CrossMark

J. Fluid Mech. (2016), *vol.* 789, *pp.* 726–749. © Cambridge University Press 2016
doi:10.1017/jfm.2015.686

Learning to school in the presence of hydrodynamic interactions

M. Gazzola[1], A. A. Tchieu[2], D. Alexeev[3], A. de Brauer[4] and P. Koumoutsakos[3,†]

[1]School of Engineering and Applied Sciences, Harvard University, MA 02138, USA

[2]SPACEX, Hawthorne, CA 90250, USA

[3]Chair of Computational Science, ETH Zürich, CH-8092, Switzerland

[4]ENSEIRB-MATMECA Bordeaux Graduate School of Engineering, Talence, France

(Received 4 April 2015; revised 27 October 2015; accepted 16 November 2015; first published online 26 January 2016)

Schooling, an archetype of collective behaviour, emerges from the interactions of fish responding to sensory information mediated by their aqueous environment. A fundamental and largely unexplored question in fish schooling concerns the role of hydrodynamics. Here, we investigate this question by modelling swimmers as vortex dipoles whose interactions are governed by the Biot–Savart law. When we enhance these dipoles with behavioural rules from classical agent-based models, we find that they do not lead robustly to schooling because of flow-mediated interactions. We therefore propose to use swimmers equipped with adaptive decision-making that adjust their gaits through a reinforcement learning algorithm in response to nonlinearly varying hydrodynamic loads. We demonstrate that these swimmers can maintain their relative position within a formation by adapting their strength and school in a variety of prescribed geometrical arrangements. Furthermore, we identify schooling patterns that minimize the individual and collective swimming effort, through an evolutionary optimization. The present work suggests that the adaptive response of individual swimmers to flow-mediated interactions is critical in fish schooling.

Key words: biological fluid dynamics, flow control, swimming/flying

1. Introduction

Schooling, encountered in over ten thousand species (Shaw 1978), is believed to provide several advantages to fish (Partridge 1982), including protection and defence against predators (Major 1978; Shaw 1978; Landeau & Terborgh 1986), enhanced foraging (Pitcher, Magurran & Winfield 1982) and mating success (Barnes & Hughes 1988). It is also plausible that fish benefit from increased hydrodynamic efficiency (Weihs 1973). Understanding the governing mechanisms in fish schooling and exploiting them for rational engineering designs (Whittlesey, Liska & Dabiri 2010) requires that we elucidate the interplay between social and hydrodynamic interactions among swimmers.

† Email address for correspondence: petros@ethz.ch

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

Investigations in natural and laboratory settings have provided invaluable insight into behavioural aspects of fish schooling. At the same time, little quantitative information is available regarding the role of hydrodynamics in collective swimming. One might expect that with today's computing capabilities it should be possible to acquire such information by numerical simulations. However, while schooling can be readily observed in natural swimmers, in simulations it is essential to equip the individuals with an appropriate behavioural model to achieve such group dynamics. Agent-based models (Reynolds 1987) that lead to schooling or flocking rely on local interaction rules handcrafted *a priori* based on empirical arguments and experimental observations (Couzin *et al.* 2002, 2005; Hubbard *et al.* 2004; Viscido, Parrish & Grunbaum 2005). These models have been a key tool in helping us to understand the influence of social traits in the emergence of schooling patterns (Aoki 1982; Huth & Wissel 1992; Niwa 1994; Vicsek *et al.* 1995). However, they do not explicitly account for the flow environment. We consider this a limitation especially in the case of large, tightly packed fish assemblies. In fact, a natural swimmer that wishes to adapt its speed and orientation to satisfy local interaction rules (e.g. move with the average velocity of its neighbours) needs to translate this into specific body gaits. These actions perturb the flow field, which in turn affects the dynamics of the neighbours.

It is also important to distinguish between self-propelled swimmers and swimmers that are towed with a specified velocity through the flow field (Jiang, Osborn & Meneveau 2002), as is usually implied in agent-based models. In the case of a towed swimmer, if hydrodynamics is included, it affects only the energy expenditure for towing the swimmer with the specified velocity, but it does not influence the swimmer's dynamics or trajectory. A self-propelled swimmer instead has to adjust its gait to compensate for nonlinearly varying hydrodynamic loads in order to propel itself in a desired direction. As fish rely on self-propulsion, it is essential to capture this trait together with the long-range fluid coupling. To the best of our knowledge, such hydrodynamic interactions have not been included in agent-based models of swimming. Hence, fundamental questions such as how fish respond to each other's wakes and to what extent schooling is the result of their synthesized vortex field or their social traits remain largely unanswered.

Swimmers influence their flow environment, which in turn affects the dynamics of the individuals at all scales. In the Stokes flow regime, it has been noted that the collective motion of microorganisms induces flow coupling that leads to transitions from ordered to disordered patterns (Brady & Bossis 1988; Brady *et al.* 1988; Aditi Simha & Ramaswamy 2002; Hatwalne *et al.* 2004). Both in the inviscid limit and at finite Reynolds numbers, recent works have demonstrated that specific body motions can propel initially stationary neighbours (Tchieu, Crowdy & Leonard 2010; Gazzola *et al.* 2012*a*), while models of rotating discs at finite Reynolds numbers have been shown to lead to the emergence of patterns (Goto & Tanaka 2015). Experimental observations indicate that some fish species arrange themselves in diagonal formations (Cullen, Shaw & Baldwin 1965; Partridge *et al.* 1980), and it has been suggested (Breder 1965; Weihs 1973; Tsang & Kanso 2013) that fish in diamond configurations can exploit the vorticity created by their neighbours to decrease their energy expenditure. This hypothesis relies on stable, periodic fish arrangements, prescribed gait and unperturbed or minimally perturbed flow conditions. At intermediate and large Reynolds numbers, the flow field synthesized by the vorticity shed by multiple swimmers (Abrahams & Colgan 1987; Liao *et al.* 2003; Weihs 2004; Ristroph & Zhang 2008) is noisy, and varying loads are induced on the swimmers depending on their relative location (Gazzola *et al.* 2011*a*). Key questions

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

remain unanswered: How can swimmers overcome this noisy environment to achieve specific behavioural or physical goals? To what extent do flow-mediated interactions affect decision-making and group behaviour?

Mathematical models have been used over the past 50 years to study fish swimming. Models of individual swimmers were pioneered by Taylor (1952), Saffman (1967) and Ted Wu (1971). More recent works have focused on models of swimmers and flyers corresponding to inviscid (Kanso & Newton 2009; Paoletti & Mahadevan 2011) and highly viscous flows (Zhu, Lauga & Brandt 2012). Such models have also been extended to account for hydrodynamic interactions between the individual swimmers (Ishikawa, Simmonds & Pedley 2006; Nair & Kanso 2007; Koch & Subramanian 2011; Lin, Thiffeault & Childress 2011; Tsang & Kanso 2013). These works also discuss the stability of various swimmer configurations and their correspondence to schooling patterns observed in nature. Here, following Tchieu, Kanso & Newton (2012), swimmers are modelled as self-propelled, finite-width dipoles capable of accelerating, decelerating and turning. The novelty of the present work lies in equipping the individual swimmers with adaptive decision-making to adjust their gaits, thus enabling them to robustly form various schooling configurations in the presence of long-range hydrodynamic interactions.

This work is inspired by the concept of vortobots (Park, Noca & Koumoutsakos 2005). Vortobots were envisioned as thin solid discs equipped with a simple rotating mechanism so that they move in swarms. Vortobots that move without change in shape and size would correspond to vortex crystals (Aref *et al.* 2003). However, such configurations are not robust with respect to noise and cannot be generalized to large numbers of vortices. One might argue that *a priori* control rules such as those encountered in agent-based models could maintain stable configurations. For the present dipole swimmer models, we show that the use of non-adaptive *a priori* defined local interaction rules does not robustly allow swimmers to maintain finite-size schooling formations, causing them to diverge from one another or to collide. In turn, we show that swimmers can learn, through a reinforcement learning algorithm (Sutton & Barto 1998), to dynamically adjust their swimming actions in response to flow-mediated interactions so as to swim in arbitrary finite-size schooling arrangements. Furthermore, we identify schooling arrangements that minimize collective swimming effort, via an evolutionary optimization technique. Finally, the relative effort of swimmers distributed within an optimal school is investigated.

Our study highlights the importance of accounting for the hydrodynamic environment in collective dynamics, and outlines a rigorous approach to identifying optimal adaptive action policies so as to respond to flow-mediated interactions.

2. Learning optimal behaviour in a fluid-mediated environment

We examine the collective behaviour of model dipole swimmers. In order to control their velocity and bearing, the swimmers can adapt their dipole strengths; in this way they affect the environment and, in turn, the dynamics of all other swimmers. In contrast to classical agent-based models (Aoki 1982; Huth & Wissel 1992; Niwa 1994; Vicsek *et al.* 1995), besides including hydrodynamics we do not specify *a priori* local interaction rules. Instead, these are automatically identified by a reinforcement learning algorithm.

The dynamics of a system of N swimmers immersed in an inviscid flow is represented by a low-order model referred to as 'finite-width dipole' (Tchieu *et al.*

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

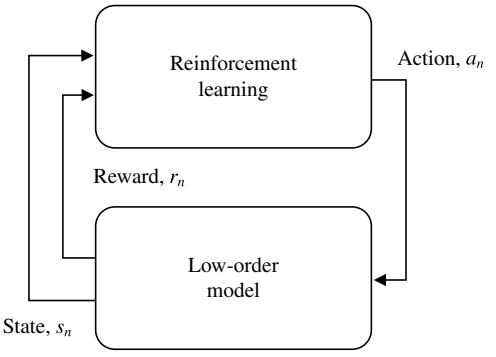


FIGURE 1. Schematic of reinforcement learning coupled with a low-order model for swimmers in an inviscid flow. The goal of the nth agent is encoded in the numerical reward r_n, and the agent learns, through a trial-and-error process called reinforcement learning (Sutton & Barto 1998), how to map states s_n into actions a_n so as to maximize the long-term reward (§ 2.4). The system is simulated via a low-order model that takes into account swimmer–swimmer dynamics mediated by an inviscid fluid medium (Tchieu *et al.* 2012).

2012). The goal of the swimmers is to swim coherently in a prescribed formation, avoiding collisions or dispersion. This is a necessary preliminary step to allow for the hydrodynamic characterization of different swimming formations in terms of energy expenditure (§ 2.5). Given the swimmers' repertoire of possible actions and sensorial representation of the environment (referred to as states), reinforcement learning (Sutton & Barto 1998) allows them, through trial and error, to discover an optimal behavioural policy (i.e. a mapping between states and actions) to maintain their relative positions within the school. Each loop in figure 1 represents a single learning instance where all agents use their learned policy to select an action which alters their state through the modelled dynamics. The reward associated with the new state aids the agents in improving their policy, which eventually converges to an optimal policy.

