Soft Matter

PAPER

View Article Online
View Journal | View Issue

Cite this: *Soft Matter*, 2025, **21**, 4488

Tailoring interactions between active nematic defects with reinforcement learning†

Carlos Floyd, *[ab] Aaron R. Dinner [ab] and Suriyanarayanan Vaikuntanathan*[ab]

Active nematics are paradigmatic active matter systems which generate micron-scale patterns and flows. Recent advances in optical control over molecular motors now allow experimenters to control the non-equilibrium activity field in space and time and, in turn, the patterns and flows. However, engineering effective activity protocols remains challenging due to the complex dynamics. Here, we explore a model-free approach for controlling active nematic fields using reinforcement learning. Combining machine learning with trial-and-error exploration of the system dynamics, reinforcement learning bypasses the need for accurate parameterization and model representation of the active nematic. We apply this technique to demonstrate how local activity fields can induce effective interactions between nematic defects, enabling them to follow designer dynamical laws. Moreover, the sufficiency of our low-dimensional system observables and actions suggests that coarse projections of the active nematic field can be used for precise feedback control, making experimental or biological implementation of such feedback loops plausible.

Received 17th January 2025,
Accepted 24th April 2025

DOI: 10.1039/d5sm00063g

rsc.li/soft-matter-journal

I. Introduction

Active nematics formed from a liquid crystalline suspension of active force dipoles such as cytoskeletal filaments and molecular motors have emerged as a useful experimental platform for creating micron-scale flows.[1,2] Recent research indicates that the positioning of defects in active nematics plays a crucial role in organizing the stress and velocity fields of the system, driving coherent flows.[3,4] In addition to presenting synthetic opportunities, control of active nematics appears useful for understanding how mechanical forces are coordinated at the tissue level in certain systems.[5,6] For example, it was recently shown that defects in the nematic ordering of cytoskeletal filaments in epithelial cells of developing hydra are precisely positioned at key global organizing centers, such as the future mouth.[7–15] It is thus of interest to explore how active nematic defects can be guided, both through through experimental manipulations *in vitro* and through mechanochemical feedback loops in living systems.[16,17]

Recent works have focused on guiding active nematic flows through the external control of spatiotemporally dynamic activity fields $\alpha(\mathbf{r},t)$. Experimental advances in optical actuation of molecular motors using light fields motivate studying activity fields as externally controlled functions.[18–26] Light-based control has also recently been developed over Ca^{2+}-powered assembly and contraction of proteins found in protists[27,28] and over cytoskeletal network growth.[29] In simulations, optimal time-dependent activity fields $\alpha(\mathbf{r},t)$ can be derived using knowledge of the nematohydrodynamic equations of motion to guide nematic and polar defects along desired paths.[30–33] Other techniques allow targeted modulation of nematic channel flow and design of localized "topological tweezers" for precise defect manipulation.[34–36]

A key difficulty in implementing the above mentioned control methods is their reliance on accurate system models and parameterization. Here, we explore controlling active nematics using a model-free machine learning technique called reinforcement learning (RL).[37,38] This approach allows a program to develop a closed-loop control policy for a dynamical system purely through trial and error, without relying on precise model specifications.

To our knowledge RL has not yet been applied to control active nematics, although previous studies used it to control other active matter systems such as flockers, driven colloids, and branched actin networks.[39–41] These studies focus on design tasks that aim to achieve a specific static property of the system, such as target net flocking motion or average cluster size. In contrast, we study control tasks that guide active nematic defects to follow prescribed virtual interactions with one another (see Fig. 1).[42,43] Our goal is to learn activity field protocols that can effectively override the natural defect dynamics (such as passive Coulomb-like attraction) to impose

[a] *Department of Chemistry, The University of Chicago, Chicago, Illinois 60637, USA. E-mail: csfloyd@uchicago.edu, svaikunt@uchicago.edu*
[b] *The James Franck Institute, The University of Chicago, Chicago, Illinois 60637, USA*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d5sm00063g

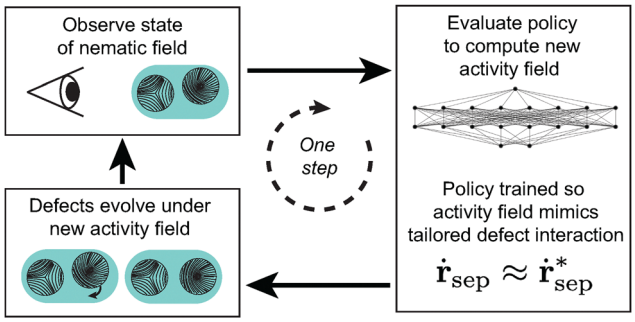

Fig. 1 Schematic overview of the closed-loop control over active nematic defect dynamics enabled by a trained RL policy. One step of the feedback loop is depicted.

a user-specified interaction law, a more general design objective termed "cyberphysics" in ref. 38. Additionally, we show that efficient protocols can be learned even when feedback is limited to imperfect information such as a coarse projection of the full nematic field configuration.

