Editors' Suggestion

Self-learning how to swim at low Reynolds number

Alan Cheng Hou Tsang,[1,2] Pun Wai Tong,[3] Shreyes Nallan,[1] and On Shun Pak[1]
[1]*Department of Mechanical Engineering, Santa Clara University, Santa Clara, California 95053, USA*
[2]*Department of Bioengineering, Stanford University, Stanford, California 94305, USA*
[3]*Clinical Genomics Program, Stanford Health Care, Stanford, California 94305, USA*

(Received 4 December 2018; revised 2 March 2019; accepted 9 June 2020; published 10 July 2020)

Designing locomotory gaits for synthetic microswimmers has been a challenge due to stringent constraints on self-propulsion at low Reynolds numbers (Re). Here, we introduce a new theoretical approach of designing a class of self-learning, adaptive (or "smart") microswimmers via reinforcement learning. Diverging from the traditional paradigm of specifying locomotory gaits *a priori*, here a self-learning swimmer can develop and adapt its propulsion strategy based on its interactions with the surrounding medium. We illustrate this new approach using a minimal but representative model swimmer consisting of an assembly of spheres connected by extensible rods. Without requiring any prior knowledge of low Re locomotion, we demonstrate that this self-learning swimmer can recover a previously known propulsion strategy, identify more effective locomotory gaits, and adapt its locomotory gaits in different media. This approach opens an alternative avenue to designing the next generation of smart microrobots with robust locomotive capabilities.

DOI: 10.1103/PhysRevFluids.5.074101

I. INTRODUCTION

Swimming at the microscopic scale encounters stringent constraints due to the dominance of viscous over inertial forces at low Reynolds numbers (Re) [1,2]. As a result of kinematic reversibility, Purcell's scallop theorem rules out reciprocal motion (i.e., strokes with time-reversal symmetry) for effective locomotion in the absence of inertia [1]. Common macroscopic propulsion strategies thus become ineffective in the microscopic world. Microorganisms have evolved diverse locomotion strategies [2,3], for instance, by rotating helical slender appendages (termed flagella) or propagating deformation waves along flagella via actions of molecular motors, to escape the constraints of the scallop theorem. Extensive efforts in the past few decades have sought to elucidate physical principles that underlie cell motility [4–8]. This has improved our general understanding of locomotion at low Re, which in recent years has engendered a variety of synthetic microswimmers [9–11].

Synthetic microswimmers capable of navigating biological environments offer exciting opportunities for biomedical applications, such as microsurgery and targeted drug delivery [12,13]. Purcell pioneered the design of synthetic microswimmers by inventing a sequence of movements with a three-link swimmer (known as Purcell's swimmer) in a nonreciprocal manner to generate self-propulsion [1,14,15]. Subsequent interdisciplinary efforts have recently resulted in major advances in the design and fabrication of synthetic microswimmers. While some designs are biomimetic or bioinspired (e.g., swimmers with appendages that resemble flagella of microorganisms [16–20]), others ingeniously exploit physical (e.g., Najafi-Golestanian's swimmer [21] and Purcell's "rotator" [22]) and/or physicochemical (e.g., catalytic Janus motors [23,24]) mechanisms available in the microscopic world to self-propel in the absence of inertia.

Successful biomedical applications of synthetic microswimmers rely on their ability to traverse vastly different biological environments, including blood-brain, gastric mucosal barriers, and

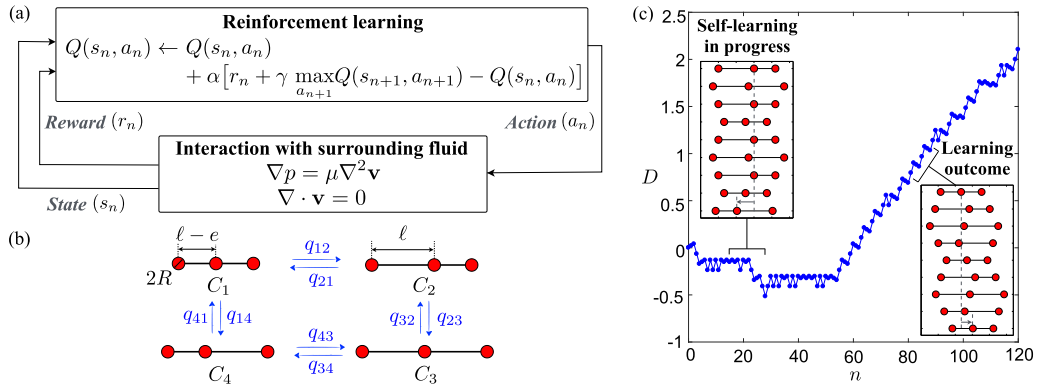

FIG. 1. A self-learning swimmer at low Reynolds numbers. (a) Schematic of reinforcement learning of a swimmer that progressively learns how to swim by interacting with the surroundings. (b) A state diagram for a three-sphere swimmer, where q_{ij} are entries in the Q-matrix that evolve based on reinforcement learning. (c) A typical learning process of a self-learning three-sphere swimmer. The learning outcome is consistent with Najafi-Golestanian's swimmer [21]. We set $\gamma = 0.8$ and $\epsilon = 0.05$ in (c) and $\ell = 10R$ and $e = 4R$ in all cases in this work.

tumor microenvironments [25–27]. Despite significant progress over the past decades, existing microswimmers are typically designed to have fixed locomotory gaits for a particular type of medium or environmental condition. However, gaits that are optimal in one medium may become ineffective in a different medium; hence, locomotion performance of synthetic microswimmers with fixed locomotory gaits may not be robust to environmental changes. In contrast, natural organisms show robust locomotion performance across varying environments by adapting their locomotory gaits to the surroundings [28–30]. Without adaptability like their biological counterparts, it remains formidable for synthetic microswimmers to operate in complex biological media with unpredictable environmental factors. Novel approaches via modular microrobotics and the use of soft active materials have been recently proposed to tackle these challenges [31–36].

