; this observation is visually more apparent for the case of $N = 6$, where the propagation of the peak and trough of the waveform is illustrated in Fig. 4(a). See also integral multimedia of Fig. 4 for animations. To help visualize the waveform propagation, we also display in Fig. 4(b) a colormap of the angles θ_i along the swimmer at different action steps n . The observed shifting of the color pattern indicates the propagation of a waveform along the swimmer. Third, for swimming gaits identified in Fig. 3, the swimmer takes increasingly more action steps to complete a cycle as N increases. In contrast, the swimmer exploits multiple hinge rotation in each action step to reduce the number of action steps required to complete a cycle. Indeed, the swimming gaits identified for $3 \leq N \leq 6$ all share the same number of action steps. The reduced number of action steps effectively contributes to the enhancement of propulsion speed.

To better characterize the performance of the new swimming gaits identified with multiple hinge rotation, we define an average propulsion velocity $V = |\Delta \mathbf{x}|/T$, where $\Delta \mathbf{x}$ is the net displacement of the swimmer in one period of T . Figure 5 shows the average propulsion velocity of the multi-link swimmer, which increases with its number of

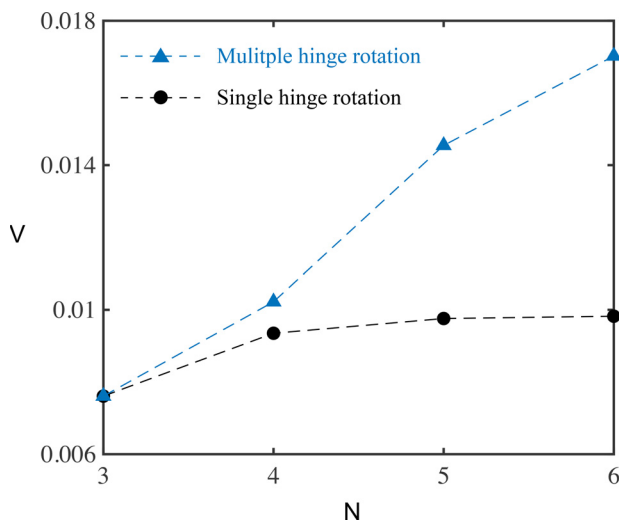


FIG. 5. Average swimming velocity V of a multi-link swimmer vs its number of links N when it is allowed to execute single (circles) and multiple (triangles) hinge rotation. Except for the special case of $N = 3$, where the swimming gaits of both single and multiple hinge rotation are identical (Purcell's swimmer), swimmers with multiple hinge rotation consistently outperform those with single hinge rotation. Multimedia view: <https://doi.org/10.1063/5.0140662.3>

links (blue triangles). For comparison, we display also the corresponding average propulsion velocity for the swimmer adopting single hinge actuation (black circles). Except for the special case of $N = 3$, where both multiple hinge actuation and single hinge actuation reproduce the gaits of a Purcell's swimmer and hence the same performance, swimmers with multiple hinge rotation substantially outperform their counterparts rotating a single hinge. See also integral multimedia of Fig. 5 for animations of the comparison for the $N = 6$ case. The enhanced performance is attributed to both the emergence of new waveforms and the reduction in the cycle period (due to reduced action steps) of the swimming gaits in the case of multiple hinge rotation.

IV. CONCLUDING REMARKS

The design of locomotory gaits at low Reynolds numbers has been a fundamental challenge that requires both ingenuity and knowledge of the physics of locomotion. Purcell first demonstrated the elegant example of how a simple three-link swimmer can perform a sequence of hinge rotation to escape the constraints of scallop theorem and generate self-propulsion. However, the consideration of more complex swimmer configurations may become intractable as the swimmer complexity increases. In this work, we employ reinforcement learning to identify effective swimming gaits of a multi-link swimmer. Without relying on prior knowledge of low-Reynolds-number locomotion, a swimmer progressively learns effective swimming strategies based on its interaction with the surrounding.