2.1. *A finite-width dipole model for hydrodynamically interacting swimmers*

The flow field generated by individual natural swimmers possesses a complex signature that is greatly affected by the swimmers' gait, morphology and size as well as by viscous and three-dimensional effects. The characterization of the group dynamics of hundreds of swimming bodies that resolves this level of detail is computationally beyond reach at present. Therefore, we study swimmers modelled as finite self-propelling dipoles (figure 2*a*,*b*) immersed in an inviscid, unbounded and incompressible flow (Tchieu *et al.* 2012). This model reflects the fact that the far field associated with a self-propelled undulating body is dipolar to leading order (Wolfgang *et al.* 1999). The finite dipole model represents a drastic idealization of a swimmer since it abstracts from morphological and kinematic traits, is massless and therefore disregards the inertia of a solid body and does not account for three-dimensional and viscous effects such as separation and vortex shedding. Nevertheless, it does capture to first order the flow coupling among self-propelled bodies that are sufficiently spaced apart (more than one body length as estimated in Tchieu *et al.* (2012)), and it is computationally effective. Bearing in mind its limitations, the dipole model is

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

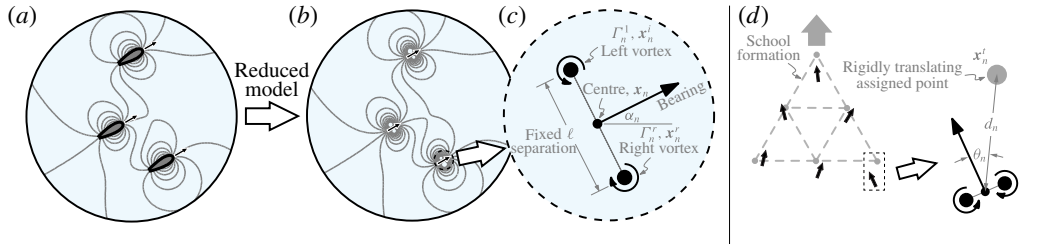


FIGURE 2. (Colour online) Streamlines of (*a*) inviscid swimmers translating in a potential flow and (*b*) the finite dipole approximation of the swimmers (Tchieu *et al.* 2012). Shown in (*c*) is the detailed view of a single finite dipole, while (*d*) illustrates the state representation of an individual dipole swimmer attempting to follow a translating lattice point (§ 2.4).

instrumental to qualitative computational investigations of fish schooling. Finally, we emphasize that self-propelled agents are distinct from agents that are towed with a certain velocity. In agent-based models the latter are usually employed, but these do not correspond to self-propelled animals responding to nonlinear interactions with the flow field.

The basis of this low-order model is depicted in figure 2. In a system of N dipole swimmers, a dipole located at $\boldsymbol{x}_n$ ($n = 1, 2, \ldots, N$) is decomposed into two vortices located at $\boldsymbol{x}_n^l$ and $\boldsymbol{x}_n^r$ with circulation strengths Γ_n^l and Γ_n^r, separated by a constant distance ℓ. The dipole swimmer travels with a bearing defined by α_n, as depicted in figure 2(*c*). Each finite dipole swimmer can change its vortex strength as a means of controlling its bearing and speed. Following Tchieu *et al.* (2012), the equations of motion of N self-propelled interacting finite dipoles are modified to allow each dipole to change its individual bearing and speed while simultaneously affecting the flow. Note that this model is different from the one used by Chate *et al.* (2008) in that the generated flow field affects the bearing of swimmers and that the swimmers directly change the flow field when performing actions. We also note that the value ℓ can be related to a characteristic width D of the swimmer by matching the far-field dipolar strength of a body moving in an inviscid fluid to that of a finite dipole, resulting in $\ell = D/(2\sqrt{2\pi})$. To proceed, a point $\boldsymbol{x}$ is mapped to the complex z-plane such that $\boldsymbol{x} = (x, y) \mapsto z = x + \mathrm{i}y$, where $\mathrm{i} = \sqrt{-1}$. Therefore, given the position of a dipole $\boldsymbol{x}_n \mapsto z_n$, its two vortices of strengths Γ_n^l and Γ_n^r, separated by a constant distance ℓ, are located at

$$z_n^l = z_n + \frac{\mathrm{i}\ell \mathrm{e}^{\mathrm{i}\alpha_n}}{2} \quad \text{and} \quad z_n^r = z_n - \frac{\mathrm{i}\ell \mathrm{e}^{\mathrm{i}\alpha_n}}{2}, \tag{2.1a,b}$$

respectively (see figure 2). The equations of motion derived from Tchieu *et al.* (2012) that govern the system of N finite dipoles are modified to read

$$\dot{\bar{z}}_n = \frac{(\Gamma_n^l + \Gamma_n^r)\mathrm{e}^{-\mathrm{i}\alpha_n}}{4\pi\ell} + \frac{w_n^o(z_n^l) + w_n^o(z_n^r)}{2}, \tag{2.2a}$$

$$\dot{\alpha}_n = \frac{\Gamma_n^l - \Gamma_n^r}{2\pi\ell^2} + \frac{\mathrm{Re}[(w_n^o(z_n^r) - w_n^o(z_n^l))\mathrm{e}^{\mathrm{i}\alpha_n}]}{\ell}, \tag{2.2b}$$

where $\mathrm{Re}[\cdot]$ denotes the real part of a complex expression and

$$w_n^o(z) = \sum_{j \neq n}^{N} \frac{1}{2\pi\mathrm{i}} \left(\frac{\Gamma_j^l}{z - z_j^l} - \frac{\Gamma_j^r}{z - z_j^r} \right) \tag{2.3}$$

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

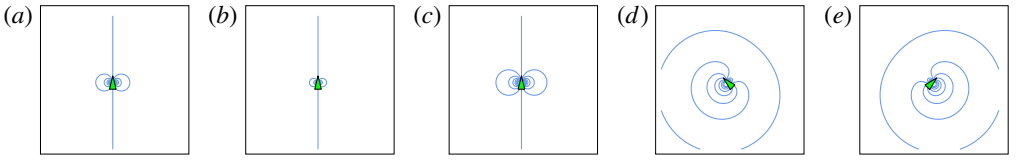


FIGURE 3. (Colour online) The set of five actions available to each dipole swimmer (represented by a triangle of width $D = 2\ell\sqrt{2\pi}$ pointing in the direction of travel), with each action mapped to the integer set $a_n = \{1, 2, 3, 4, 5\}$: (*a*) travelling straight at the nominal speed v^0 ($a_n = 1$); (*b*) travelling straight at a slower speed v^- ($a_n = 2$); (*c*) travelling straight at a faster speed v^+ ($a_n = 3$); (*d*) making a left turn at a specified turn radius ρ_T ($a_n = 4$); and (*e*) making a right turn with turn radius ρ_T ($a_n = 5$). In each panel we also show the streamlines demonstrating how a swimmer affects the background flow field in the absence of all other dipole swimmers.

is the interaction term due to all other dipoles in the environment. We emphasize that the first terms of (2.2*a*) and (2.2*b*) correspond to the dipole self-induced velocity and bearing rate when no other swimmers or background flow are present. Therefore, these terms relate to the ability of an agent to affect its own speed and bearing. Moreover, it is worth noting that the swimming speed is directly proportional to the dipolar circulation strength and inversely proportional to the swimmers' size D through ℓ. This is consistent with the fact that for a fixed amount of circulation, a small fish swims faster than a large one.

We stress the fact that the dipole model allows us to evolve the system in time by actually solving the Euler equations for an incompressible inviscid flow, and so the presence of a liquid environment is not simplistically modelled through *ad hoc* local interaction rules. The major advantage of this formulation is that it provides a neat distinction between social and hydrodynamic effects, unlike previous modelling approaches.

2.2. *Swimming gaits and manoeuvring through circulation change*

We extend the original finite dipole model to equip each swimmer with a set of gaits or actions, as depicted in figure 3. A swimmer can travel forward at three distinct speeds, v^0 (nominal speed), $v^- = v^0 - v^A$ (slow) and $v^+ = v^0 + v^A$ (fast), and can turn left or right with turn radius ρ^T while travelling at speed v^0. These actions are realized by allowing each dipole swimmer to instantaneously adjust its vortex circulations Γ_n^l and Γ_n^r. The five actions, mapped to integer values, adjust the circulation according to the rule

$$\text{if} \quad \begin{cases} a_n = 1\text{:} & \text{travel straight at } v^0 & \longrightarrow & \Gamma_n^l = \Gamma_n^r = \Gamma^0, \\ a_n = 2\text{:} & \text{travel straight at } v^- & \longrightarrow & \Gamma_n^l = \Gamma_n^r = \Gamma^0 - \Gamma^A, \\ a_n = 3\text{:} & \text{travel straight at } v^+ & \longrightarrow & \Gamma_n^l = \Gamma_n^r = \Gamma^0 + \Gamma^A, \\ a_n = 4\text{:} & \text{turn left with radius } \rho^T & \longrightarrow & \Gamma_n^l = \Gamma^0 + \Gamma^T, \quad \Gamma_n^r = \Gamma^0 - \Gamma^T, \\ a_n = 5\text{:} & \text{turn right with radius } \rho^T & \longrightarrow & \Gamma_n^l = \Gamma^0 - \Gamma^T, \quad \Gamma_n^r = \Gamma^0 + \Gamma^T, \end{cases} \tag{2.4}$$

where $\Gamma^0, \Gamma^A, \Gamma^T > 0$. The nominal vortex strength Γ^0 is related to the cruise velocity v^0 and the characteristic size ℓ through $\Gamma^0 = 2\pi\ell v^0$.

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

Behaviour	Zone of repulsion	Zone of orientation	Zone of attraction
Highly parallel	1	12	15
Dynamically parallel	1	6	10

TABLE 1. Radii of zones of interaction relative to the 'dynamically parallel school' and 'highly parallel school' behaviours as detailed in Couzin *et al.* (2002); the radii are normalized by ℓ.

The additional circulations $\pm\Gamma^A$ and $\pm\Gamma^T$ due to travelling fast or slow and turning right or left are fixed by the swimming parameters v^0, v^A and ρ^T. These values are related to the nominal circulation by

$$\Gamma^A = \left(\frac{v^A}{v^0}\right)\Gamma^0, \tag{2.5a}$$

$$\Gamma^T = \left(\frac{\rho^T}{\sqrt{2\pi}\ell}\right)^{-1}\Gamma^0. \tag{2.5b}$$

The change in circulation in turn modifies the flow field and thus influences all swimmers in the system. Note that these actions are exclusive, i.e. a swimmer can select only one action at a time.

We emphasize that the use of five actions is a simplification of the movements of naturally occurring swimmers, which are characterized by a large number of kinematic degrees of freedom, so that swimmers have the ability to fine tune their gaits in response to environmental cues. However, the use of a small number of actions drastically reduces the computational costs associated with identifying an optimal behavioural policy through reinforcement learning, hence our choice to equip the agents with a limited repertoire of gaits.

2.3. *The Aoki–Couzin behavioural model with hydrodynamics interactions*

We examine how the classical Aoki–Couzin model (Aoki 1982; Couzin *et al.* 2002) with *a priori* defined interaction rules would perform in the presence of hydrodynamics. In particular, we considered the so-called 'dynamically parallel school' and 'highly parallel school' behavioural rules, as detailed in (Couzin *et al.* 2002).