II. Methods

Here we describe how we use RL to impose customized interaction laws between active nematic defects. In the Appendix we outline the overdamped active nematohydrodynamic equations of motion that we numerically integrate. In Section II A we describe the geometry and interaction laws governing active nematic defects. We then define the states, actions, rewards, and experiment structure which make up our RL setting in Section II B. In Section II C we describe the specific RL algorithm that we use to optimize the policy.

A. Defect geometry and interactions

We consider throughout a pair of $\pm 1/2$ defects in a periodic two-dimensional (2D) domain. The defect positions are labeled $\mathbf{p}_{\pm}$ and we define the separation vector $\mathbf{r}_{\mathrm{sep}} = \mathbf{p}_{+} - \mathbf{p}_{-}$ (see Fig. 2). As described for example in ref. 35, each defect has an orientation in addition to its position. The +1/2 defect orientation is described by a vector $\hat{\mathbf{e}} = (\cos(\phi_{+}), \sin(\phi_{+}))$ given by

$$\hat{\mathbf{e}} = \frac{\nabla \cdot \mathbf{Q}(\mathbf{r}, t)}{|\nabla \cdot \mathbf{Q}(\mathbf{r}, t)|}, \tag{1}$$

where $\mathbf{Q}$ is the symmetric and traceless nematic order parameter, and the expression is evaluated in the limit approaching the center of the defect core. The orientation of the $-1/2$ defect has a three-fold symmetry and cannot be represented by a vectorial quantity. Instead, it is represented by the rank-three tensor

$$\Theta_{ijk} = \frac{\langle \partial_i Q_{jk} + \partial_j Q_{ik} + \partial_k Q_{ij} \rangle}{3\left|\langle \partial_k Q_{ij} \rangle\right|}, \tag{2}$$

where the brackets denote an angular average around the defect core. This quantity can also be expressed in terms of triple outer products of a vector $\hat{\mathbf{t}} = (\sin(\phi_{-}), \cos(\phi_{-}))$ as

$$\Theta_{ijk} = \hat{t}_i \hat{t}_j \hat{t}_k - \frac{1}{4}(\delta_{ij}\hat{t}_k + \delta_{kj}\hat{t}_i + \delta_{ik}\hat{t}_j), \tag{3}$$

where δ_{ij} is the Kronecker delta. As required, Θ_{ijk} is invariant under $\phi_{-} + n2\pi/3$ for any integer n. In simulation we compute the nematic positions and orientations following methods described in ref. 44.

Although defects in an active nematic fluid evolve under complex hydrodynamics (see eqn (10) and (11) below), theoretical work has shown that their motion can be approximated using simpler one- and two-body dynamical equations.[35,45,46] The positions $\mathbf{p}_{\pm}$ and orientations $\phi_{\pm}$ approximately obey differential equations which are coupled to each other through elastic interactions, and they are also coupled to the local activity profile $\alpha(\mathbf{r})$ and its gradients. These equations are derived from the hydrodynamic Stokes flow of the active nematic fluid and involve numerical prefactors such as the effective defect size, friction coefficient, elasticity constant, shear viscosity, and others. As predicted by these equations, in the absence of activity (*i.e.*, for passive nematics) a pair of oppositely-charged defects will attract each other without rotating in a Coulomb-like interaction until they annihilate, preserving the total topological charge. This default behavior is illustrated in Fig. 3. These effective dynamical equations enable the design of precise "activity tweezers" that manipulate defect dynamics, but this technique is complicated by the difficulty of estimating the model parameters involved.[35,47] Here, we simply use the fact that defect positions and orientations couple in some way to activity gradients to find tweezer-like activity protocols purely through exploration (*i.e.*, RL). This approach is thus agnostic to the underlying model parameters and dynamics.

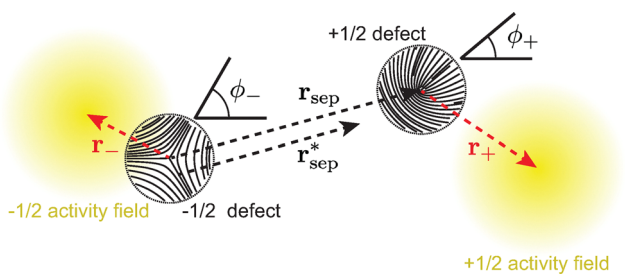

Fig. 2 Illustration of the geometric quantities used to define the positions and orientations of a pair of $\pm 1/2$ nematic defects in a general configuration. Two activity fields are also shown in the vicinity of the defects. The vectors $\mathbf{r}_{\pm}$ point from the defect locations to the center of their nearby activity fields, while $\mathbf{r}_{\mathrm{sep}}$ points from the $-1/2$ defect to the +1/2 defect. The target separation $\mathbf{r}^{*}_{\mathrm{sep}}$ is denoted with an asterisk. The angles $\phi_{\pm}$ describe the orientation of the defects with respect to the horizontal axis.