In this work, we demonstrated the use of this reinforcement learning approach on a multi-link swimmer. We first considered the case of single hinge rotation, where the multi-link swimmer is allowed to only rotate a single hinge in each action step. In this case, a class of swimming gaits corresponding to the sequential rotation of the hinges along the swimmer is identified, generalizing the classical gaits of Purcell's three-link swimmer. As a remark, the swimming strategy

based on the propagation of a traveling wave of actuation observed here has been also consistently found in multi-sphere systems that generate self-propulsion²⁷ and net rotation,³¹ suggesting some generality of the locomotion strategy. We next relaxed the restriction of single hinge rotation and allow multiple hinges to rotate simultaneously in one action step. In this case, the reinforcement learning algorithm has identified a new class of swimming gaits that resembles the propagation of a transverse flagellar waveform along the swimmer. Moreover, the swimming gaits with multiple hinge rotation display substantially enhanced propulsion performance compared with those using only single hinge rotation.

Taken together, our results demonstrate the use of a simple reinforcement learning technique in identifying swimming gaits at low Reynolds numbers. The approach is not tied to a specific swimmer design but applies to different reconfigurable swimmer models. We remark that we use only a simple reinforcement learning algorithm, Q-learning, in this work for its simplicity. The use of this value-based method nevertheless imposes constraints on the scalability and capabilities of the approach to handle systems with increased degrees of freedom and complex maneuvers, such as continuous and time-dependent actions. Moreover, the variability in the learning process due to the emergence of vertical motion for swimmers with more than three links remains an issue to be addressed in subsequent studies, potentially with more complex rewards functions that penalize vertical swimmer displacements. Due to these constraints, here we limit our investigation to swimmers up to six links, specified and equal amplitudes of rotation for all hinges, as well as a constant rate for the hinge rotations. More advanced reinforcement learning approaches such as the combination of reinforcement learning with artificial neural network^{46–50} may be pursued in the future work to remove these limitations and handle systems with increased complexity. In addition, it would be an interesting direction for future works to consider different reward functions to examine different swimming gaits harvested by reinforcement learning. In particular, power dissipation is another important indicator for locomotion performance and a reward based on minimizing power dissipation could lead to energetically efficient swimming gaits different from those identified here. We also foresee the use of these reinforcement learning approaches in identifying locomotory gaits in complex fluids, where general design guidelines and principles of microswimmers remain elusive.