In the Aoki–Couzin model, collective behaviour emerges due to three *a priori* specified rules among agents: each agent tries to avoid collision with neighbours, aligns to the moving direction of the agents contained in a larger neighbourhood and, finally, tends to approach the agents of an even larger neighbourhood. Given these three rules, each agent first computes its desired direction and then turns in accordance with that direction (Couzin *et al.* 2002). By varying the size of the interaction regions, qualitatively different behaviours can be observed. In table 1, we summarize the radii characteristic of the 'dynamically parallel school' and 'highly parallel school', as detailed in Couzin *et al.* (2002) and used here.

In order to cast the Aoki–Couzin model in the present dipole framework, each dipole agent determines its desired direction $\alpha_{desired}$ by following the specifications of

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

Couzin *et al.* (2002), and then adjusts its vortex circulations as follows:

$$\left.\begin{aligned} \dot{\alpha}_{desired} &= \frac{\alpha_{desired} - \alpha_0}{\tau} \\ \Gamma_{add} &= \pi \ell^2 (\dot{\alpha}_{desired} - \dot{\alpha}_0) \\ \Gamma^l &= \Gamma^l + \Gamma_{add} \\ \Gamma^r &= \Gamma^r - \Gamma_{add} \end{aligned}\right\} \tag{2.6}$$

where τ, α_0 and $\dot{\alpha}_0$ are, respectively, the simulation's time step and the agent's current bearing and bearing rate. Computing the new Γ^r and Γ^l corresponds to adapting the gaits to match exactly the desired bearings, assuming the absence of all other swimmers. Once the circulations of the individual dipoles are determined, the system of N swimmers is evolved by accounting for hydrodynamic coupling via the governing equations (2.1)–(2.3).

The behaviour of this model in the presence of hydrodynamic interactions is examined in § 3.2.

2.4. *Reverse engineering of dynamic interaction rules via reinforcement learning*

Swimmers are modelled as finite-width dipoles with varying circulation strength. Their presence and actions affect the flow field and, in turn, all other swimmers. Because of this highly nonlinear coupling, it is virtually impossible to handcraft local interaction rules that allow the dipoles to coherently swim in any predefined, finite-size schooling arrangement. Therefore agent-based models, with *a priori* defined rules, cannot help us to assess the hydrodynamic properties of different schooling configurations. With this in mind, we employ a reverse engineering approach to obtain the interaction rules among swimmers. We specify for the agents the goal of maintaining a given geometric arrangement and employ a reinforcement learning technique to identify an appropriate interaction policy. This approach relies on four key components: the reward that encodes the agent's goal; the state that formalizes what the dipole can sense of the surrounding environment; the actions, i.e. the repertoire of gaits at each swimmer's disposal (§ 2.2); and, finally, a learning strategy based on trial and error.

In this study we employ a particular reinforcement learning technique, namely the one-step Q-learning algorithm (Sutton & Barto 1998). Besides its algorithmic simplicity, Q-learning has been proven to converge to an optimal behavioural policy for finite Markov decision processes (Watkins & Dayan 1992). In this setting, the swimming agent explores the environment and its experience is represented by the tuple $\langle s_n, a_n, r_n, s'_n \rangle$, where s'_n is the next state given the action a_n taken from the current state s_n and r_n is the corresponding reward. An agent estimates by trial and error the action-value function $Q_n(s_n, a_n)$, i.e. the expected long-term reward for taking action a_n given the state s_n (a schematic of this approach is depicted in figure 1). The action-value function Q_n can be understood as a table or a matrix in which for every state–action entry the corresponding expected reward (estimated through the reward history) is stored. This table is consulted by the agent whenever an action has to be taken, and it is continuously updated as the system evolves. Therefore Q_n encodes the swimmers' adaptive decision-making intelligence, and the corresponding behaviour is determined by choosing from Q_n, with probability $1 - \epsilon$, the best action a_n such that $a_n \to \max_{a_n} Q_n(s_n, a_n)$ from the current state s_n (an ϵ-greedy selection scheme). The ϵ-probability of choosing a non-optimal action allows the agent to explore new state–action (s_n, a_n) pairs (Sutton & Barto 1998). Therefore, reinforcement learning

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

intrinsically accounts for noise through ϵ, which can be related to the noise of natural schooling systems. Here, we use a shared policy approach among all swimmers to accelerate the learning process; thus $Q_n = Q$ and all agents update Q based on their personal experience. At every learning time interval δt, the swimmers update the action-value function according to $Q(s_n, a_n) = Q(s_n, a_n) + \varphi(\Delta Q)_n$ for $n = 1, \ldots, N$, where $0 \leqslant \varphi \leqslant 1$ is the learning rate and $(\Delta Q)_n = r_n + \gamma \max_{a_n} Q(s'_n, a_n) - Q(s_n, a_n)$, with $0 \leqslant \gamma < 1$ being the discount parameter which corresponds to the weight given to past experiences. We emphasize that learning individual policies, as opposed to the shared approach employed here, may allow agents finer behavioural tuning. For example, in the case of schooling, swimmers may adapt their policies depending on their location within the group. However, it has been shown empirically that the use of a shared policy reduces the time to convergence linearly with the number of agents (Gazzola 2013). In our study this entails a hundreds-fold reduction in computational cost, which is the rationale behind our choice of employing a shared approach. In the following paragraphs the definitions of reward, state and action are formalized.

Reward. Since ultimately we are interested in investigating the hydrodynamic properties associated with different schooling geometries, we must first have the dipoles learn to swim in a given formation. This is achieved by setting the goal of each swimmer as following a specified target point $\boldsymbol{x}_n^t$ in a predefined arrangement, as depicted in figure 2(*d*). This allows hundreds of self-propelled dipoles to learn how to swim coherently, a task that is beyond the reach of models based on handcrafted *a priori* interaction rules (in §§ 2.5 and 3.4 we detail how optimal arrangements can be obtained). The swimmer's goal is translated mathematically into a numerical reward signal. The numerical reward is chosen to reflect how well the dipole swimmer can follow its assigned target point while doing the minimum amount of manoeuvring so that $r_n = w_d(1 - d_n/\ell) + w_a \cdot \eta_a$, where $\eta_a = 0$ for travelling at v_0, $\eta_a = -1$ for accelerating and turning and $\eta_a = 1$ for decelerating. Weights are set to $w_d = 0.9$ and $w_a = 0.1$; thus $\max(r_n) = 1$. We note how the second term of r_n penalizes swimmers that take unnecessary actions, while favouring those that reduce their effort by slowing down.

State. Swimmers can sense their distance $d_n = |\boldsymbol{x}_n^t - \boldsymbol{x}_n|$ and orientation $\theta_n = \arg(\boldsymbol{x}_n^t - \boldsymbol{x}_n) - \alpha_n$ with respect to their assigned target point $\boldsymbol{x}_n^t$ within the school, as shown in figure 2(*d*). We stress the fact that the dynamics of a swimmer is not mapped onto a lattice. Swimmers are in fact free to move in the continuum two-dimensional space, while they adaptively adjust their gait in the attempt to maintain their relative position within the school. Instead, the quantities d_n and θ_n are each mapped into a set of $L = 30$ discrete states within the range $\Delta d = 10D$ and $\Delta\theta = 2\pi$ such that $s_n = \{\min(L, \max(0, \lfloor d_n L/\Delta d \rfloor)), \min(L, \max(0, \lfloor \theta_n L/\Delta\theta \rfloor))\}$. In total, each swimmer has a state space that consists of 900 states. The choice of a target point over the sensing of the neighbours dramatically reduces the state-space dimension (the curse of dimensionality), allowing us to tackle the problem computationally. Moreover, it enables the study of structured schooling arrangements while still capturing the influence of neighbours as they directly affect each other's dynamics through long-range hydrodynamic interactions.

Actions. Each dipole swimmer can perform five actions: it can travel forward at three distinct speeds, v^0 (its nominal travelling speed), $v^+ = 1.1v^0$ and $v^- = 0.9v^0$, or turn left or right with radius $r^T = 10\ell$ while travelling at v^0. These actions are realized by allowing the dipole swimmer to adjust its vortex circulations Γ_n^l and Γ_n^r accordingly

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

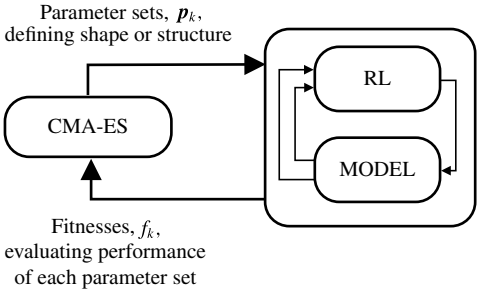


FIGURE 4. Schematic of CMA-ES optimizer coupled with reinforcement learning (RL) and the low-order model. The CMA-ES dispatches $k = 1, \ldots, p$ parameter sets $\boldsymbol{p}_k$ defining the school configuration. The reinforcement learning framework allows agents to learn to swim in the formation characterized by $\boldsymbol{p}_k$. In return, the CMA-ES receives a fitness f_k that captures the effectiveness of collective swimming relative to each school $\boldsymbol{p}_k$.

(§ 2.2). Since for every state five different gaits are available, the overall state–action space has 4500 entries, to be stored in a table representing the action-value function $Q(s_n, a_n)$.

An example further illustrating the working principles of reinforcement learning is reported in § 3.1.

2.5. *Optimization of schooling patterns via evolution strategies*

In the previous sections we established the algorithmic framework that allows self-propelled swimmers to learn how to keep their relative position within a school. We now look for optimal swimming configurations according to a desired metric (cost function), using the 'covariance matrix adaptation evolutionary strategy' (CMA-ES) of Hansen, Muller & Koumoutsakos (2003). The CMA-ES has been proven effective in a number of fluid mechanics and biological problems, from the optimization of gait and morphology in swimmers (Kern & Koumoutsakos 2006; Gazzola, van Rees & Koumoutsakos 2012*b*; van Rees, Gazzola & Koumoutsakos 2013, 2015) to the identification of virus traffic mechanisms (Gazzola *et al.* 2009).

The CMA-ES is a stochastic optimization algorithm that samples at each generation p parameter vectors from a multivariate Gaussian distribution $\mathcal{N}$. Here each parameter vector encodes the geometric configuration of a schooling arrangement (appendix A). The covariance matrix of the distribution $\mathcal{N}$ is then adapted based on successful past schools, chosen according to their corresponding cost function value f. In the present context, the CMA-ES evolves schooling configurations based on a metric of swimming effectiveness. In order to evaluate the cost function, i.e. the performance, of each school geometry, the dipole swimmers must first learn through reinforcement learning how to swim coherently in that specific arrangement. Then, after a learning period $\Delta T_{learning}$, the average cost function is evaluated by simulating the school for the time ΔT_{eval} (appendix A). The value f so computed is then returned to the optimizer, which uses it to select the best configurations and produce a new, more performant generation of school arrangements, until convergence to an optimal solution is achieved. This process is depicted in figure 4, while the definition of the cost function is given in the next paragraph.