Here we leverage the prowess of machine learning to investigate a new approach in designing low Re swimmers. Machine learning has enabled the design of artificial intelligent systems that can perform complex tasks without being explicitly programed [37]. This approach has also sparked several novel directions in fluid mechanics, including modeling of turbulence [38,39], fish schooling [40–42], soaring birds [43,44], bacterial swarms [45], wake detection [46], and navigation problems [47,48]. Unlike other pioneering works that coarse-grained microswimmers as active particles and assumed *a priori* the existence of a self-propelling velocity [47,48], here we focus on a more fundamental challenge of generating self-propulsion at low Re. We ask the general questions: Without any prior knowledge on low Re locomotion, can a swimmer learn how to escape the constraints of the scallop theorem for self-propulsion via a simple machine learning algorithm? How well does this self-learning approach perform for a system with multiple degrees of freedom? Can such a self-learning swimmer adapt its locomotory gaits to traverse media with vastly different properties?

This work contributes to the development of microswimmers by leveraging machine learning to move away from the traditional paradigm of specifying locomotory gaits in advance; instead, a self-learning swimmer can develop and adapt its propulsion strategy based on its interactions with the surrounding environment. Through a minimal but representative canonical swimmer, we demonstrate how a standard reinforcement learning technique already equips the swimmer with previously unattainable capabilities. Specifically, we show that, without requiring any prior knowledge of low Re locomotion, a self-learning swimmer can (1) recover the swimming strategy by Najafi and Golestanian [21] (Fig. 1), (2) identify more effective locomotory gaits with increased degrees of freedom (Figs. 2 and 3), and (3) adapt locomotory gaits in different media (Fig. 4).

This approach offers a new avenue for resolving outstanding challenges in the application of microswimmers in complex environments.

II. SWIMMING AT LOW RE VIA REINFORCEMENT LEARNING

As a model swimmer, we consider here a simple reconfigurable system consisting of N spheres connected by $N-1$ extensible rods of negligible diameters [Fig. 1(b)]. Each configuration of the N-sphere system can transition to $N-1$ different configurations by extending or contracting one of the connecting rods. Previous studies have used similar reconfigurable systems to generate net translation (e.g., Najafi-Golestanian's swimmer [21] and its variants [49–52]), rotation [22], and combined motion [53,54]. Unlike these traditional approaches where the swimming strokes were specified, here the spheres will self-learn propulsion policies based on knowledge gained by interacting with the surroundings via reinforcement learning [55].

A. Reinforcement learning

The use of reinforcement learning enables the swimmer to progressively learn how to act by interacting with the surrounding fluid [Fig. 1(a)]. For a given configuration of the swimmer (the state, s_n) in the n-th learning step, the swimmer can extend or contract one of its rods (the action, a_n) to transforms from the current state to a new state. Such an action results in a displacement of the swimmer's body centroid (the reward, r_n), which measures the immediate success of the action relative to its goal. Here we scale the cumulative displacement d of the swimmer's body centroid by the sphere radius R to track the dimensionless cumulative displacement $D = d/R$ of the swimmer.

We implemented reinforcement learning by a standard Q-learning algorithm for its simplicity and expressiveness [56]. The experience gained by the swimmer is stored in a Q-matrix, $Q(s_n, a_n)$, which is an action-value function that captures the expected long-term reward for taking the action a_n given the state s_n. After each learning step,

$$Q(s_n, a_n) \leftarrow Q(s_n, a_n) + \alpha\big[r_n + \gamma \max_{a_{n+1}} Q(s_{n+1}, a_{n+1}) - Q(s_n, a_n)\big]. \tag{1}$$

Here, α is the learning rate ($0 \leqslant \alpha \leqslant 1$), which determines to what extent new information overrides old information in the Q-matrix. Unless otherwise specified, we fixed $\alpha = 1$ to maximize learning in a fully deterministic system. The Q-matrix encodes the adaptive decision-making intelligence of the swimmers by accounting for both immediate reward r_n and maximum future reward at the next state, $\max_{a_{n+1}} Q(s_{n+1}, a_{n+1})$. The discount factor γ assigns a weight to immediate versus future rewards ($0 \leqslant \gamma < 1$). When γ is small, the swimmer is shortsighted and tends to maximize the immediate reward; when γ is large, the swimmer is farsighted and focuses more on future rewards. In addition, we incorporated an ϵ-greedy selection scheme: in each learning step, the swimmer chooses the best action advised by the Q-matrix with a probability $1-\epsilon$ or takes a random action with a small probability ϵ, which allows the swimmer to explore new solutions and avoids being trapped in only locally optimal policies.

B. Hydrodynamic interactions

The interaction between the spheres and the surrounding viscous fluid is governed by the Stokes equation, $\nabla p = \mu \nabla^2 \mathbf{v}$. For incompressible flows, $\nabla \cdot \mathbf{v} = 0$. Here, p and $\mathbf{v}$ represent, respectively, the pressure and velocity fields, and μ represents the dynamic viscosity. We used the Oseen tensor to consider the hydrodynamic interaction between spheres that are spaced far apart ($R/\ell \ll 1$) [21,57]. The linearity of the Stokes equation allows us to relate the velocities of the sphere $\mathbf{V}_j$ and the forces $\mathbf{F}_j$ acting on them as

$$\mathbf{V}_i = \sum_{j=1}^{N} \mathbf{H}_{ij} \mathbf{F}_j. \tag{2}$$

Here, the Oseen tensor $\mathbf{H}_{ij}$ for spheres is given by

$$\mathbf{H}_{ij} = \begin{cases} \mathbf{I}/6\pi\mu R, & \text{if } i = j \\ (1/8\pi\mu|\mathbf{x}_{ij}|)(\mathbf{I} + \mathbf{x}_{ij}\mathbf{x}_{ij}/|\mathbf{x}_{ij}|^2), & \text{if } i \neq j \end{cases}, \tag{3}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{x}_{ij} = \mathbf{x}_j - \mathbf{x}_i$ denotes the vector between spheres i and j. The instantaneous positions of the spheres $\mathbf{x}_i$ are determined by enforcing, respectively, the force-free and torque-free conditions,

$$\sum_{i=1}^{N} \mathbf{F}_i = \mathbf{0}, \tag{4}$$

$$\sum_{i=1}^{N} \mathbf{F}_i \times \mathbf{x}_i = \mathbf{0}. \tag{5}$$

The linearity and time-independence of the Stokes equation imply that the displacement of a swimmer depends only on the sequence of configurations (or states) changes of the swimmer. The sequence of state changes hence defines the propulsion policy of a low Re swimmer [1,2].