B. States, actions, rewards, and experiment structure

An RL algorithm produces a so-called policy $\pi_{\theta}(S)$ which is a function that maps from a state S into an action A. Training occurs through many trial-and-error iterations in which the policy is incrementally improved *via* updates to its parameters $\boldsymbol{\theta}$, in the sense that the actions which it learns to choose

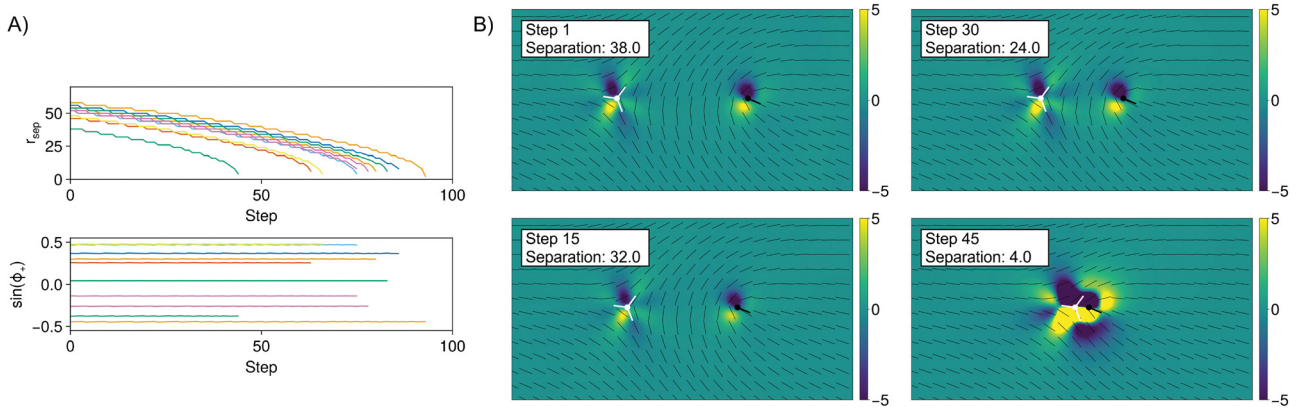

Fig. 3 Without control, defects attract and annihilate without rotating. (A) Top: Trajectories of r_{sep} shown as different colors with no activity field, illustrating the passive behavior of the defect pairs. Defects pairs are initiated with random horizontal separations and the nematic field is rotated by a random amount. Bottom: Corresponding trajectories of $\sin(\phi_+)$. (B) Snapshots of the nematic field at four different steps during one of the trajectories. Each black line represents the local nematic orientation; one line is drawn for every 16 lattice sites. Color represents the local vorticity of the fluid velocity field. The −1/2 defect and its orientation are depicted as a white dot and lines, and the +1/2 defect and its orientation are depicted as a black dot and a line. The 100 × 100 simulation domain is cropped to 100 × 50 to improve visibility.

produce future states that optimize a user-defined reward function. We demonstrate RL-based control over active nematic defect dynamics through several tasks in this paper, each corresponding to a different choice of states, actions, and reward function. We describe each choice as we discuss the tasks below.

For every task we run several experiments using different task parameters or different seeds for the pseudo-random number generator, and the outcome of each experiment is a trained policy function $\pi_{\boldsymbol{\theta}}(S)$. An experiment is divided into episodes, each of which begins by resetting the active nematic field using some distribution of initial conditions (Fig. 4). Every episode consists of a fixed number of steps, N_{steps}, although an episode may terminate before this number of steps if the number of defects changes through annihilation or creation. At every step, the RL agent views the state S and chooses a new action A, corresponding to a given configuration of the activity field $\alpha(\mathbf{r},t)$, which is applied for the duration of the step. A step consists of $N_{\mathrm{d}t}$ iterations of the numerical integrator. The fixed time interval which elapses during one step is thus $N_{\mathrm{d}t}\mathrm{d}t$ where $\mathrm{d}t$ is the time resolution of the integrator; in simulation units we have $\mathrm{d}t = 1$. The experiment ends when a number of episodes, N_{eps}, is reached or when the wall time of the program exceeds a specified value. Throughout this paper we set $N_{\mathrm{d}t} = 50$, $N_{\text{steps}} = 75$, and we run experiments for 3 hours, corresponding to roughly $N_{\text{eps}} = 150$ episodes per experiment. The choice to fix wall time rather than N_{eps} is arbitrary and made simply for convenience.

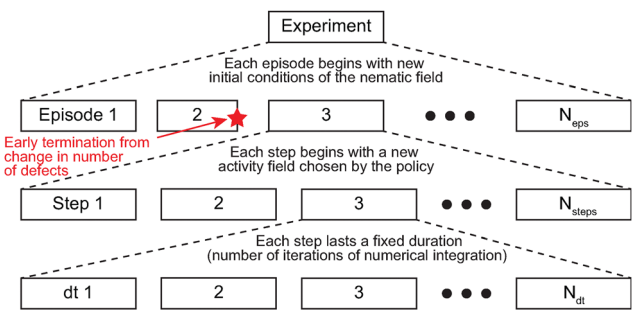

Fig. 4 Schematic illustration of the structure of an experiment, episode, and step as used in our RL training program.