Cost function. Here the cost function to be minimized implements a metric of swimming effort for the entire school. The effort of an individual swimmer relative

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

to its nominal cruise effort, i.e. its additional circulation expenditure, is defined by

$$\Delta\Gamma_n^e = \frac{1}{\Delta T_{eval}} \int_t^{t+\Delta T_{eval}} (\underbrace{|\Gamma_n^l - \Gamma_n^r|}_{\text{turning effort}} + \underbrace{\Gamma_n^l + \Gamma_n^r}_{\text{forward effort}} - 2\Gamma^0)\, \mathrm{d}t, \tag{2.7}$$

where Γ^0 is the nominal strength of each vortex. Hence, the cost function for the entire collection of swimmers is defined as $f = 1/(N\Gamma^0) \sum_{n=1}^{N} \Delta\Gamma_n^e$. The change in circulation can be related to the production of vorticity involved in accelerations or turning manoeuvres of the swimmers and, therefore, to the swimming effort. We also note for future reference that $f = 0$ corresponds to cruise swimming of isolated dipoles.

3. Results

3.1. *Learning process for an individual dipole swimmer*

Despite the formalism, the working principles of reinforcement learning are rather simple. We illustrate them here with the aid of a simple but representative problem, before proceeding further.

We consider a dipole whose goal is to follow a prescribed trajectory. The trajectory is represented as a target point $\boldsymbol{x}^t$ that moves by alternating straight runs and random turns of fixed radius ρ^T. The dipole is aware at all times of its own bearing α and position $\boldsymbol{x}$ as well as of the target position $\boldsymbol{x}^t$. At regular intervals δt, the agent is faced with the problem of choosing whether to turn left or right, accelerate, slow down or keep straight in order to accurately follow the target point. The dipole has not been instructed how to act given its relative position to the target, i.e. no *a priori* local rules are enforced. Instead, at every time step the swimmer estimates its relative distance d and orientation θ with respect to $\boldsymbol{x}^t$, as illustrated in figure 2(*d*). This is equivalent to determining the state s, i.e. the agent's current situation. Since the intelligence or behaviour of the swimmer is encoded as a multidimensional table or matrix, the continuous values of d and θ are discretized into a number of integer values, as described in § 2.4. Once the state matrix entry is determined, the agent can consult the expected rewards stored in the matrix associated with taking each of the five aforementioned actions a. These values $Q(s, a)$ are constantly updated by the dipole and represent its past experience; initially they are all set to zero. At this point the agent chooses with probability $1 - \epsilon$ the best action, i.e. the one with the largest $Q(s, a)$ value, and after pursuing it the new distances (d', θ') from the target are estimated, defining the new state s'. Moreover, based on d', the reward r is assigned to the dipole. The policy is then improved by discounting the old estimate of the expected reward $Q(s, a)$ and complementing it with the new information r according to the update rule described in § 2.4. This process is repeated indefinitely until convergence to an optimal behavioural policy.

The evolution of the swimmer's reward over time is shown in figure 5(*a*), and examples of the agent trajectories during the learning process are given in figure 5(*b–d*). The process of improving the policy can be seen in figure 5(*a*), where, after an initial transient, the agent progressively learns how to maximize its numerical reward by accurately following $\boldsymbol{x}^t$. In figure 5(*b*), the swimmer fails rather quickly, as demonstrated by its path diverging from the target path at the first turn. Subsequently, in figure 5(*c*), the swimmer follows adequately for a longer time before failing, and in figure 5(*d*) the swimmer learns a policy that allows it to follow a pseudo-random path indefinitely.

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

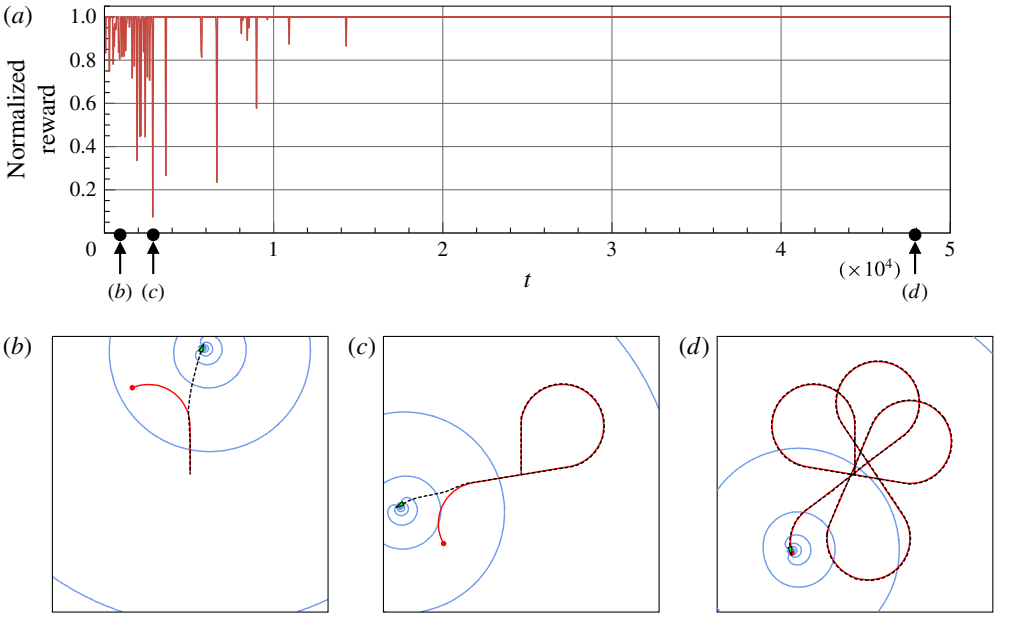


FIGURE 5. (Colour online) A single agent learns to optimally follow a moving target point $\boldsymbol{x}^t$. Shown in (*a*) is the time evolution of the agent's reward normalized by its maximum attainable value, based on the reward definition given in § 2.4. Panels (*b–d*) correspond, respectively, to non-adaptive, intermediate-adaptive and well-adaptive learning stages. The starting times for panels (*b–d*) are marked on the *x*-axis of (*a*). Black dashed and solid red lines correspond, respectively, to the trajectories of the agent (green triangle) and the target point (red dot). Instantaneous streamlines (blue lines) are displayed for reference. Simulations are run in the domain $[0, 1] \times [0, 1]$ with $\ell = 3 \times 10^{-3}$, $v^0 = 5\ell$, $v^A = 0.1v^0$, $\rho^T = 10\ell$, $\delta t = 0.1$, $\varphi = 0.01$, $\gamma = 0.98$ and $\epsilon = 0.01$. The notation is defined in § 2 and appendix A.

3.2. *Classical agent-based models versus learning agents in the presence of hydrodynamics*

We first show that prescribed schooling patterns, including the diamonds and squares which have been proposed as favourable schooling patterns (Weihs 1973), are not robustly maintained without an adaptive dynamic response of the swimmers to the flow field. In figure 6 we display the results of 16 swimming agents attempting to school in several formations initialized (at $t = 0$) as shown in figure 6(*a*,*d*,*g*). These initial patterns are characterized by diamond-like, square-like and random arrangements. With a prespecified forward swimming gait, the relative locations of swimmers will result in varying hydrodynamic loads, thus implying a dynamic rearrangement of the swimmers. Indeed, when the agents are assigned a specified swimming configuration, the simulated swimmers diverge from their relative positions and are prone to collisions with their neighbours due to flow-mediated interactions, which is consistent with the findings of Tchieu *et al.* (2012). As shown in figure 6(*b*) at $t = 80$, no collisions occur in the diamond-like configuration, but the swimmers are strained apart substantially. In figure 6(*e*) it can be seen that initializing with the square arrangement causes all agents to collide, while figure 6(*h*) shows that random configurations lead simultaneously to straining and collision

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

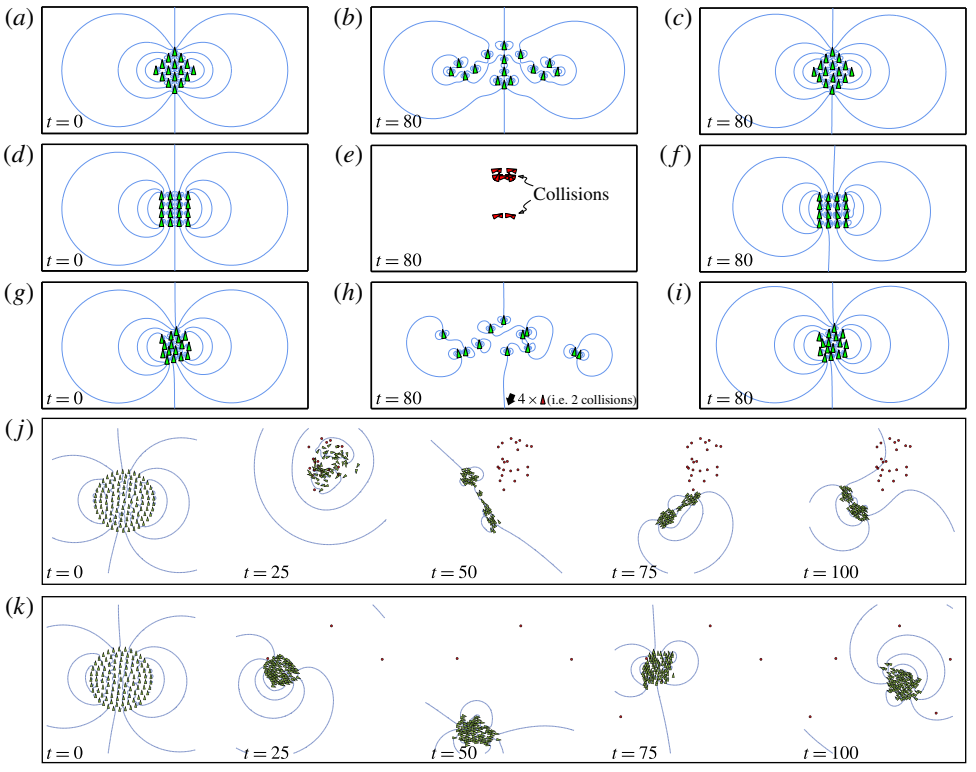


FIGURE 6. (Colour online) Swimming configurations for 16 dipole swimmers at specified times: (*a*,*d*,*g*) initial/desired configurations; (*b*,*e*,*h*) non-adaptive configurations; (*c*,*f*,*i*) adaptive (with policy learned from reinforcement learning) configurations. Red dipole swimmers have experienced a collision (after which they no longer move). Instantaneous streamlines (blue lines) are displayed for reference. Dipole swimmers are initialized on a diamond lattice in (*a–c*), a square lattice in (*d–f*) and randomly in a circular region of radius of 17.5ℓ in (*g–i*). In all initial configurations, a minimum interdipole spacing of 10ℓ is enforced. Also shown are the time evolutions of a school of 100 agents obeying (*j*) 'dynamically parallel group' and (*k*) 'highly parallel group' models (Couzin *et al.* 2002) enhanced with hydrodynamic interactions. Simulations are run in the domain $[0, 1] \times [0, 1]$ with $\ell = 5 \times 10^{-4}$, $v^0 = 5\ell$, $v^A = 0.1v^0$, $\rho^T = 10\ell$, $\delta t = 0.1$, $\varphi = 0.01$, $\gamma = 0.98$ and $\epsilon = 0.01$. The notation is defined in § 2 and appendix A.

effects. A qualitative hydrodynamic explanation for the disruption of square-like and diamond-like formations is provided in § 3.4 and figure 10. From figure 6(*c*,*f*,*i*) we see that schooling patterns can be maintained by swimmers through an adaptive modification of their swimming gaits that accounts for hydrodynamic interactions using a reinforcement learning algorithm. Therefore, through reinforcement learning the agents learn to adjust their swimming gaits to compensate for the varying hydrodynamic loads typical of a liquid environment. This can be related to the noisy environments and the corresponding response of swimmers in natural schooling systems.