III. RESULTS AND DISCUSSION

A. A self-learning three-sphere swimmer

We first considered a three-sphere swimmer ($N = 3$), which has the minimal degrees of freedom for swimming at low Re [1,21] [Figs. 1(b) and 1(c), Movie S1]. The swimmer has four different configurations [Fig. 1(b)]. In each learning step, the swimmer switches from one configuration to another, and updates the corresponding entry in the Q-matrix according to Eq. (1) (see Sec. II in the Supplemental Material [58] for the evolution of the entries). Figure 1(c) depicts a typical self-learning process: the swimmer initially struggles to find a policy to swim forward and thus moves back and forth (left inset), where D remains close to 0. The swimmer keeps exploring the surrounding medium by taking different actions and adapting its propulsion policy. After accumulating enough knowledge, the swimmer develops an effective propulsion policy that repeats the same sequence of action (except with ϵ probability, at which a random action is chosen), and swims with increasing D (right inset). The propulsion policy obtained by our learning algorithm for a three-sphere swimmer is consistent with Najafi-Golestanian's swimmer [21]. This example demonstrates, for the first time, how reinforcement learning enables a swimmer to self-learn how to swim with no prior knowledge of low Re locomotion.

B. Extending to N-sphere swimmers

We applied this self-learning approach to systems with increased number of spheres. Unlike the three-sphere ($N = 3$) swimmer, where only one propulsion policy leads to net translation, multiple propulsion policies are possible when the number of spheres increases [53]. In Fig. 2(a), we first use a four-sphere system ($N = 4$) to illustrate that a swimmer equipped with reinforcement learning is not only able to identify multiple propulsion policies (e.g., policies I–III) but also self-improves and evolves a better policy during the learning process (e.g., from policy I to policy II; see Movie S2). We illustrate the effect of discount factor γ on the learning outcomes: first, with $\gamma = 0.6$, the swimmer learns one propulsion policy in the initial stage [Fig. 2(a), policy I]. Through continuous learning, the swimmer keeps modifying its Q-matrix and eventually identifies a better propulsion policy [Fig. 2(a), policy II], as indicated by the increased slope of D in the left panel of Fig. 2(a). We note that this propulsion policy [Fig. 2(a), policy II] is reminiscent of the propagation of a longitudinal traveling wave along a cell body [59] and was shown to be optimal for a four-sphere system [53]. We note that continuous improvement of the propulsion policy does not always happen, as

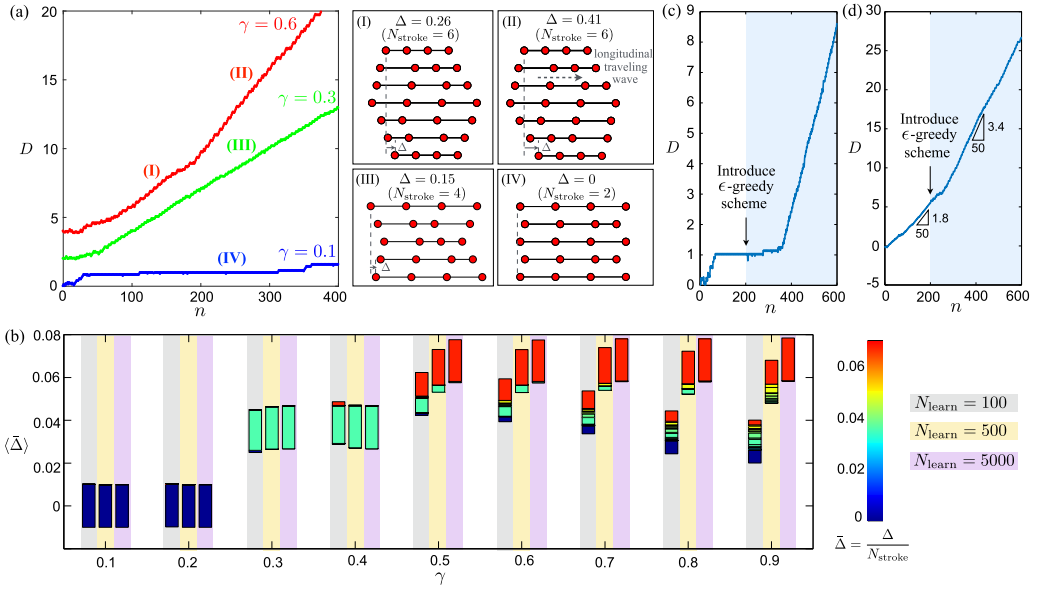

FIG. 2. Self-learning of N-sphere swimmers. (a) A four-sphere swimmer learns several propulsion policies depending on γ. We set $\epsilon = 0.05$. Panels (I)–(IV) depict four different policies obtained throughout the learning process, showing number of strokes (N_{stroke}) and net displacement over one cycle (Δ, scaled by R): (II) is the optimal policy with a longitudinal traveling wave pattern and (IV) is a failed policy with which the swimmer does not swim. (b) The learning outcome as a function of γ and total number of learning steps N_{learn}. Here, we varied γ from 0.1 to 0.9 in increments of 0.1 and N_{learn} between 100 (shaded regions in gray), 500 (yellow), and 5000 (purple), with a fixed $\epsilon = 0.1$. Each bar represents the results of 1,000 individual simulations and the colors of each segment represent the net displacement over one cycle divided by the number of strokes, $\bar{\Delta}$. Since distinct propulsion policies can have the same $\bar{\Delta}$ and hence the same color, they are separated by a black border in the bar. The middle of each bar represents the averaged $\bar{\Delta}$ for the 1,000 simulations, $\langle\bar{\Delta}\rangle$. (c, d) Trade-off between exploration and exploitation. Introducing an ϵ-greedy scheme with $\epsilon = 0.05$ allows the swimmer to escape from locally trapped policies and evolve a better propulsion policy: (c) The swimmer transitions from a failed to a successful policy at $\gamma = 0.4$. (d) The swimmer transitions from a sub-optimal to an optimal propulsion policy at $\gamma = 0.9$.

demonstrated with $\gamma = 0.3$ in Fig. 2(a). In this case, the swimmer learns a different but suboptimal propulsion policy [Fig. 2(a), policy III] and cannot improve the policy further. When γ is even smaller (e.g., $\gamma = 0.1$), the swimmer fails to learn any effective propulsion policy; the swimmer moves back-and-forth with zero net translation [Fig. 2(a), policy IV].