C. Reinforcement learning algorithm

To train the RL policy $\pi_{\boldsymbol{\theta}}(S)$ we use a variant of the actor-critic algorithm[37] called deep deterministic policy gradient (DDPG),[48,49] which is suited for continuous actions in deterministic environments. Four neural networks are used in this approach: two copies of an actor network with parameters $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$, which implement the policy function, and two copies of a critic network with parameters $\mathbf{w}$ and $\mathbf{w}'$, which estimate the value function $Q_{\mathbf{w}}(S,A)$, *i.e.*, the expected cumulative future reward of choosing action A in state S. The main networks, $\mu_{\boldsymbol{\theta}}(A)$ and $Q_{\mathbf{w}}(S,A)$, are updated during training and used to select actions and train the actor, while the target networks track the parameters of the main networks with slow updates to provide stable function estimates for training the critic. Specifically, the target network parameters are updated as $\boldsymbol{\theta}' \leftarrow \rho\boldsymbol{\theta}' + (1-\rho)\boldsymbol{\theta}$ and $\mathbf{w}' \leftarrow \rho\mathbf{w}' + (1-\rho)\mathbf{w}$ for $\rho \lesssim 1$. Using these slowly updated target networks in addition to the main networks improves stability and convergence of learning.[49]

The main actor and critic networks are updated using stochastic gradient descent (SGD) with mini-batches of $N_{\text{batch}} = 32$ samples from a replay buffer, which stores tuples $(S(s), A(s), R(s), T(s))$ (state, action, reward, and termination flag) collected at every step s. For each experiment, an initial random policy is used for $10N_{\text{batch}}$ steps to populate the replay buffer, after which the main actor network $\mu_{\boldsymbol{\theta}}(A)$ is used for action selection, with SGD updates performed every 5 steps. How the mini-batches are used to update the network parameters is described in ref. 48 and 49. All neural networks have two hidden layers with 32 neurons each, and they are trained using the ADAM

optimizer with a learning rate of 0.001 and a weight norm clip of 1.0. The discount factor is $\gamma = 0.99$ (see ref. 48 and 49 for a definition) and the weight transfer factor is $\rho = 0.995$. These parameter values were chosen to be consistent with previous implementations of the DDPG method.[48,49]

The actions are chosen as $A = \pi_\theta(S) + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0,0.05)$ is a small Gaussian noise added to the policy output to encourage exploration of the available actions. The resulting action A is then clipped to the range $[-1,1]$. State values provided as inputs to the network are normalized to lie approximately within the same range. For each task, the action values are linearly mapped from this range to the corresponding scaled parameters of the activity field $\alpha(\mathbf{r},t)$, such that -1 maps to the lowest scaled values and 1 maps to the largest.

To numerically integrate the active nematic hydrodynamics equations in eqn (10), we use a custom Julia implementation of Heun's finite difference method. The implementation was previously described and validated in ref. 50–52. We parameterize the nematic system so that the unit of length is equal to the equilibrium nematic persistence length (see Appendix).[53] To train the RL policy, we combine this numerical solver with an implementation of the DDPG algorithm provided by the ReinforcementLearning.jl package.[54]

III. Results

To demonstrate the feasibility of using RL to control active nematic defect interactions through spatiotemporal activity fields, we show here three test cases of varying difficulty.

A. Translating a +1/2 defect

In the first task, we do not impose a virtual dynamics but instead a static property that the defects should exhibit. Specifically, the goal is to move the defects so that they are horizontally separated by an amount l_0 (Fig. 5A). To do this we apply a disk of nearly uniform activity centered on the +1/2 defect, and we leverage the fact that such a defect self-propels with a velocity vector parallel to its orientation vector and approximately proportional to the local activity: $\dot{\mathbf{p}}_+ \sim -\alpha\hat{\mathbf{e}}$.[35,45,55] We initialize each episode by nucleating a pair of defects that are vertically aligned in the 100×100 periodic domain and horizontally separated by a random offset l_{init} selected uniformly from the range [37.5,62.5].

The state used in the RL algorithm at step s is $S(s) = (r_{\text{sep}}(s) - l_0)/50$ (where 50 is a rough scale factor), and the action is the amplitude α_0 of the activity profile scaled to the range $[-5,5]$. We note that allowing the activity to change sign is unphysical as a typical active nematic fluid has either only extensile ($\alpha > 0$) or contractile ($\alpha < 0$) force dipoles. We allow activity to change sign in this task as a first illustration of a simple control mechanism, and in the remaining tasks we constrain the sign of the activity and only vary its magnitude or position. The shape of the activity profile is an azimuthally symmetric function centered on $\mathbf{p}_+$ with radial dependence

$$\alpha(r;\alpha_0,c,m) = \frac{\alpha_0}{2}\left(1 - \tanh\left(\frac{r-c}{m}\right)\right), \tag{4}$$

where, for this task, $c = 5$ is a cutoff parameter and $m = 1$ is the width of the logistic profile. The reward for this task is computed at step s as $R(s) = -|r_{\text{sep}}(s+1) - l_0|$.