We note that while the stability of diamond and square configurations has also been investigated by Tsang & Kanso (2013), our approach is fundamentally different. In fact, Tsang & Kanso (2013) considered infinite, doubly periodic lattices of dipoles

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

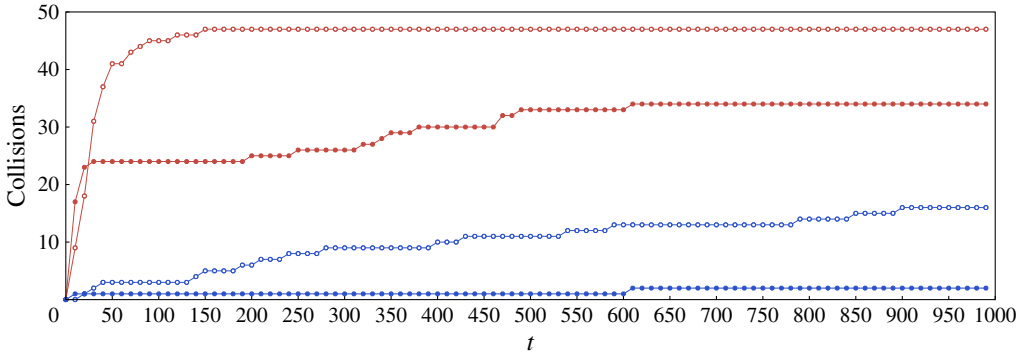


FIGURE 7. (Colour online) Comparison of the quantitative response of the Aoki–Couzin model with and without hydrodynamic interactions in terms of the number of agent collisions over time. A collision is detected whenever the distance between two agents is less than 2ℓ. Red and blue correspond, respectively, to the 'dynamically parallel group' and 'highly parallel group' behavioural rules (Couzin *et al.* 2002). Blank and solid circles refer, respectively, to simulations with and without hydrodynamics.

characterized by a single gait. The dipoles are then arranged so that the resulting flow field passively stabilizes the lattice, removing altogether the need to respond to varying loads by adapting swimming gaits, and the associated energetic costs. Here, the assumption of infinite schools is discarded in favour of a more realistic description. As a consequence, passive stabilization due to a stationary global flow field is no longer an option, hence the introduction of varying gaits and adaptive decision-making. Therefore, our approach complements the results of Tsang & Kanso (2013), allowing us to study the hydrodynamic and energetic features associated with arbitrarily shaped, finite-size schools.

We also investigated the dynamics of the Aoki–Couzin behavioural model (Aoki 1982; Couzin *et al.* 2002) in the presence of hydrodynamic interactions (§ 2.3). This model relies on *a priori* handcrafted local interaction rules among agents and does not explicitly account for the flow environment during the decision-making process. The time evolutions of a school of 100 agents obeying 'dynamically parallel group' and 'highly parallel group' behavioural rules (Couzin *et al.* 2002) are shown in figure 6(*j*,*k*). We find that the swimmers experience substantial straining and collisions, due to the hydrodynamic coupling. This behaviour is a drastic departure from the schooling patterns observed when employing the original models (Couzin *et al.* 2002). Indeed, the number of collisions increases by 40 % and 700 % in, respectively, the 'dynamically parallel group' and 'highly parallel group' models in the presence of hydrodynamics (figure 7). These findings emphasize the role of the environment, especially in a hydrodynamic setting in which all agents are doubly connected through the flow. The fact that the nonlinear response of the hydrodynamic system cannot be anticipated makes the definition of interaction rules by hand cumbersome, tedious and, ultimately, not robust.

Agent-based models with *a priori* specified rules, such as the ones considered herein, are characterized by a large parameter space (size of each zone, attraction and repulsion weights, time step etc.), and their results are known to be sensitive to these settings. In this study we have not explored the full parameter space and so cannot rule out the possibility that particular parameter combinations may allow

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

dipole swimmers to maintain structured arrangements or exhibit robust schooling dynamics. Nevertheless, the present investigation raises two key issues. First, the introduction of the flow environment modifies the dynamics associated with classical agent-based model settings. This implies that to reproduce the behaviour observed for a given instance of the Aoki–Couzin model, a new set of parameters has to be discovered. Since zone sizes and attractive and repulsive forces have a well-defined 'social' meaning, the presence of the fluid affects these quantities and alters the nature of social interactions. Therefore, the characterization of social traits cannot be prescinded from accounting for the environment. Second, the sensitivity of classical models to parameter settings supports the need for rigorous, automatic procedures for identifying local rules in the context of collective behaviour. Indeed, we have been unable to handcraft or derive through a direct optimization process any parameter set that would enable dipole swimmers to maintain a structured schooling formation, whereas this was readily achieved through the reinforcement learning framework. We reiterate that this finding does not represent a mathematical proof that handcrafted *a priori* models do not lead robustly to schooling, but it strongly highlights the need for computational methods that can guide the systematic exploration of the models' parameter space.

3.3. *Optimal internal structure of a school of dipole swimmers*

Since it has been suggested that diamond lattices embedded in an infinite school are energetically favourable (Weihs 1973; Tsang & Kanso 2013), we first characterize in terms of collective effort only the internal structure of schooling formations, disregarding edge effects related to the finite size of the school. The swimming effort of an individual is quantified by the variation of its dipole strength from a nominal value, $\Delta\Gamma_n^e$ (see § 2.5). This effort changes during swimming, according to the reinforcement learning algorithm, to overcome hydrodynamic noise. We therefore optimize the bulk school structure so as to minimize the cost function f defined as the linear sum of the efforts of all the swimmers in the group (§ 2.5), with $f = 0$ corresponding to cruise swimming of isolated swimmers. We consider three different parameterizations for generalized internal structure: (*a*) diamond, (*b*) rectangle and (*c*) hexagon configurations, as depicted in figure 8(*a–c*). The parameter β characterizes the angle between the swimming direction and the axis defining the lattice structure (see the insets of figure 8*a–c*). The hexagon formations constitute a subset of diamond formations, with the additional restriction that swimmers should be equidistant to their nearest neighbours. To minimize edge effects, we generate a circular shape and fill it with a lattice of $N \approx 200$ agents. Collective effort is evaluated only for the interior agents, and our criteria selects for the 50 agents closest to the centre of the school.

The starting configurations are shown in figure 8(*d–f*), while the corresponding optimal solutions are displayed in figure 8(*g–i*). The swimmers form striated patterns and get closely packed with one another in their travelling direction while separating as far as possible (given the bounds of the optimization search space) in the orthogonal direction. The packing in the direction of travel is limited by the capacity of the agents to stay in formation due to the strong flow-mediated interactions.

For the hexagon case (equidistant neighbours), from a starting configuration that is slightly detrimental to the school ($f_1 = 0.001$), the optimizer finds a configuration that offers no added benefit to the collective, as the fitness of the case shown in figure 8(*i*) is $f_{best} \approx 0$. We conclude that the constraint of equidistant swimmers does not allow

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

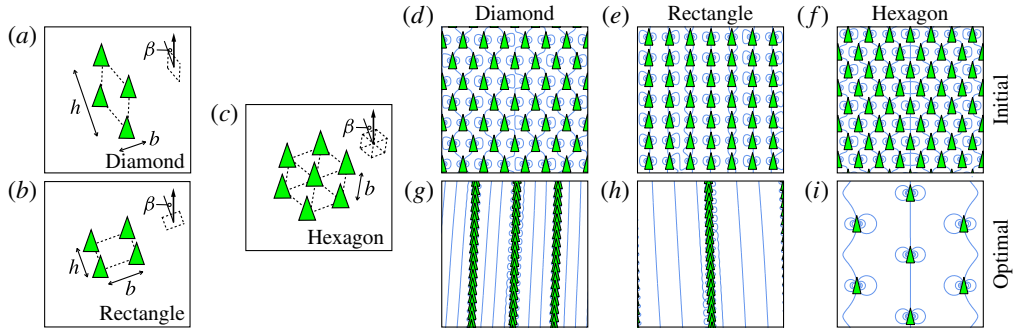


FIGURE 8. (Colour online) The three different parameterizations for the internal structure of the school: we choose to investigate (*a*) diamond configurations parameterized by $\boldsymbol{p} = \{b, h, \beta\}$, (*b*) rectangular configurations parameterized by $\boldsymbol{p} = \{b, h, \beta\}$ and (*c*) hexagon configurations parameterized by $\boldsymbol{p} = \{b, \beta\}$. In (*c*), all agents are equidistant to their nearest neighbours. The angle β represents the difference in direction of travel with respect to the axis defining the structure. Panels (*d–f*) show the initial guesses, where the fitnesses $f \approx 0$, and panels (*g–i*) show the best optimized solutions for the parameters of the internal structure defined according to (*a–c*). The population size is $p = 100$. Streamlines are shown in blue. The optimal parameters are (*g*) $\boldsymbol{p}_{best} = \{49.70\ell, 2.55\ell, 0.02\}$, (*h*) $\boldsymbol{p}_{best} = \{45.90\ell, 5.20\ell, -0.03\}$ and (*i*) $\boldsymbol{p}_{best} = \{37.25\ell, 0.52\}$, corresponding to the fitnesses $f_{best} = -0.175, -0.184$ and 0, respectively. Swimmers tend to form striated patterns in diamond and rectangular configurations; in the case where dipole swimmers are required to be equidistant from one another, there is no apparent hydrodynamic benefit from staying in the collective. Simulations are run in a $[0, 1] \times [0, 1]$ box with $\ell = 5 \times 10^{-4}$ (or $\ell = 1 \times 10^{-4}$), $v^0 = 5\ell$, $\rho^T = 10\ell$, $\delta t = 0.1$, $\varphi = 0.01$, $\gamma = 0.98$ and $\epsilon = 0.01$. The notation is defined in § 2 and appendix A. We note that the model relies on the assumption that swimmers are separated more than one characteristic length ℓ from one another (Tchieu *et al.* 2012). This condition is met here with the minimum distance of 2.55ℓ for (*g*), corresponding to the spacing b in our parameterization. This implies that swimmers could pack even tighter, but such a scenario is found to be suboptimal.

the agents to pack in the direction of travel and thus is not favourable to reducing circulation expenditure.