C. Influences of learning parameters on self-learning swimmers

In this section, we systematically investigated the influences of learning parameters on self-learning swimmers. Here we consider a four-sphere swimmer as an example; as shown in Fig. 2(a), the four-sphere swimmer can evolve to multiple propulsion strategies as a result of the reinforcement learning. The learning outcome depends on how much the swimmer values an immediate reward (γ), how often the swimmer explores randomly (ϵ), and how many learning steps the swimmer takes (N_{learn}). We explored different possibilities by performing simulations with random initial states, where we fixed $\epsilon = 0.1$ and varied γ and N_{learn}. We performed 1000 simulations for each set of parameters and extracted the resulting propulsion policies given by the Q-matrix after learning. To distinguish propulsion policies with different numbers of strokes per cycle, we characterized each

propulsion policy by the displacement per stroke, $\bar{\Delta} = \Delta/N_{\text{stroke}}$, which divides the net displacement over one cycle Δ by the number of strokes in the cycle N_{stroke}. The average of $\bar{\Delta}$ over 1000 simulations is defined as $\langle\bar{\Delta}\rangle$. The simulation results revealed three main findings [Fig. 2(b)]:

First, given sufficiently large N_{learn} (e.g., $N_{\text{learn}} = 5000$), the learning process evolves into three major outcomes, depending on the value of γ [Fig. 2(b)]. At a small γ ($\leqslant$0.2), the swimmer fails to swim [e.g., a two-stroke policy in Fig. 2(a), policy IV]. At an intermediate γ (e.g., $\gamma = 0.3$ to 0.4), the swimmer identifies an effective but suboptimal policy [e.g., a four-stroke policy in Fig. 2(a), policy III]. At a large γ ($\geqslant$0.5), the swimmer learns the optimal policy [a six-stroke policy in in Fig. 2(a), policy II], which corresponds to a longitudinal traveling wave pattern.

Second, there exists a threshold of γ, below which the swimmer cannot learn the optimal propulsion policy (e.g., for a four-sphere system, the critical $\gamma \lesssim 0.5$). The learning process leads to suboptimal policies even with many learning steps. This occurs because compared to the suboptimal policies, the optimal policy involves more swimming strokes, including those that contribute immediate, negative rewards in the cycle. Therefore, only a far-sighted swimmer (large γ) can learn the optimal propulsion policy. Before the propulsion policy converges (e.g., when $N_{\text{learn}} = 100$), a small portion of swimmers at $\gamma = 0.4$ follow the optimal policy due to random initialization of the Q-matrix, but the policy eventually converges to a suboptimal policy at large N_{learn} (e.g., $N_{\text{learn}} = 5000$).

Third, for a given number of learning steps N_{learn}, an optimal γ maximizes the portion of swimmers that can acquire the optimal propulsion policy [Fig. 2(b)]. As N_{learn} increases, the optimal γ increases its value from $\gamma \approx 0.5$ for $N_{\text{learn}} = 100$ to $\gamma \approx 0.7$ for $N_{\text{learn}} = 500$. These results illustrate that while a sufficiently large γ is necessary to acquire the optimal policy, an excessively large γ (e.g., $\gamma = 0.9$) can delay the learning of the optimal policy because the swimmer becomes too far-sighted and largely ignores immediate rewards for new possibilities. In addition, when there is only a small number of learning steps [e.g., $N_{\text{learn}} = 100$ in Fig. 2(b)], this emphasis on long-term benefits results in harvesting more distinct policies as γ increases, which can hamper the overall learning outcomes [see decrease in $\langle\bar{\Delta}\rangle$ for learning processes with $N_{\text{learn}} = 100$ and $\gamma > 0.5$ in Fig. 2(b)].

The effects of ϵ on self-learning the propulsion policy are obvious when we compared $\epsilon = 0$ and $\epsilon > 0$ in Figs. 2(c) and 2(d). When $\epsilon = 0$ (greedy policy), the swimmer can get trapped in certain suboptimal propulsion policies. For instance, a swimmer may be trapped in a failed policy that yields no net displacement [white region in Fig. 2(c)] or an effective but suboptimal policy [white region in Fig. 2(d)]. In either case, the introduction of $\epsilon > 0$ (epsilon-greedy policy) helps kick the swimmer away from these locally trapped policies, thus enabling the swimmer to continually improve its propulsion policy to its fullest extent for a given value of γ [blue regions in Figs. 2(c) and 2(d)]. We remark that, however, the ϵ-greedy scheme does not show any significant effect in the regime of small γ ($\leqslant$0.2), where the swimmer fails to learn any effective policy as shown in Fig. 2(a), even with a fourfold increase in ϵ.

Taken together, these findings reveal how learning parameters influence the robustness of the self-learning approach for identifying effective propulsion policies.

D. Benchmarking the self-learning approach

Next we probed the performance of the Q-learning approach for swimmers consisting of up to ten spheres (i.e., $N = 3$ to 10) in Fig. 3. With sufficiently large number of learning steps, we found that these N-sphere swimmers all learn the propulsion policy reminiscent of the longitudinal traveling wave pattern illustrated in Fig. 2(a), policy II. For every N considered, we evaluated the minimum number of learning step required for a swimmer to learn the traveling-wave propulsion policy (N_{QL}, blue box plots in Fig. 3). More learning steps are required for systems with increased number of spheres as expected, but the rate of increase levels off for larger values of N. We note that the learning algorithm harvests effective policies from a pool consisting of a tremendous number of stroke combinations as the number of spheres increases. A crude estimate of the size of the pool of combinations goes as follows: the number of combinations that have the same number of