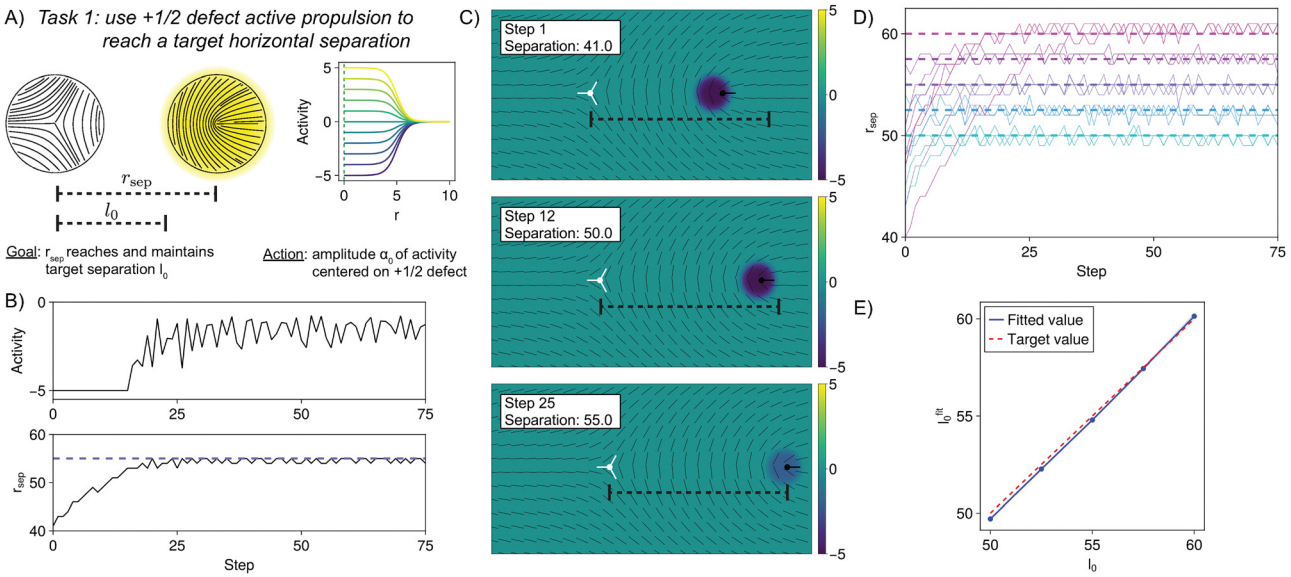

Fig. 5 Results for task 1: to reach a target horizontal separation l_0. (A) Schematic illustration of the task, in which a nearly uniform disk of activity is applied to the center of the +1/2 defect. The amplitude of the activity is controlled as an action, and its radial profile for different amplitudes is shown on the right. The dashed green line denotes the position of the +1/2 defect in the activity field at the start of each step. (B) Trajectories of the activity amplitude (top) and resulting horizontal separation r_{sep} (bottom) with a trained RL policy. The dashed line represents $l_0 = 55$ for this task. (C) Snapshots of the active nematic field at the end of three different steps for the episode in panel (B). See Fig. 3B caption for a description of the visualization. Color here denotes the amplitude of the activity. The black dashed line represents the target separation $l_0 = 55$. (D) Two trajectories of r_{sep} with a trained RL policy for five values of l_0, shown as different colors. The values of l_0 are shown as dashed lines. (E) Fitted values of l_0^{fit}, plotted against their target values of l_0. Standard deviations are shown as shaded areas (smaller than symbols).

A typical episode for $l_0 = 55$ is shown in Fig. 5B and C. The policy applies a strong negative activity field to the +1/2 defect, which propels it away from the −1/2 defect. Upon reaching the desired separation of 55 lattice units the applied activity weakens in magnitude and, in a feedback control-like manner, nudges the +1/2 defect further backward whenever the −1/2 defect moves toward it due a to the attractive elastic interaction between them. This maintains the separation at $r_{sep} = 55$ for the remainder of the episode. Although in principle the activity could be applied when the separation is at the spring equilibrium, $r_{sep} = l_0$, to exactly balance the attractive force, the actual separation oscillates slightly around this point due to the discretization of the simulation domain. We also do not expect the algorithm to learn a numerically exact policy.

We train RL policies for a range of values of l_0 and observe that in each case the algorithm converges to a policy in which the target separation is reached by the end of each episode (Fig. 5D). We fit the learned l_0^{fit} by taking the average of r_{sep} over the last 25 steps of each episode, over the last 40 episodes of each experiments, and over 10 experiments using different random initial seeds. The learned l_0^{fit} matches the target l_0 (Fig. 5E) in each case. Thus, RL can learn how to adjust the activity amplitude as a function of defect position to reach a target separation, for a simple physical set-up in which the defects are aligned and move primarily due to the activity-induced propulsion of the +1/2 defect.

B. Translating a −1/2 defect

We next consider a more complicated task, in which the horizontal defect separation r_{sep} should not just maintain a user-defined static value but should evolve under a user-defined dynamics. We specify an overdamped spring dynamics, in which

$$\dot{r}^*_{sep} = -k_0\left(r_{sep} - l_0\right), \tag{5}$$

where the asterisk denotes that this is the target dynamics of r_{sep}. For steps of duration $\Delta t = N_{dt}dt$ (*cf.* Fig. 4), we use the finite-difference approximation to the change in r_{sep} from the target dynamics at step s,

$$r^*_{sep}(s+1) = r_{sep}(s) - \Delta t k_0\left(r_{sep}(s) - l_0\right), \tag{6}$$

to form the reward function

$$R(s) = -\left|r^*_{sep}(s+1) - r_{sep}(s+1)\right|, \tag{7}$$

where $r_{sep}(s+1)$ is the actual separation obtained by evolving the dynamics under the activity field chosen at step s. As the RL program learns to better mimic the prescribed dynamics the reward approaches zero from below. The state observed by the program and the episode initialization procedure are the same as in the previous task.