The results reported in this section may be viewed in the light of recent experiments that systematically investigated the thrust, power and efficiency performance of two side-by-side (Dewey *et al.* 2014) and in-line pitching airfoils (Boschitsch, Dewey & Smits 2014). In the side-by-side case (Dewey *et al.* 2014), it is found that the performances of individual airfoils are always anticorrelated, except for perfectly in-phase or out-of-phase actuation (which may not be realistic or robust in a schooling system). This arrangement entails an overall constant system efficiency and the generation of a net torque, which needs to be compensated for in stable schooling arrays. In the context of our study, these results suggest no foreseeable benefits associated with parallel swimming dipoles. Indeed, the striated patterns identified tend to minimize anticorrelation effects, by stretching lateral spacings as much as possible, effectively decoupling parallel dipoles. The case of in-line pitching airfoils is more complex. For small spacings $s/\ell < 0.5$ (where s is the linear coordinate in the direction of travel and ℓ is the airfoil chord), the performances of the leading and trailing airfoils are anticorrelated. For larger spacings, on the other hand, only the trailing body is affected and the vortex shedding from the leading airfoil plays a

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

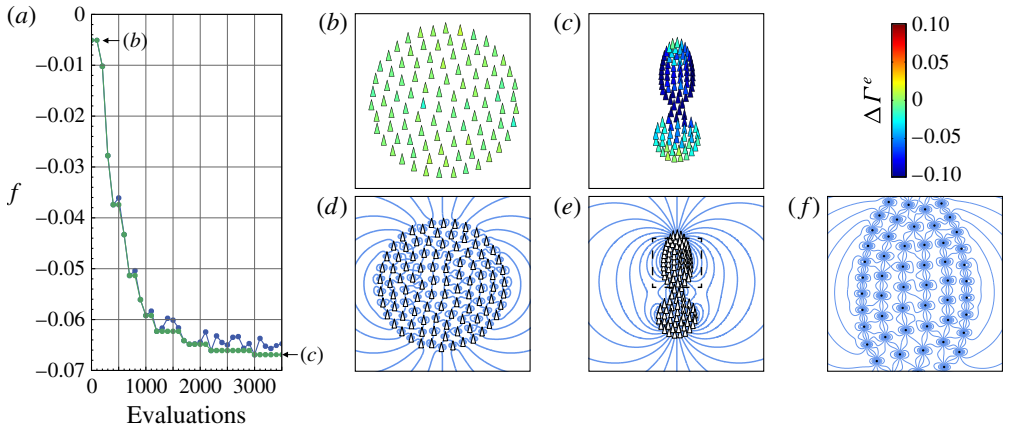


FIGURE 9. (Colour online) (*a*) Evolution of the cost function f versus the number of cost function evaluations. The population size per generation is $p = 100$; blue and green lines correspond, respectively, to the best solution in the current generation and the best solution ever. Also shown are the school formations corresponding to (*b*) the starting search point ($\boldsymbol{p}_1 = \{3,\ 1,\ 1,\ 1.57\}$, $f_1 = -0.005$) and (*c*) the best point ($\boldsymbol{p}_{best} = \{1.35,\ 0.5,\ 1.98,\ 2.42\}$, $f_{best} = -0.067$). The additional circulation change is computed for each swimmer and coloured accordingly. Simulations are run in the domain $[0, 1] \times [0, 1]$ with $\ell = 5 \times 10^{-4}$, $v^0 = 5\ell$, $\rho^T = 10\ell$, $\delta t = 0.1$, $\varphi = 0.01$, $\gamma = 0.98$ and $\epsilon = 0.01$. All configurations correspond to a simulation time of $t = 100$. The notation is as defined in § 2 and appendix A. Panel (*d*) shows streamlines of the starting search point and panel (*e*) the best solution configuration. Given in panel (*f*) is a detailed view of the dashed box in (*e*), with the dipole swimmers represented as black dots. As can be seen from (*f*), it is advantageous for the dipole swimmers to line up in striated patterns to aid in creating a favourable flow in the direction of travel. Indeed, the higher streamline density in this direction is the footprint of a stronger longitudinal velocity field. We note that the dipole model relies on the assumption that swimmers are separated more than one characteristic length ℓ from one another (Tchieu *et al.* 2012). This condition is met here with a minimum distance of 1.35ℓ (first entry of the parameter vector $\boldsymbol{p}_{best}$). This implies that swimmers could pack even tighter, but such a scenario is found to be suboptimal.

prominent role in the observed dynamics. Extrapolating to multiple airfoils, we might expect small spacings to be suboptimal. In fact, owing to the strong anticorrelation, every airfoil would experience both enhancing and disruptive effects. A larger spacing would instead allow, under the appropriate phase lag, all airfoils to benefit from flow coupling. The dipole model does not account explicitly for vortex shedding. As a consequence, there is no phase lag or cutoff separation distance that controls the interactions of the leading and trailing dipoles. Nevertheless, the optimization process discards configurations characterized by small spacing, since the strong anticorrelated dipole interaction poses control and learning problems and increases the swimming effort. Dipoles settle for larger spacing ($s/\ell \geqslant 2.55$ in figure 8, $s/\ell \geqslant 1.35$ in figure 9) to weaken anticorrelation effects, allowing for better stability and reduced effort, consistent with the above observations. We conclude that our findings, within the limits of our modelling approach, qualitatively capture the salient features associated with two-body swimming interactions (Boschitsch *et al.* 2014; Dewey *et al.* 2014).

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

3.4. *Optimal shape of a school of dipole swimmers*

We proceed by optimizing for the collective swimming effort f of the school configuration as a whole, where both internal structure and edge effects compete simultaneously. We initialize a school formation by arranging $N = 100$ swimmers in a prescribed shape so as to maximize the distance from each other as well as from the shape boundary (appendix A). The schooling arrangement is specified by four parameters $\boldsymbol{p} = \{k_1, k_2, \phi_1, d_{avg}\}$, where k_1, k_2 and ϕ_1 characterize its shape and d_{avg} determines its area, $A = N \cdot d_{avg}^2$. Again the parameters corresponding to the optimal schooling arrangement are identified through the CMA-ES algorithm (§ 2.5) by minimizing f, with $f = 0$ corresponding to cruise swimming of isolated swimmers.

The course of the stochastic optimization is shown in figure 9(*a*). The schooling arrangement evolves from the initial circular shape in figure 9(*b*) to the optimal 'hourglass' solution of figure 9(*c*), reminiscent of shapes observed in nature for medium-sized schools (Misund, Aglen & Fronaes 1995; Parrish, Viscido & Grünbaum 2002). The 'hourglass' shape is associated with an ∼80 % area contraction and with a 10-fold reduction in collective effort. The colour coding in figure 9(*b*,*c*) signifies the average swimming effort required by the dipoles to maintain their position in the school and illustrates how the circulation strength decreases for all swimmers, due to the more favourable arrangement.

The streamlines of the collective flow fields are illustrated in figure 9(*d*–*f*). The swimmers in the 'hourglass' school are found to align in striated patterns (figure 9*f*), consistent with the findings of figure 8, and to synchronize to induce a stronger global dipolar field (figure 9*e*), which is associated with a higher streamline density in the direction of travel and implies a stronger forward velocity. This formation allows swimmers to maintain their forward speed with a reduced individual effort (smaller circulation strength). Furthermore, dipoles in the centre of the necking region (figure 9*c*,*e*) get the benefit of drafting from the swimmers in the front while at the same time being pushed by the swimmers in the rear. In summary, packing to a smaller area while elongating the school shape favours striated swimming patterns (figure 9*c*,*f*), so that swimmers benefit from flow-mediated interactions. We note that the optimal swimming pattern exhibits a swimming effort which outperforms that of the Aoki–Couzin models by 190 % ($f_{a\,priori\,models} \simeq 0.06$).

Moreover, specified square-like and diamond-like formations are also shown to be detrimental in terms of collective and individual effort (figure 10); nevertheless, their analysis is revealing of the hydrodynamic mechanisms at play in the 'hourglass' optimal solution. Indeed, although the fitness of the entire school is $f = 0.004$ for both formations in figure 10(*a*,*b*), swimmers exposed on the left and right edges of the diamond formation suffer from high circulation expenditures, whereas those in the square formation generally do not. In the square formation, swimmers line up near the left and right edges of the school to help create greater net flow in the direction of travel. Conversely, swimmers at the front or rear of the diamond school benefit from schooling, whereas those in the square formation do not because of the counterflow produced from their immediate left and right neighbours. The optimal school shape solution of figure 10(*c*) takes advantage of these two effects: the 'hourglass' shape allows agents to line up near the boundary of the school and elongates in the travelling direction to reduce the counterflow from agents on the right and left edges.

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

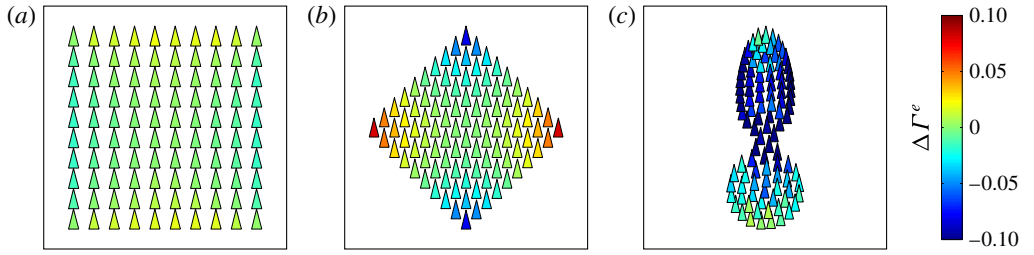


FIGURE 10. (Colour online) Comparison between (*a*) the square formation ($f = 0.004$), (*b*) the diamond formation ($f = 0.004$) and (*c*) the optimal solution ($f_{best} = -0.067$). The additional circulation change is computed for each swimmer and coloured accordingly. Simulations are run in the domain $[0, 1] \times [0, 1]$ with $\ell = 5 \times 10^{-4}$, $v^0 = 5\ell$, $\rho^T = 10\ell$, $\delta t = 0.1$, $\varphi = 0.01$, $\gamma = 0.98$ and $\epsilon = 0.01$. The notation is as defined in § 2 and appendix A.

4. Discussion

We present a novel approach to studying schooling that involves coupling reinforcement learning and stochastic optimization with swimming agents (Tchieu *et al.* 2012) that explicitly account for hydrodynamic interactions. Swimmers are modelled as self-propelled finite-width dipoles and have the capability to adjust their speed and orientation to cope with varying hydrodynamic loads. The dipoles represent the far-field vortex wake of self-propelled natural swimmers, so that agents impart long-range velocity fields on all other agents via their dipolar strengths. These nonlinear hydrodynamic interactions critically affect schooling dynamics and individual swimming patterns. We show that classical agent models, which rely on *a priori* specified social rules, are not robust in the presence of hydrodynamics. In order to compensate for the hydrodynamics and to allow for schooling in the present agent-based model, we reverse-engineer the rules that are followed by the swimmers. This reverse engineering is achieved through a reinforcement learning algorithm that creates mappings between the dynamic environment of the agents and their actions so as to maximize a numerical reward. Our approach differs from the widely used handcrafted *a priori* behavioural rules, and allows us to examine how hydrodynamics affects swimmers' decision-making in schooling. We find that adaptive swimming policies are crucial for maintaining schooling formations.