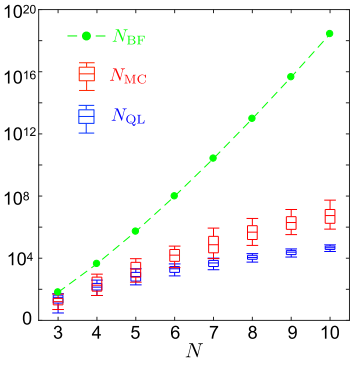

FIG. 3. The number of learning steps required for an N-sphere swimmer to learn the traveling-wave propulsion policy is shown in the box plot. Here N_{QL} (blue box plots), N_{MC} (red box plots), and N_{BF} (green circles) represent results for Q-learning, Monte Carlo method, and brute-force search, respectively. In the box plots, the midlines represent the median, and the box's upper and lower bounds indicate the interquartile range; the upper and lower whiskers denote the 9th and 91st percentile of the simulation data, respectively. We performed 100 simulations for each value of N, where $\gamma = 0.9$ and $\epsilon = 0.3$.

strokes as the traveling wave policy [i.e., $2(N-1)$ number of strokes] is $(N-1)^{2(N-1)}$. To perform a brute-force search, a total number of $N_{\mathrm{BF}} = 2(N-1)^{2N-1}$ steps are required to go through all combinations (green dashed line, Fig. 3). The brute-force search becomes quickly intractable as N increases and N_{BF} are orders of magnitude greater than N_{QL}. We also benchmark the results with another commonly used scheme based on the Monte Carlo method [55]. Q-learning clearly outperforms the Monte Carlo method as the degree of freedom of the system increases (e.g., $N \geqslant 5$). See Sec. III in the Supplemental Material [58] for more details. These results indicate how standard machine learning methods can be leveraged in Stokesian locomotion to harvest effective propulsion policies from a tremendous number of stroke combinations. We remark that while the standard Q-learning algorithm employed here can be extended to consider even larger values of N, there exists a vast potential for improving the scalability for large state and action spaces using more advanced machine learning approaches (e.g., combining Q-learning with deep convolutional neural networks [60]).

E. Swimming performance in a noisy environment

We assessed how a self-learning swimmer behaves under the influence of random noises from the environment. In each learning step, we introduced noise to the displacement Δd (and hence reward) of the swimmer: $r_n = \Delta d(1 + \xi U)$, where ξ is the noise level and U is a random variable with a uniform distribution in $[-1, 1]$. To illustrate, we considered a three-sphere swimmer and used cumulative displacement D at $n = 200$ as a metric for the performance of the swimmer under increasing noise levels ξ [Fig. 4(a), Movie S3]. When the noise is weak ($\xi \leqslant 0.5$), the swimmer with a learning rate $\alpha = 1$ (blue line) performs the best. As the noise level increases, $\alpha = 1$ no longer guarantees the best performance and different optimal values of α exist depending on the noise level. Remarkably, when the noise level is as large as 100% of the instantaneous displacement ($\xi = 1$), a self-learning swimmer with $\alpha = 0.8$ still reaches over 85% of the mean displacement compared with the noise-free environment ($\xi = 0$); even when the noise level is twice as much as the instantaneous displacement ($\xi = 2$), a swimmer with $\alpha = 0.4$ to 0.6 can retain over 60% of its noise-free performance [Fig. 4(a)]. Thus, a self-learning swimmer can robustly adapt to swim in a noisy environment, and further improve its performance by adjusting its learning rate.

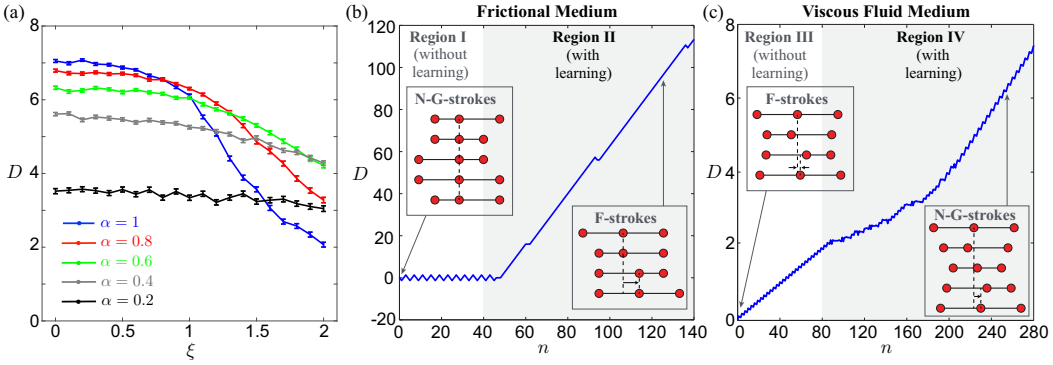

FIG. 4. A self-learning swimmer can adapt to different environments. (a) The performance of a self-learning swimmer in a noisy environment. We measured the swimmer's displacement D after $n = 200$ under increasing noise level ξ and various learning rates α. Each data point represents the mean of D for 1000 simulations. The error bars denote the standard error of the mean. (b, c) A self-learner adapts its locomotory gait to propel effectively in vastly different media: (b) In a frictional medium, Najafi-Golestanian's strokes (termed N-G-strokes; left inset) fails to propel (region I). When the learning algorithm is turned on, the self-learner rapidly identifies new locomotory gaits (termed F-strokes; right inset) to propel effectively (region II); (c) in a viscous fluid medium, F-strokes (left inset) also propels the swimmer (region III), but reinforcement learning (region IV) enables the swimmer to re-learn the more effective N-G-strokes (right inset). In all cases, $\gamma = 0.7$, $\epsilon = 0.05$.

F. Adaptive locomotion across different media

Finally, we demonstrated the adaptivity of this self-learning approach in a medium vastly different from viscous fluids—a frictional environment [30,61,62] where motion arises by the interaction of surface friction $\mathbf{F}_i$ and net driving forces $\mathbf{f}_i$ exerted on each sphere by the rods. We considered here again a three-sphere swimmer for illustration. As will be shown below, the lack of hydrodynamic interaction causes the Najafi-Golestanian's strokes to become ineffective in the frictional medium. We therefore incorporate additional degrees of freedom into the three-sphere swimmer by allowing both rods to move in the same step, thereby enabling free transitions between the four states in Fig. 1(b). We restricted our analysis to a standard Coulomb sliding friction law [61,62]: when the magnitude of the net driving force on a sphere is greater than the sliding friction F_s (i.e., $|\mathbf{f}_i| > F_s$), the friction acting on the sphere is given by $\mathbf{F}_i = -F_s\hat{\mathbf{V}}_i$, where $\hat{\mathbf{V}}_i = \mathbf{V}_i/|\mathbf{V}_i|$ is the velocity direction of the sphere. When $|\mathbf{f}_i| \leqslant F_s$, the static friction balances the net driving force, $\mathbf{F}_i = -\mathbf{f}_i$.