We train the RL program to achieve these target dynamics by leveraging a recently demonstrated modality of defect motion under activity gradients, rather than the modality of +1/2 defect propulsion under constant activity used in the previous task. −1/2 defects have been shown to couple to second-order and higher gradients in $\alpha(\mathbf{r})$,[35] and, using this fact, we parameterize the activity field in the vicinity of the −1/2 defect as $\alpha(r;\mathrm{d}e\alpha_0,1,5)$ in eqn (4), which has non-zero gradients of all orders. We allow the RL program to adjust the amplitude α_0 of the activity field (in the range [0,12]) and the horizontal offset r_- between the −1/2 defect and the center of the activity field (in the range [−10,10]); see Fig. 2 and 6A.

We fix $l_0 = 50$ in eqn (5) and vary k_0, the overdamped spring constant. We show typical trajectories obtained with a trained policy for $k_0 = 0.0015$ in Fig. 6B and C. The policy simultaneously adjusts the amplitude and offset of the activity profile relative to the −1/2 defect so that the defect separation r_{sep} gradually approaches the "rest length" of $l_0 = 50$. Viewing several episodes for policies trained on $k_0 = 0.001$ and $k_0 = 0.004$, we see that the deviation from the target decays roughly exponentially with a rate that depends on the target stiffness (Fig. 6D).

We consider a range of k_0 values and average k_0^{fit}, which we obtain as the fitted rate of exponential approach of $r_{sep}(s)$ to l_0, over the last 40 episodes of each experiment and over 10 experiments using different random initial seeds (Fig. 6E). We observe good agreement between k_0 and k_0^{fit}, albeit with deviations at high values of k_0. At these values the policy has difficulty pulling the defects faster than the nematic material timescales allow. Despite these limitations, the generally good agreement between k_0 and k_0^{fit} indicates that the RL program can utilize a range of physical effects (including coupling to gradients of activity) to pull defects, and these can be made to obey target dynamics rather than just target static configurations.

C. Rotating a +1/2 defect

Finally, we consider the task of reorienting the +1/2 defect by coupling its orientation vector to gradients in the activity field. Describing the defect's orientation by $\zeta_+ \equiv \sin(\phi_+)$, we specify the dynamical law

$$\dot{\zeta}^*_+ = -k_\theta \zeta_+, \tag{8}$$

corresponding to a overdamped spring with rest length 0 acting on the variable ζ_+. We use $\sin(\phi_+)$ instead of ϕ_+ to avoid the discontinuity at $\phi_+ = 0$ and $\phi_+ = 2\pi$. As in the previous task we use the finite-difference approximation for $\zeta^*_+(s+1)$ to compute the reward $R(s) = -\left|\zeta^*_+(s+1) - \zeta_+(s+1)\right|$. The state observed by the RL program is $S(s) = \zeta_+(s)$, and its action is the angle ϕ_α, depicted in Fig. 2 and 7A, between $\hat{\mathbf{e}}$ and the center of the activity profile $\alpha(r;-7.5,5,1)$, which is centered $r_+ = 5$ lattice units away from the defect. We draw initial conditions by taking a pair of defects vertically aligned and horizontally separated by 50 lattice units, and then rotating the nematic director everywhere by a random angle chosen uniformly from the range [−0.4,0.4] rads.

We vary the stiffness k_θ in eqn (8). We show typical trajectories obtained with a trained policy for $k_\theta = 0.0007$ in Fig. 7B and C. Under the policy, the angle ϕ_α fluctuates around a mean