We evaluate the effectiveness of various formation patterns through an evolution strategy algorithm that identifies optimal schooling shapes and swimmer arrangements. We find that schools exhibiting minimal collective effort are 'hourglass'-shaped and elongated in the swimming direction. Elongated shapes allow for drafting and pushing of swimmers arranged in internal striated patterns. Such internal striated patterns are found to be optimal independently of the overall shape of the school, which is qualitatively consistent with experimental observations (Misund *et al.* 1995; Parrish *et al.* 2002; Boschitsch *et al.* 2014; Dewey *et al.* 2014).

Moreover, it is found that a tight packing of swimmers inside the school allows them to exploit flow-mediated interactions in terms of collective effort by enhancing the global dipolar field. At the same time, there is a limit to the amount of packing, as it becomes increasingly difficult for the swimmers to stay in formation due to stronger flow-mediated effects and increased probability of collisions. Such flow-mediated interactions can help to explain how certain fish travel in dense, elongated packs when migrating and foraging.

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

We remark that the present reverse-engineering approach to automatically identify interaction rules relative to a goal can be readily generalized to other forms of collective behaviour, from car traffic to social aggregations. In the context of schooling, future work is concerned with extending the use of these learning and optimization techniques to two- and three-dimensional viscous flows of multiple swimmers at intermediate Reynolds numbers (Gazzola, Hejazialhosseini & Koumoutsakos 2014). Reverse-engineering techniques, such as the ones proposed herein, can then be used to identify the various evolutionary traits that may have led to fish schooling.

Appendix A

A.1. *Learning optimal behaviour in a fluid-mediated environment*

A.1.1. *Time integration and handling of collisions*

The set of ODEs given in (2.2*a*,*b*) is numerically integrated using the forward Euler method so as to retain flexibility in the decision selection process of the reinforcement learning algorithm. Since the equations become stiff as swimmers approach one another, the time step $\mathrm{d}t$ is computed according to the minimum distance d_{min} between all dipole swimmers present in the environment. We bound $\mathrm{d}t$ with $\mathrm{d}t = \mathrm{d}t_{max}$ if $d_{min} \geqslant 2\ell$ and $\mathrm{d}t = \mathrm{d}t_{min} = 5e^{-4}$ if $d_{min} \leqslant \ell/2$. We require $\mathrm{d}t_{max}$ to be dependent on the nominal velocity v^0 such that $\mathrm{d}t_{max} = (5\ell C)/v^0$. For the simulations here, $C = 0.005$.

Dipoles tend to collide if they are in close proximity with one another (Tchieu *et al.* 2012). In a collision of two dipoles, both dipoles stop and annihilate their respective circulation strengths. We deal with this phenomenon by labelling colliding dipoles as 'dead' if $\mathrm{d}t = \mathrm{d}t_{min}$. The dipoles' circulations are artificially set to zero; hence dead dipoles no longer influence the other swimmers in the flow.

A.1.2. *Error analysis of non-adaptive and adaptive schools of dipole swimmers*

As observed in figure 6 of the main text and figure 11 here, in the presence of hydrodynamic interaction a given schooling arrangement is not maintained robustly without an adaptive dynamic response to the flow. We quantitatively demonstrate that schooling patterns can be maintained by swimmers through reinforcement learning. This is illustrated in figure 11, in which the errors (average distances to the target points, $e = \sum_n^N d_n/\ell$) of the simulations in figure 6 are compared.

A.2. *Optimal schooling formations*

We seek optimal schooling configurations by combining reinforcement learning with an evolutionary strategy. The algorithm of choice is the covariance matrix adaptation evolutionary strategy (CMA-ES) in its multihost, rank-μ and weighted recombination form (Hansen *et al.* 2003; Gazzola, Vasilyev & Koumoutsakos 2011*b*). The robustness of the CMA-ES is mainly controlled by the population size p (Hansen *et al.* 2003). In this work, as a tradeoff between robustness and fast convergence, we set $p = 100$ for all optimization computations. Bounds of the search space are enforced during the sampling through a rejection algorithm.

In this strategy, the CMA-ES determines the optimal configuration based on the metric of swimming effectiveness f, while reinforcement learning determines the optimal policy for an agent to follow its target point under any configuration

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

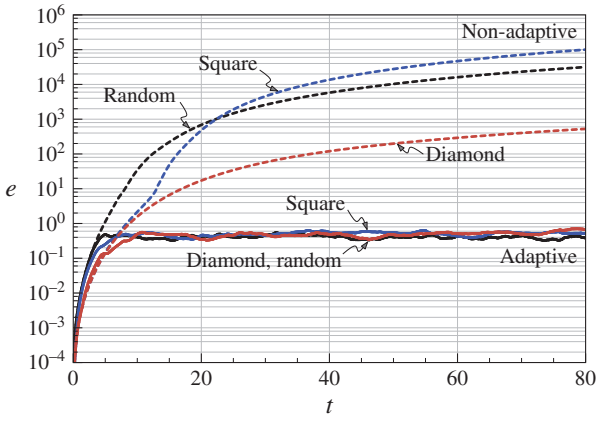


FIGURE 11. (Colour online) Time evolution of the sum of all errors between all agents and their respective assigned target points. The non-adaptive school simulations (dotted lines) of diamond (red), square (blue) and random (black) formations correspond to figure 6(*b*,*e*,*h*), respectively. The adaptive school simulations (bottom set of solid lines) of diamond (red), square (blue), and random (black) formations correspond to figure 6(*c*,*f*,*i*), respectively. The error is defined as $e = \sum_n^N d_n/\ell$, i.e. the average distance to the target points.

requested by the CMA-ES. Therefore, every cost function evaluation entails the CMA-ES dispatching a parameter set defining the geometry of the school, and then a reinforcement learning training period ($\Delta T_{training} = 10\,000$) that allows the dipoles to learn how to swim in the given arrangement, followed by an evaluation interval ($\Delta T_{eval} = 100$) in which the school effectiveness is measured and returned to the CMA-ES.

A.2.1. *CMA-ES settings and parameterization for the optimization of school shape*

We evaluate how the overall school shape affects collective effort and optimize its effectiveness in terms of (2.7). To do so, we create a school by designing a general external shape and arranging the swimmers inside the boundary by placing them maximally distant from one another and the boundary. The shape of the school is dictated by a cubic spline-based parameterization introduced in Rossinelli, Chatagny & Koumoutsakos (2011). According to this parameterization, the external school shape is represented as $c = S(\phi)$, where c is the radial distance from the origin in the polar plane, $\phi \in [0, \pi]$ is the angle and S is the piecewise polynomial of the cubic spline. The spline control points (red dots in figure 12), expressed in polar coordinates, are $(c_0, 0)$, (c_1, ϕ_1) and $(c_2, 0)$, with the radii defined as $c_1 = k_1 \cdot c_0$ and $c_2 = k_2 \cdot b_0$ where k_1 and k_2 are constants. The school shape is completed by reflecting the obtained spline profile. Unlike Rossinelli *et al.* (2011), the area of the school shape $A(S)$, and therefore c_0, is controlled via an extra parameter, namely the average distance d_{avg} between swimmers, such that $b_0 : A(S) = N \cdot d_{avg}^2$, where N is the number of dipoles. Therefore, the shape of the school depends on the four parameters $\boldsymbol{p} = \{d_{avg}, k_1, k_2, \phi_1\}$. These free parameters are varied within the search space $[1, 5] \times [0.1, 2] \times [0.1, 2] \times [\pi/6, 5\pi/6] \in \mathbb{R}^4$ during the optimization.

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

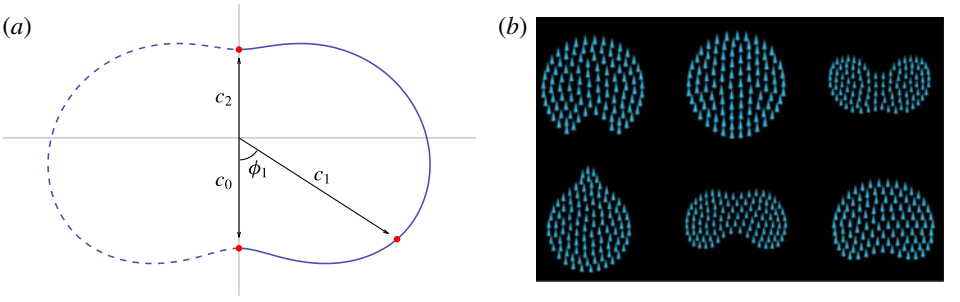


FIGURE 12. (Colour online) Parameterization of the shape of the school and a representative subset of candidate solutions. The parameterization generates the solid line (blue) and is reflected to create the symmetric complement (dashed line). The control points are depicted as red dots.

A.2.2. *CMA-ES settings for the optimization of the internal lattice structure of a school*

The free parameters h, b and β for the diamond, rectangular and hexagonal search space (see figure 8*a–c*) are varied in the ranges $[0, 50\ell] \times [0, 50\ell] \times [0, \pi/2] \in \mathbb{R}^3$, $[0, 50\ell] \times [0, 50\ell] \times [0, \pi/2] \in \mathbb{R}^3$ and $[0, 50\ell] \times [0, \pi/6] \in \mathbb{R}^2$, respectively, during the optimization.

REFERENCES

ABRAHAMS, M. V. & COLGAN, P. W. 1987 Fish schools and their hydrodynamic function: a reanalysis. *Environ. Biol. Fishes* **20** (1), 79–80.

ADITI SIMHA, R. & RAMASWAMY, S. 2002 Hydrodynamic fluctuations and instabilities in ordered suspensions of self-propelled particles. *Phys. Rev. Lett.* **89** (5), 058101.

AOKI, I. 1982 A simulation study on the schooling mechanism in fish. *Bull. Japan. Soc. Sci. Fisheries* **48** (8), 1081–1088.

AREF, H., NEWTON, P., STREMLER, M. A., TOKIEDA, T. & VAINCHTEIN, D. 2003 Vortex crystals. *Adv. Appl. Mech.* **39**, 1–79.

BARNES, R. S. K. & HUGHES, R. N. 1988 *An Introduction to Marine Ecology*. Blackwell.

BOSCHITSCH, B. M., DEWEY, P. A. & SMITS, A. J. 2014 Propulsive performance of unsteady tandem hydrofoils in an in-line configuration. *Phys. Fluids* **26** (5), 051901.

BRADY, J. F. & BOSSIS, G. 1988 Stokesian dynamics. *Annu. Rev. Fluid Mech.* **20**, 111–157.