We consider again the inertialess regime, where inertial forces are subdominant to frictional forces [62]. Frictional forces are therefore transduced directly to velocities instead of accelerations, similar to the low Re regime in viscous flows. As a result, locomotion in frictional media in this inertialess regime is kinematic (also a feature of Stokesian locomotion), in that the net displacement is independent of its rate but only the sequence of deformations [63]. Despite the similarities between frictional and viscous fluid media, a key difference is the absence of hydrodynamic interactions in the frictional medium. This difference renders Najafi-Golestanian's strokes (N-G-strokes) of a three-sphere swimmer ineffective in a frictional medium [region I in Fig. 4(b); without learning, Movie S4]. Nevertheless, when we turn on reinforcement learning and allow simultaneous actuation of both rods, the self-learner rapidly adapts to the frictional medium and identifies a new, effective propulsion policy (region II; F-strokes). We note that it takes significantly fewer steps to learn propulsion policies in the frictional medium than in the viscous medium because the F-strokes do not involve steps that contribute intermediate, negative rewards. Finally, we found that the new locomotory gaits identified in the frictional medium (F-strokes) also propel a swimmer in

a viscous fluid [region III in Fig. 4(c); without learning, Movie S5]; nevertheless, a swimmer with reinforcement learning will explore, relearn, and evolve a better propulsion policy to adapt to the surrounding medium (region IV; returning to the N-G-strokes). The adaptivity demonstrated here represents a first step in realizing a smart "amphibian" microrobot that can move effectively in both liquid and solid terrains by adjusting its locomotory gaits.

IV. CONCLUDING REMARKS

This work presents the first integration of machine learning into the design of locomotory gaits at low Reynolds numbers. Several novel features emerge as a result of such integration, from the self-learning of effective propulsion strategies to adaptive locomotion in response to environmental changes. We considered a canonical example of N-sphere swimmer to illustrate this new approach and set a benchmark for future applications of this new approach in other biological or synthetic systems. There exists vast potential in further optimizing the learning algorithms or pursuing more advanced machine learning approaches. Subsequent works will also focus on designing self-learning swimmers with more complex maneuvers than net translation. We will also extend the investigation to account for more realistic biological environments, which are typically highly heterogeneous and display complex (non-Newtonian) rheological properties. We envision that this new approach would be particularly relevant for complex environments where the physics of locomotion remains not poorly understood, or when the properties of the medium are unpredictable or impossible to characterize in advance.

Finally, while technological implementation is beyond the scope of this work, we discuss several possible experimental platforms for implementation. Swimmers comprising of linked spheres have already been realized experimentally by various actuation mechanisms, which enable the transformation of a swimmer between different configurations, via optical tweezers [64] or external magnetic fields [65–67]. In addition, soft active materials, including hydrogels [68,69] and liquid-crystal elastomers [70], which exhibit configurational changes in response to heat or light, form another class of potential actuation mechanisms for our model system [32,71]. These actuation mechanisms are programmable and can be integrated with real-time microscopy [48] to generate different locomotory gaits informed by reinforcement learning. To conclude, we have leveraged the prowess of machine learning to demonstrate an alternative avenue to designing the next generation of smart microrobots with robust locomotive capabilities. These initial theoretical efforts and plausible implementation ideas call for future experimental realizations.

ACKNOWLEDGMENTS

Funding by the National Science Foundation (Grants No. EFMA-1830958 and No. CBET-1931292 to O.S.P.) is gratefully acknowledged. We also thank Wai Tong (Louis) Fan and Yi Fang for useful discussion.

[1] E. M. Purcell, Life at low Reynolds number, Am. J. Phys. **45**, 3 (1977).
[2] E. Lauga and T. R. Powers, The hydrodynamics of swimming microorganisms, Rep. Prog. Phys. **72**, 096601 (2009).
[3] L. J. Fauci and R. Dillon, Biofluidmechanics of reproduction, Annu. Rev. Fluid Mech. **38**, 371 (2006).
[4] J. M. Yeomans, D. O. Pushkin, and H. Shum, An introduction to the hydrodynamics of swimming microorganisms, Eur. Phys. J.: Spec. Top. **223**, 1771 (2014).
[5] J. Elgeti, R. G. Winkler, and G. Gompper, Physics of microswimmers—Single particle motion and collective behavior: A review, Rep. Prog. Phys. **78**, 056601 (2015).