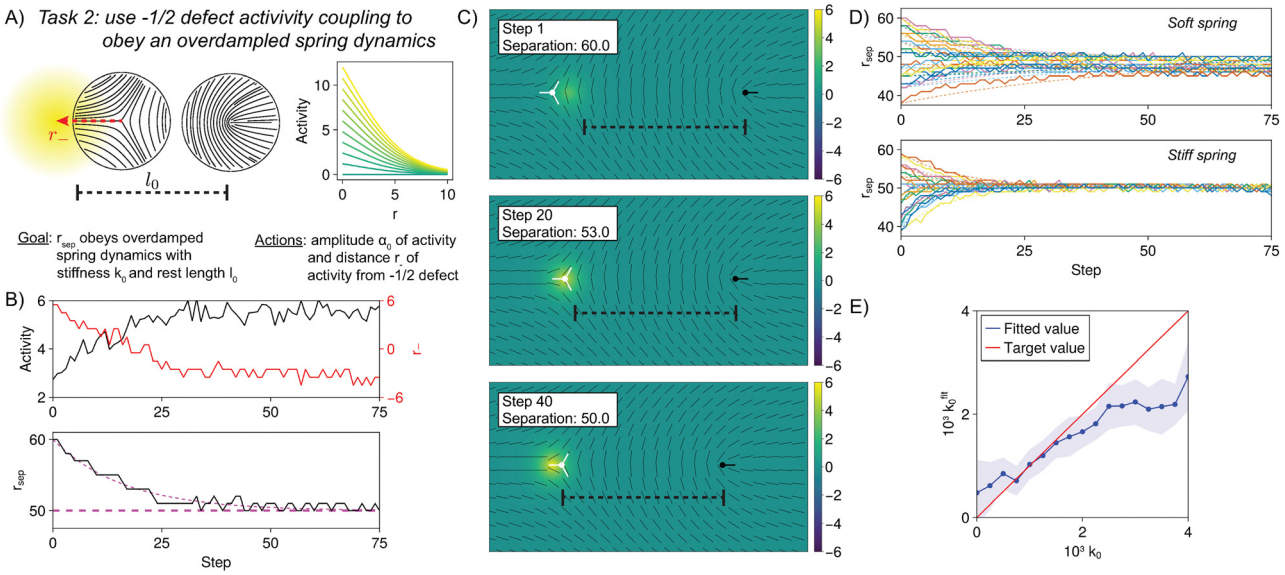

Fig. 6 Results for task 2: for r_{sep} to obey the dynamics of an overdamped spring with rest length $l_0 = 50$ and varying stiffness k_0. (A) Schematic illustration of the task, in which an inhomogeneous activity field is applied near the −1/2 defect. The position of the activity field and its amplitude are controlled as actions, and its radial profile for different amplitudes is shown on the right. (B) Top: Trajectories of the activity amplitude (black) and horizontal distance from the −1/2 defect (red) for a trained RL policy. Bottom: The resulting horizontal separation r_{sep} trajectory. The thick dashed line on the bottom represents $l_0 = 50$, the target stiffness is $k_0 = 0.0015$, and the thin dashed line denotes an exponential fit with $k_0^{fit} = 0.00124$. (C) Snapshots of the active nematic field at the end of three different steps for the episode in panel (B). See Fig. 3B caption for a description of the visualization. The black dashed line represents the target separation $l_0 = 50$. (D) Top: A set of 20 trajectories of r_{sep} with a trained RL policy for $k_0 = 0.001$, with exponential fits shown as dashed lines. Bottom: Same as top, but with $k_0 = 0.004$. (E) Fitted values k_0^{fit} plotted against their target values of k_0. Standard deviations are shown as shaded areas. For this plot we exclude episodes which randomly start within 2 of $r_{sep} = 50$ to focus on trajectories in which r_{sep} changes appreciably during the episode. See Video S1 (ESI†) for a movie of the trained RL algorithm performing this task.

value so as to cause the +1/2 defect to rotate at approximately the desired exponential rate, k_θ, toward $\phi_+ = \zeta_+ = 0$. In Fig. 7D we show several episodes for $k_\theta = 0.00025$ and $k_\theta = 0.0007$, indicating a clear difference in the learned decay rate of ζ_+ toward 0. Fitting k_θ^{fit} for the last 40 episodes of each experiment and over 10 experiments with different random initial seeds, we observe good agreement with the target value of k_θ (Fig. 7E). We thus see that RL can make both the positions and orientations of defects appear to follow dynamical laws that differ from those intrinsic to the system.

IV. Discussion

Here, we explored a computational strategy for tailoring activity fields to guide the dynamics of active nematic defects through closed-loop control policies learned by RL. RL offers a model-free approach for learning feedback control policies that implement a desired dynamical interaction law *via* trial-and-error exploration of the action space so as to maximize the reward. This approach presents a practically viable alternative to optimal control methods[30,32] or human-designed techniques,[35] which require accurate model specification. Relatedly, recent work suggests that RL can offer performance advantages over optimal control in steering racing drones by better accommodating model uncertainty.[56] Our focus here was on proof of principle using simple tasks. Future research could directly compare RL and optimal control in terms of the robustness of their policies for controlling active nematics.

Our current implementation has several limitations that could be improved in future work. First, we employed a fairly restrictive parameterization of the controlled activity field, chosen to ensure a low-dimensional and easily trainable action space. This design supports the idea that simple spatio-temporal patterns can couple effectively to defect dynamics in active nematics. However, future work could refine this approach, for example, by using multi-agent control[40,57] to explore higher-dimensional settings that might allow reconstruction of more spatially intricate control fields such as those shown to function as precise "topological tweezers".[35,36] Second, our model of active nematics includes several simplifying assumptions: we adopt a high-friction limit,[35,58] assume negligible density variations,[45] and omit detailed modeling of the dynamics of the activity-carrying agents, such as dispersal.[59,60] These choices were made for computational efficiency, as the RL algorithm requires many trial iterations to train. Our algorithm converged in under three hours without heavy optimization, however, suggesting substantial room to scale up in future implementations. A key advantage of the RL framework is its agnosticism to the specific physical model, which means that, at least in principle, it can learn effective control strategies across a range of active fluid models. Finally, our reward computation could be improved. We observe small deviations from ideal spring-like behavior in cases with soft spring constants (near $k_0 = 0$) or when the defects are near their