BRADY, J. F., PHILLIPS, R. J., LESTER, J. C. & BOSSIS, G. 1988 Dynamic simulation of hydrodynamically interacting suspensions. *J. Fluid Mech.* **195** (1), 257–280.

BREDER, C. M. 1965 Vortices and fish schools. *Zool. New York* **50** (2), 97–114.

CHATE, H., GINELLI, F., GREGOIRE, G., PERUANI, F. & RAYNAUD, F. 2008 Modeling collective motion: variations on the Vicsek model. *Eur. Phys. J.* B **64** (3–4), 451–456.

COUZIN, I. D., KRAUSE, J., FRANKS, N. R. & LEVIN, S. A. 2005 Effective leadership and decision-making in animal groups on the move. *Nature* **433** (7025), 513–516.

COUZIN, I. D., KRAUSE, J., JAMES, R., RUXTON, G. D. & FRANKS, N. R. 2002 Collective memory and spatial sorting in animal groups. *J. Theoret. Biol.* **218** (1), 1–11.

CULLEN, J. M., SHAW, E. & BALDWIN, H. A. 1965 Methods for measuring the three-dimensional structure of fish schools. *Animal Behaviour* **13** (4), 534–543.

DEWEY, P. A., QUINN, D. B., BOSCHITSCH, B. M. & SMITS, A. J. 2014 Propulsive performance of unsteady tandem hydrofoils in a side-by-side configuration. *Phys. Fluids* **26** (4), 041903.

GAZZOLA, M. 2013 Simulation, optimization and learning of artificial swimmers. PhD thesis, ETH Zurich.

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

GAZZOLA, M., BURCKHARDT, C. J., BAYATI, B., ENGELKE, M., GREBER, U. F. & KOUMOUTSAKOS, P. 2009 A stochastic model for microtubule motors describes the in vivo cytoplasmic transport of human adenovirus. *PLoS Comput. Biol.* **5** (12), e1000623.

GAZZOLA, M., CHATELAIN, P., VAN REES, W. M. & KOUMOUTSAKOS, P. 2011*a* Simulations of single and multiple swimmers with non-divergence free deforming geometries. *J. Comput. Phys.* **230** (19), 7093–7114.

GAZZOLA, M., HEJAZIALHOSSEINI, B. & KOUMOUTSAKOS, P. 2014 Reinforcement learning and wavelet adapted vortex methods for simulations of self-propelled swimmers. *SIAM J. Sci. Comput.* **36** (3), B622–B639.

GAZZOLA, M., MIMEAU, C., TCHIEU, A. A. & KOUMOUTSAKOS, P. 2012*a* Flow mediated interactions between two cylinders at finite Re numbers. *Phys. Fluids* **24** (4), 043103.

GAZZOLA, M., VAN REES, W. M. & KOUMOUTSAKOS, P. 2012*b* C-start: optimal start of larval fish. *J. Fluid Mech.* **698**, 5–18.

GAZZOLA, M., VASILYEV, O. V. & KOUMOUTSAKOS, P. 2011*b* Shape optimization for drag reduction in linked bodies using evolution strategies. *Comput. Struct.* **89** (11–12), 1224–1231.

GOTO, Y. & TANAKA, H. 2015 Purely hydrodynamic ordering of rotating disks at a finite Reynolds number. *Nat. Commun.* **6**.

HANSEN, N., MULLER, S. D. & KOUMOUTSAKOS, P. 2003 Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evol. Comput.* **11** (1), 1–18.

HATWALNE, Y., RAMASWAMY, S., RAO, M. & SIMHA, R. A. 2004 Rheology of active-particle suspensions. *Phys. Rev. Lett.* **92** (11), 118101.

HUBBARD, S., BABAK, P., SIGURDSSON, S. T. & MAGNUSSON, K. G. 2004 A model of the formation of fish schools and migrations of fish. *Ecol. Model.* **174** (4), 359–374.

HUTH, A. & WISSEL, C. 1992 The simulation of the movement of fish schools. *J. Theoret. Biol.* **156** (3), 365–385.

ISHIKAWA, T., SIMMONDS, M. P. & PEDLEY, T. J. 2006 Hydrodynamic interaction of two swimming model micro-organisms. *J. Fluid Mech.* **568**, 119–160.

JIANG, H. S., OSBORN, T. R. & MENEVEAU, C. 2002 The flow field around a freely swimming copepod in steady motion. Part I. Theoretical analysis. *J. Plankton Res.* **24** (3), 167–189.

KANSO, E. & NEWTON, P. N. 2009 Passive locomotion via normal-mode coupling in a submerged spring–mass system. *J. Fluid Mech.* **641**, 205–215.

KERN, S. & KOUMOUTSAKOS, P. 2006 Simulations of optimized anguilliform swimming. *J. Expl Biol.* **209** (24), 4841–4857.

KOCH, D. L. & SUBRAMANIAN, G. 2011 Collective hydrodynamics of swimming microorganisms: living fluids. *Annu. Rev. Fluid Mech.* **43**, 637–659.

LANDEAU, L. & TERBORGH, J. 1986 Oddity and the 'confusion effect' in predation. *Animal Behaviour* **34** (5), 1372–1380.

LIAO, J. C., BEAL, D. N., LAUDER, G. V. & TRIANTAFYLLOU, M. S. 2003 Fish exploiting vortices decrease muscle activity. *Science* **302** (5650), 1566–1569.

LIN, Z., THIFFEAULT, J.-L. & CHILDRESS, S. 2011 Stirring by squirmers. *J. Fluid Mech.* **669**, 167–177.

MAJOR, P. F. 1978 Predator–prey interactions in two schooling fishes, *Caranx ignobilis* and *Stolephorus purpureus*. *Animal Behaviour* **26** (0), 760–777.

MISUND, O. A., AGLEN, A. & FRONAES, E. 1995 Mapping the shape, size, and density of fish schools by echo integration and a high-resolution sonar. *ICES J. Marine Sci.* **52** (1), 11–20.

NAIR, S. & KANSO, E. 2007 Hydrodynamically coupled rigid bodies. *J. Fluid Mech.* **592**, 393–411.

NIWA, H. S. 1994 Self-organizing dynamic-model of fish schooling. *J. Theoret. Biol.* **171** (2), 123–136.

PAOLETTI, P. & MAHADEVAN, L. 2011 Planar controlled gliding, tumbling and descent. *J. Fluid Mech.* **689**, 489–516.

PARK, H., NOCA, F. & KOUMOUTSAKOS, P. 2005 Vortex-generating microscopic robots would move in swarms. *Tech. Rep.*, NASA Jet Propulsion Laboratory.

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press

PARRISH, J. K., VISCIDO, S. V. & GRÜNBAUM, D. 2002 Self-organized fish schools: an examination of emergent properties. *Biol. Bull.* **202** (3), 296–305.

PARTRIDGE, B. L. 1982 The structure and function of fish schools. *Sci. Am.* **246** (6), 114–123.

PARTRIDGE, B. L., PITCHER, T., CULLEN, J. M. & WILSON, J. 1980 The three-dimensional structure of fish schools. *Behav. Ecol. Sociobiol.* **6** (4), 277–288.

PITCHER, T. J., MAGURRAN, A. E. & WINFIELD, I. J. 1982 Fish in larger shoals find food faster. *Behav. Ecol. Sociobiol.* **10** (2), 149–151.

VAN REES, W. M., GAZZOLA, M. & KOUMOUTSAKOS, P. 2013 Optimal shapes for anguilliform swimmers at intermediate reynolds numbers. *J. Fluid Mech.* **722**, R3.

VAN REES, W. M., GAZZOLA, M. & KOUMOUTSAKOS, P. 2015 Optimal morphokinematics for undulatory swimmers at intermediate Reynolds numbers. *J. Fluid Mech.* **775**, 178–188.

REYNOLDS, C. W. 1987 Flocks, herds and schools: a distributed behavioural model. In *SIGGRAPH'87: Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 71–87.

RISTROPH, L. & ZHANG, J. 2008 Anomalous hydrodynamic drafting of interacting flapping flags. *Phys. Rev. Lett.* **101** (19), 194502.

ROSSINELLI, D., CHATAGNY, L. & KOUMOUTSAKOS, P. 2011 Evolutionary optimization of scalar transport in cylinder arrays on multiGPU/multicore architectures. In *Evolutionary and Deterministic Methods for Design, Optimization and Control*, pp. 773–784. Proc. Euro-Gen.

SAFFMAN, P. G. 1967 The self-propulsion of a deformable body in a perfect fluid. *J. Fluid Mech.* **28**, 385–389.

SHAW, E. 1978 Schooling fishes. *Am. Sci.* **66** (2), 166–175.

SUTTON, S. R. & BARTO, A. G. 1998 *Reinforcement Learning: An Introduction*. MIT Press.

TAYLOR, G. I. 1952 Analysis of the swimming of long and narrow animals. *Proc. R. Soc. Lond.* A **214**, 158–183.

TCHIEU, A. A., CROWDY, D. & LEONARD, A. 2010 Fluid-structure interaction of two bodies in an inviscid fluid. *Phys. Fluids* **22** (10), 107101.

TCHIEU, A. A., KANSO, E. & NEWTON, P. K. 2012 The finite-dipole dynamical system. *Proc. R. Soc. Lond.* A **468**, 3006–3026.

WU, T. Y. 1971 Hydrodynamics of swimming fish and cetaceans. *Adv. Appl. Maths* **11**, 1–63.

TSANG, A. C. H. & KANSO, E. 2013 Dipole interactions in doubly periodic domains. *J. Nonlinear Sci.* **23** (6), 971–991.

VICSEK, T., CZIROK, A., BENJACOB, E., COHEN, I. & SHOCHET, O. 1995 Novel type of phase-transition in a system of self-driven particles. *Phys. Rev. Lett.* **75** (6), 1226–1229.

VISCIDO, S. V., PARRISH, J. K. & GRUNBAUM, D. 2005 The effect of population size and number of influential neighbors on the emergent properties of fish schools. *Ecol. Model.* **183** (2–3), 347–363.

WATKINS, C. & DAYAN, P. 1992 Q-learning. *Mach. Learn.* **8** (3–4), 279–292.

WEIHS, D. 1973 Hydromechanics of fish schooling. *Nature* **241** (5387), 290–291.

WEIHS, D. 2004 The hydrodynamics of dolphin drafting. *J. Biol.* **3** (2), 8.

WHITTLESEY, R., LISKA, S. & DABIRI, J. 2010 Fish schooling as a basis for vertical axis wind turbine farm design. *Bioinspir. Biomim.* **5** (3), 035005.

WOLFGANG, M. J., ANDERSON, J. M., GROSENBAUGH, M. A., YUE, D. K. & TRIANTAFYLLOU, M. S. 1999 Near-body flow dynamics in swimming fish. *J. Exp. Biol.* **202** (17), 2303–2327.

ZHU, L., LAUGA, E. & BRANDT, L. 2012 Self-propulsion in viscoelastic fluids: pushers vs. pullers. *Phys. Fluids* **24**, 051902.

https://doi.org/10.1017/jfm.2015.686 Published online by Cambridge University Press