[6] O. S. Pak and E. Lauga, Chapter 4 Theoretical models of low-Reynolds-number locomotion, in *Fluid-Structure Interactions in Low-Reynolds-Number Flows* (The Royal Society of Chemistry, Cambridge, 2016), pp. 100–167.
[7] E. Lauga, Bacterial hydrodynamics, Annu. Rev. Fluid Mech. **48**, 105 (2016).
[8] D. Saintillan, Rheology of active fluids, Annu. Rev. Fluid Mech. **50**, 563 (2018).
[9] S. J. Ebbens and J. R. Howse, In pursuit of propulsion at the nanoscale, Soft Matter **6**, 726 (2010).
[10] S. Sengupta, M. E. Ibele, and A. Sen, Fantastic voyage: Designing self-powered nanorobots, Angew Chem. Int. Ed. **51**, 8434 (2012).
[11] C. Hu, S. Pané, and B. J. Nelson, Soft micro- and nanorobotics, Annu. Rev. Control Robot. Auton. Syst. **1**, 53 (2018).
[12] B. J. Nelson, I. K. Kaliakatsos, and J. J. Abbott, Microrobots for minimally invasive medicine, Annu. Rev. Biomed. Eng. **12**, 55 (2010).
[13] W. Gao and J. Wang, Synthetic micro/nanomotors in drug delivery, Nanoscale **6**, 10486 (2014).
[14] L. E. Becker, S. A. Koehler, and H. A. Stone, On self-propulsion of micromachines at low Reynolds number: Purcell's three-link swimmer, J. Fluid Mech. **490**, 15 (2003).
[15] D. Tam and A. E. Hosoi, Optimal Stroke Patterns for Purcell's Three-Link Swimmer, Phys. Rev. Lett. **98**, 068105 (2007).
[16] R. Dreyfus, J. Baudry, M. L. Roper, M. Fermigier, H. A. Stone, and J. Bibette, Microscopic artificial swimmers, Nature **437**, 862 (2005).
[17] J. J. Abbott, K. E. Peyer, M. C. Lagomarsino, L. Zhang, L. Dong, I. K. Kaliakatsos, and B. J. Nelson, How should microrobots swim? Int. J. Robotics Res. **28**, 1434 (2009).
[18] O. S. Pak, W. Gao, J. Wang, and E. Lauga, High-speed propulsion of flexible nanowire motors: Theory and experiments, Soft Matter **7**, 8169 (2011).
[19] L. Zhang, J. J. Abbott, L. Dong, K. E. Peyer, B. E. Kratochvil, H. Zhang, C. Bergeles, and B. J. Nelson, Characterizing the swimming properties of artificial bacterial flagella, Nano Lett. **9**, 3663 (2009).
[20] A. Ghosh and P. Fischer, Controlled propulsion of artificial magnetic nanostructured propellers, Nano Lett. **9**, 2243 (2009).
[21] A. Najafi and R. Golestanian, Simple swimmer at low Reynolds number: Three linked spheres, Phys. Rev. E **69**, 062901 (2004).
[22] R. Dreyfus, J. Baudry, and H. A. Stone, Purcell's "rotator": Mechanical rotation at low Reynolds number, Euro. Phys. J. B **47**, 161 (2005).
[23] W. F. Paxton, S. Sundararajan, T. E. Mallouk, and A. Sen, Chemical locomotion, Angew. Chem. Int. Ed. **45**, 5420 (2006).
[24] J. L. Moran and J. D. Posner, Phoretic self-propulsion, Annu. Rev. Fluid Mech. **49**, 511 (2017).
[25] X. Nassif, S. Bourdoulous, E. Eugène, and P.-O. Couraud, How do extracellular pathogens cross the blood–brain barrier? Trends Microbiol. **10**, 227 (2002).
[26] J. P. Celli, B. S. Turner, N. H. Afdhal, S. Keates, I. Ghiran, C. P. Kelly, R. H. Ewoldt, G. H. McKinley, P. So, S. Erramilli, and R. Bansil, Helicobacter pylori moves through mucus by reducing mucin viscoelasticity, Proc. Natl. Acad. Sci. USA **106**, 14321 (2009).
[27] S. A. Mirbagheri and H. C. Fu, Helicobacter Pylori Couples Motility and Diffusion to Actively Create a Heterogeneous Complex Medium in Gastric Mucus, Phys. Rev. Lett. **116**, 198101 (2016).
[28] O. R. Barclay, The mechanics of amphibian locomotion, J. Exp. Biol. **23**, 177 (1946).
[29] C. Fang-Yen, M. Wyart, J. Xie, R. Kawai, T. Kodger, S. Chen, Q. Wen, and A. D. T. Samuel, Biomechanical analysis of gait adaptation in the nematode caenorhabditis elegans, Proc. Natl. Acad. Sci. USA **107**, 20323 (2010).
[30] R. D. Maladen, Y. Ding, C. Li, and D. I. Goldman, Undulatory swimming in sand: Subsurface locomotion of the sandfish lizard, Science **325**, 314 (2009).
[31] U. K. Cheang, F. Meshkati, H. Kim, K. Lee, H. C. Fu, and M. J. Kim, Versatile microrobotics using simple modular subunits, Sci. Rep. **6**, 30472 (2016).
[32] S. Palagi, A. G. Mark, S. Y. Reigh, K. Melde, T. Qiu, H. Zeng, C. Parmeggiani, D. Martella, A. Sanchez-Castillo, N. Kapernaum, F. Giesselmann, D. S. Wiersma, E. Lauga, and P. Fischer, Structured light enables