ACKNOWLEDGMENTS

L.Z. acknowledges the Singapore Ministry of Education Academic Research Fund Tier 2 (No. MOE-T2EP50221-0012) grant, the start-up grant (No. R-265-000-696-133) provided by the National University of Singapore, and the A*Star Advanced Manufacturing and Engineering Young Individual Research Grants, AME-YIRG (No. A2084c0175). O.S.P acknowledges funding support by the National Science Foundation (Grant Nos. 1830958 and 1931292). Computations were performed using the WAVE computing facility at Santa Clara University, enabled by the Wiegand Foundation.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Ke Qin: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Zonghao Zou:** Formal analysis (equal); Investigation (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal). **Lailai Zhu:** Conceptualization (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **On Shun Pak:** Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- ¹G. I. Taylor, “Analysis of the swimming of microscopic organisms,” *Proc. R. Soc. London, Ser. A* **209**, 447–461 (1951).
- ²L. J. Fauci and R. Dillon, “Biofluidmechanics of reproduction,” *Annu. Rev. Fluid Mech.* **38**, 371–394 (2006).
- ³C. Brennen and H. Winet, “Fluid mechanics of propulsion by cilia and flagella,” *Annu. Rev. Fluid Mech.* **9**, 339–398 (1977).
- ⁴E. Lauga and T. R. Powers, “The hydrodynamics of swimming microorganisms,” *Rep. Prog. Phys.* **72**, 096601 (2009).
- ⁵J. Elgeti, R. G. Winkler, and G. Gompper, “Physics of microswimmers—Single particle motion and collective behavior: A review,” *Rep. Prog. Phys.* **78**, 056601 (2015).
- ⁶J. M. Yeomans, D. O. Pushkin, and H. Shum, “An introduction to the hydrodynamics of swimming microorganisms,” *Eur. Phys. J. Spec. Top.* **223**, 1771–1785 (2014).
- ⁷O. S. Pak and E. Lauga, “Theoretical models of low-Reynolds-number locomotion,” in *Fluid-Structure Interactions in Low-Reynolds-Number Flows* (The Royal Society of Chemistry, 2016), Chap. 4, pp. 100–167.
- ⁸E. M. Purcell, “Life at low Reynolds number,” *Am. J. Phys.* **45**, 3 (1977).
- ⁹E. Lauga, “Bacterial hydrodynamics,” *Annu. Rev. Fluid Mech.* **48**, 105–130 (2016).
- ¹⁰M. Arroyo, L. Heltai, D. Millán, and A. DeSimone, “Reverse engineering the euglenoid movement,” *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17874–17879 (2012).
- ¹¹B. J. Nelson, I. K. Kaliakatsos, and J. J. Abbott, “Microrobots for minimally invasive medicine,” *Annu. Rev. Biomed. Eng.* **12**, 55–85 (2010).
- ¹²J. Wang and W. Gao, “Nano/microscale motors: Biomedical opportunities and challenges,” *ACS Nano* **6**, 5745–5751 (2012).
- ¹³Z. Wu, Y. Chen, D. Mukasa, O. S. Pak, and W. Gao, “Medical micro/nanorobots in complex media,” *Chem. Soc. Rev.* **49**, 8088–8112 (2020).
- ¹⁴A. C. H. Tsang, E. Demir, Y. Ding, and O. S. Pak, “Roads to smart artificial microswimmers,” *Adv. Intell. Syst.* **2**, 1900137 (2020).
- ¹⁵L. E. Becker, S. A. Koehler, and H. A. Stone, “On self-propulsion of micro-machines at low Reynolds number: Purcell’s three-link swimmer,” *J. Fluid Mech.* **490**, 15–35 (2003).
- ¹⁶D. Tam and A. E. Hosoi, “Optimal stroke patterns for Purcell’s three-link swimmer,” *Phys. Rev. Lett.* **98**, 068105 (2007).
- ¹⁷O. Wiesel and Y. Or, “Optimization and small-amplitude analysis of Purcell’s three-link microswimmer model,” *Proc. R. Soc. London, Ser. A* **472**, 20160425 (2016).
- ¹⁸L. Giralaldi, P. Martinon, and M. Zoppello, “Optimal design of Purcell’s three-link swimmer,” *Phys. Rev. E* **91**, 023012 (2015).
- ¹⁹R. Marchello, M. Morandotti, H. Shum, and M. Zoppello, “The N-link swimmer in three dimensions: Controllability and optimality results,” *Acta Appl. Math.* **178**, 6 (2022).
- ²⁰F. Alouges, A. DeSimone, L. Giralaldi, and M. Zoppello, “Self-propulsion of slender micro-swimmers by curvature control: N-link swimmers,” *Int. J. Non-Linear Mech.* **56**, 132–141 (2013).
- ²¹F. Alouges, A. DeSimone, L. Giralaldi, and M. Zoppello, “Can magnetic multi-layers propel artificial microswimmers mimicking sperm cells?,” *Soft Rob.* **2**, 117–128 (2015).
- ²²C. Moreau, L. Giralaldi, and H. Gadéha, “The asymptotic coarse-graining formulation of slender-rods, bio-filaments and flagella,” *J. R. Soc. Interface* **15**, 20180235 (2018).