biomimetic swimming and versatile locomotion of photoresponsive soft microrobots, Nat. Mater. **15**, 647 (2016).
[33] A. v. Rohr, S. Trimpe, A. Marco, P. Fischer, and S. Palagi, Gait learning for soft microrobots controlled by light fields, 2018 IEEE Int. C. Int. Robot. 6199 (2018).
[34] W. Hu, G. Z. Lum, M. Mastrangeli, and M. Sitti, Small-scale soft-bodied robot with multimodal locomotion, Nature **554**, 81 (2018).
[35] H.-W. Huang, F. E. Uslu, P. Katsamba, E. Lauga, M. S. Sakar, and B. J. Nelson, Adaptive locomotion of artificial microswimmers, Sci. Adv. **5**, eaau1532 (2019).
[36] A. C. H. Tsang, E. Demir, Y. Ding, and O. S. Pak, Roads to smart artificial microswimmers, Adv. Intell. Syst. **n/a**, 1900137 (2020).
[37] M. I Jordan and T. M. Mitchell, Machine learning: Trends, perspectives, and prospects, Science **349**, 255 (2015).
[38] J. Ling, A. Kurzawski, and J. Templeton, Reynolds averaged turbulence modeling using deep neural networks with embedded invariance, J. Fluid Mech. **807**, 155 (2016).
[39] J. Nathan Kutz, Deep learning in fluid dynamics, J. Fluid Mech. **814**, 1 (2017).
[40] M. Gazzola, B. Hejazialhosseini, and P. Koumoutsakos, Reinforcement learning and wavelet adapted vortex methods for simulations of self-propelled swimmers, SIAM J. Sci. Comput. **36**, B622 (2014).
[41] M. Gazzola, A. A. Tchieu, D. Alexeev, A. de Brauer, and P. Koumoutsakos, Learning to school in the presence of hydrodynamic interactions, J. Fluid Mech. **789**, 726 (2016).
[42] S. Verma, G. Novati, and P. Koumoutsakos, Efficient collective swimming by harnessing vortices through deep reinforcement learning, Proc. Natl. Acad. Sci. USA **115**, 5849 (2018).
[43] G. Reddy, A. Celani, T. J. Sejnowski, and M. Vergassola, Learning to soar in turbulent environments, Proc. Natl. Acad. Sci. USA **113**, E4877 (2016).
[44] G. Reddy, J. Wong-Ng, A. Celani, T. J. Sejnowski, and M. Vergassola, Glider soaring via reinforcement learning in the field, Nature **562**, 236 (2018).
[45] H. Jeckel, E. Jelli, R. Hartmann, P. K. Singh, R. Mok, J. F. Totz, L. Vidakovic, B. Eckhardt, J. Dunkel, and K. Drescher, Learning the space-time phase diagram of bacterial swarm expansion, Proc. Natl. Acad. Sci. USA **116**, 1489 (2019).
[46] B. Colvert, M. Alsalman, and E. Kanso, Classifying vortex wakes using neural networks, Bioinspir. Biomim. **13**, 025003 (2018).
[47] S. Colabrese, K. Gustavsson, A. Celani, and L. Biferale, Flow Navigation by Smart Microswimmers via Reinforcement Learning, Phys. Rev. Lett. **118**, 158004 (2017).
[48] S. Muiños-Landin, K. Ghazi-Zahedi, and F. Cichos, Reinforcement learning of artificial microswimmers, arXiv:1803.06425 (2018).
[49] J. E. Avron, O. Kenneth, and D. H. Oaknin, Pushmepullyou: An efficient microswimmer, New J. Phys. **7**, 234 (2005).
[50] R. Golestanian and A. Ajdari, Stochastic low Reynolds number swimmers, J. Phys.: Condens. Matter **21**, 204104 (2009).
[51] F. Alouges, A. DeSimone, and A. Lefebvre, Optimal strokes for low Reynolds number swimmers: An example, J. Nonlinear Sci. **18**, 277 (2008).
[52] B. Nasouri, A. Vilfan, and R. Golestanian, Efficiency limits of the three-sphere swimmer, Phys. Rev. Fluids **4**, 073101 (2019).
[53] D. J. Earl, C. M. Pooley, J. F. Ryder, I. Bredberg, and J. M. Yeomans, Modeling microscopic swimmers at low Reynolds number, J. Chem. Phys. **126**, 064703 (2007).
[54] F. Alouges, A. DeSimone, L. Heltai, A. Lefebvre-Lepot, and B. Merlet, Optimally swimming Stokesian robots, Discrete Cont. Dyn. Syst. Ser. B. **18**, 1189 (2012).
[55] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, 1998).
[56] C. Watkins and P. Dayan, Q-learning, Mach. Learn. **8**, 279 (1992).
[57] J. Happel and H. Brenner, *Low Reynolds Number Hydrodynamics: with Special Applications to Particulate Media* (Noordhoff International Publishing, Leiden, 1973).
[58] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevFluids.5.074101 for more details on the evolution of the action-value function, the learning performance, and the

implementation of the learning algorithms. For theoretical analyses of Q-learning algorithms, see also: Szepesvári, Cs., "The Asymptotic Convergence-rate of Q-learning," *Proceedings of the 10th International Conference on Neural Information Processing Systems*, NIPS'97, 1064–1070 (1997); E. Even-Dar and Y. Mansour, Learning rates for Q-learning, J. Mach. Learn. Res. **5**, 1 (2003); and M. Ghavamzadeh, H. J. Kappen, M. G. Azar, and R. Munos, Speedy Q-learning, Adv. Neural Info. Process. Syst. **24**, 2411 (2011).

[59] K. M. Ehlers, A. D. Samuel, H. C. Berg, and R. Montgomery, Do cyanobacteria swim using traveling surface waves? Proc. Natl. Acad. Sci. USA **93**, 8340 (1996).

[60] V. 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).

[61] Z. V. Guo and L. Mahadevan, Limbless undulatory propulsion on land, Proc. Natl. Acad. Sci. USA **105**, 3179 (2008).

[62] D. L. Hu, J. Nirody, T. Scott, and M. J. Shelley, The mechanics of slithering locomotion, Proc. Natl. Acad. Sci. USA **106**, 10081 (2009).

[63] R. L. Hatton, Y. Ding, H. Choset, and D. I. Goldman, Geometric Visualization of Self-Propulsion in a Complex Medium, Phys. Rev. Lett. **110**, 078101 (2013).

[64] M. Leoni, J. Kotar, B. Bassetti, P. Cicuta, and M. C. Lagomarsino, A basic swimmer at low Reynolds number, Soft Matter **5**, 472 (2009).

[65] G. Grosjean, M. Hubert, G. Lagubeau, and N. Vandewalle, Realization of the najafi-golestanian microswimmer, Phys. Rev. E **94**, 021101 (2016).

[66] F. Box, E. Han, C. R. Tipton, and T. Mullin, On the motion of linked spheres in a Stokes flow, Exp. Fluids **58**, 29 (2017).

[67] S. Klumpp, C. T. Lefèvre, M. Bennet, and D. Faivre, Swimming with magnets: From biological organisms to synthetic devices, Phys. Rep. **789**, 1 (2019).

[68] A. W. Hauser, A. A. Evans, J.-H. Na, and R. C. Hayward, Photothermally reprogrammable buckling of nanocomposite gel sheets, Angew. Chem. Int. Ed. **54**, 5434 (2015).

[69] H.-W. Huang, M. S. Sakar, A. J. Petruska, S. Pané, and B. J. Nelson, Soft micromachines with programmable motility and morphology, Nat. Commun. **7**, 12263 (2016).

[70] C. Ohm, M. Brehmer, and R. Zentel, Liquid crystalline elastomers as actuators and sensors, Adv. Mater. **22**, 3366 (2010).

[71] H. Zeng, P. Wasylczyk, C. Parmeggiani, D. Martella, M. Burresi, and D. S. Wiersma, Light-fueled microscopic walkers, Adv. Mater. **27**, 3883 (2015).