- ²³K. Qin, Z. Peng, Y. Chen, H. Nganguia, L. Zhu, and O. S. Pak, “Propulsion of an elastic filament in a shear-thinning fluid,” *Soft Matter* **17**, 3829–3839 (2021).
- ²⁴Y. Jiao, F. Ling, S. Heydari, N. Heess, J. Merel, and E. Kanso, “Learning to swim in potential flow,” *Phys. Rev. Fluids* **6**, 050505 (2021).
- ²⁵M. Gazzola, B. Hejazi Hosseini, and P. Koumoutsakos, “Reinforcement learning and wavelet adapted vortex methods for simulations of self-propelled swimmers,” *SIAM J. Sci. Comput.* **36**, B622–B639 (2014).
- ²⁶S. Colabrese, K. Gustavsson, A. Celani, and L. Biferale, “Flow navigation by smart microswimmers via reinforcement learning,” *Phys. Rev. Lett.* **118**, 158004 (2017).
- ²⁷A. C. H. Tsang, P. W. Tong, S. Nallan, and O. S. Pak, “Self-learning how to swim at low Reynolds number,” *Phys. Rev. Fluids* **5**, 074101 (2020).
- ²⁸M. Mirzakhani, S. Esmailzadeh, and M.-R. Alam, “Active cloaking in Stokes flows via reinforcement learning,” *J. Fluid Mech.* **903**, A34 (2020).
- ²⁹J. K. Alageshan, A. K. Verma, J. Bec, and R. Pandit, “Machine learning strategies for path-planning microswimmers in turbulent flows,” *Phys. Rev. E* **101**, 043110 (2020).
- ³⁰S. Muñinos-Landin, A. Fischer, V. Holubec, and F. Cichos, “Reinforcement learning with artificial microswimmers,” *Sci. Rob.* **6**, eabd9285 (2021).
- ³¹Y. Liu, Z. Zou, A. C. H. Tsang, O. S. Pak, and Y.-N. Young, “Mechanical rotation at low Reynolds number via reinforcement learning,” *Phys. Fluids* **33**, 062007 (2021).
- ³²B. Hartl, M. Hübl, G. Kahl, and A. Zöttl, “Microswimmers learning chemotaxis with genetic algorithms,” *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2019683118 (2021).
- ³³J. 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).
- ³⁴M. Nasiri and B. Lieberman, “Reinforcement learning of optimal active particle navigation,” *arXiv:2202.00812* (2022).
- ³⁵G. Zhu, W.-Z. Fang, and L. Zhu, “Optimising low-Reynolds-number predation via optimal control and reinforcement learning,” *arXiv:2203.07196* (2022).
- ³⁶Z. Zou, Y. Liu, Y.-N. Young, O. S. Pak, and A. C. Tsang, “Gait switching and targeted navigation of microswimmers via deep reinforcement learning,” *Commun. Phys.* **5**, 158 (2022).
- ³⁷A. Najafi and R. Golestanian, “Simple swimmer at low Reynolds number: Three linked spheres,” *Phys. Rev. E* **69**, 062901 (2004).
- ³⁸R. Dreyfus, J. Baudry, and H. A. Stone, “Purcell’s ‘rotator’: Mechanical rotation at low Reynolds number,” *Eur. Phys. J. B* **47**, 161–164 (2005).
- ³⁹L. Amoudruz and P. Koumoutsakos, “Independent control and path planning of microswimmers with a uniform magnetic field,” *Adv. Intell. Syst.* **4**, 2100183 (2022).
- ⁴⁰M. R. Behrens and W. C. Ruder, “Smart magnetic microrobots learn to swim with deep reinforcement learning,” *Adv. Intell. Syst.* **4**, 2270049 (2022).
- ⁴¹J. Gray and G. J. Hancock, “The propulsion of sea-urchin spermatozoa,” *J. Exp. Biol.* **32**, 802–814 (1955).
- ⁴²J. Lighthill, *Mathematical Biofluidynamics* (SIAM, Philadelphia, 1975).
- ⁴³O. S. Pak and E. Lauga, “Theoretical models in low-Reynolds-number locomotion,” in *Fluid-Structure Interactions in Low-Reynolds-Number Flows*, edited by C. Duprat and H. A. Stone (Royal Society of Chemistry, 2014), p. 100.
- ⁴⁴C. Watkins and P. Dayan, “Q-learning,” *Mach. Learn.* **8**, 279–292 (1992).

- ⁴⁵R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, 2018).
- ⁴⁶J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” [arXiv:1707.06347](#) (2017).
- ⁴⁷V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, “Human-level control through deep reinforcement learning,” *Nature* **518**, 529 (2015).
- ⁴⁸V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, “Asynchronous methods for deep reinforcement learning,” in *ICML* (PMLR, 2016), pp. 1928–1937.
- ⁴⁹Y. Duan, X. Chen, R. Houthoofd, J. Schulman, and P. Abbeel, “Benchmarking deep reinforcement learning for continuous control,” in *ICML* (PMLR, 2016), pp. 1329–1338.
- ⁵⁰Z. Wang, V. Bapst, N. Heess, V. Mnih, R. Munos, K. Kavukcuoglu, and N. de Freitas, “Sample efficient actor-critic with experience replay,” [arXiv:1611.01224](#) (2016).