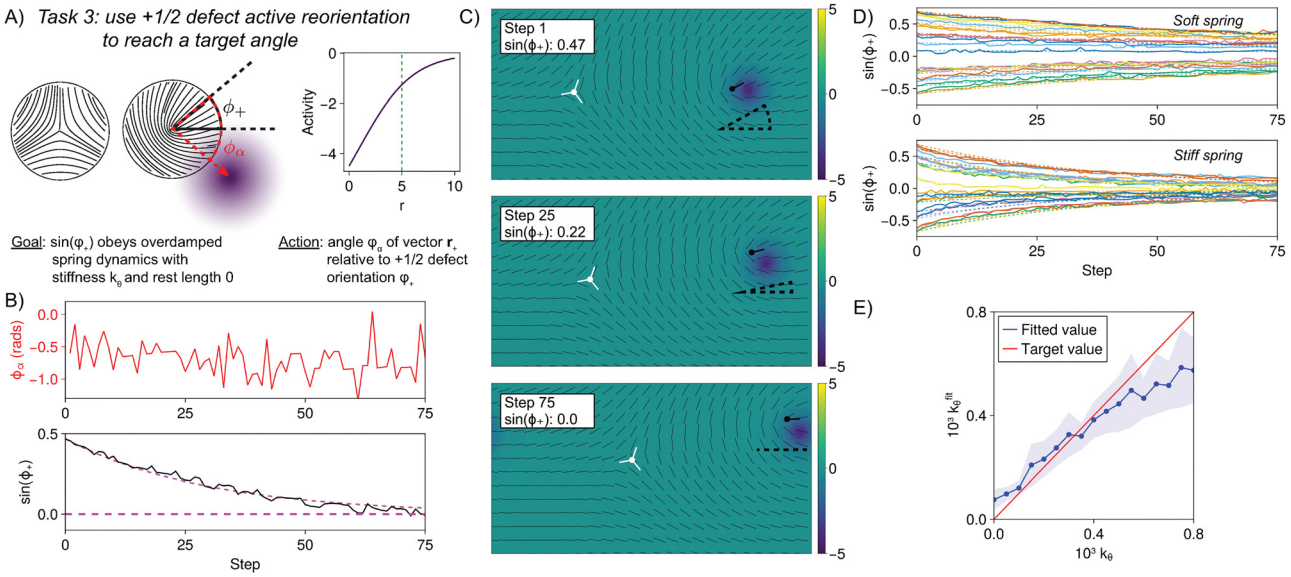

Fig. 7 Results for task 3: for $\sin(\phi_+)$ to obey the dynamics of an overdamped spring with zero rest length and varying stiffness k_θ. (A) Schematic illustration of the task, in which an inhomogeneous activity field is applied near the +1/2 defect. The negative angle ϕ_α is indicated. This angle is controlled as an action, and the activity field's radial profile is shown on the right. The dashed green line denotes the position of the +1/2 defect in the activity field at the start of each step. (B) Top: Trajectory of the angular position ϕ_α of the activity field for a trained RL policy. Bottom: The resulting trajectory for orientation of the +1/2 defect $\sin(\phi_+)$. The thick dashed line on the bottom represents $\sin(\phi_+) = 0$, the target stiffness is $k_\theta = 0.0007$, and the thin dashed line denotes an exponential fit with $k_\theta^{\text{fit}} = 0.00067$. (C) Snapshots of the active nematic field at the end of three different steps for the episode in panel (B). See Fig. 3B caption for a description of the visualization. The black dashed wedge represents the current orientation $\sin(\phi_+)$ which approaches the target value of 0. (D) Top: A set of 20 trajectories of $\sin(\phi_+)$ with a trained RL policy for $k_\theta = 0.00025$, with exponential fits shown as dashed lines. Bottom: Same as top, but with $k_\theta = 0.0008$. (E) Fitted values k_θ^{fit} plotted against the target value of k_θ. For this plot we exclude episodes which randomly start within 0.05 of $\sin(\phi_+) = 0$ to focus on trajectories in which $\sin(\phi_+)$ changes appreciably during the episode. See Video S2 (ESI†) for a movie of the trained RL algorithm performing this task.

equilibrium separation. In these regimes, the expected defect motions are minimal, so small erroneous motions are weakly penalized. Achieving more uniform adherence to spring-like behavior may require a reweighted reward function, and further refinements could enable more complex, multi-objective control tasks.

The scientific merit of demonstrating control over active nematics *via* RL is at least two-fold. First, as mentioned earlier, RL offers a practical method for engineering soft active matter systems with minimal reliance on accurate model specification, although it does present several technical challenges. When deploying RL to control a new system, careful attention must be given to parameterizing the control fields and designing the reward functions in a way that allows the algorithm to efficiently learn to solve the abstract optimization problem. Like other machine learning applications, this process involves fine-tuning various hyperparameters such as learning rates, neural network architectures, *etc.* Additionally, reliable state observations are important (although there exist stochastic variants of RL using uncertain state measurements), and accurately measuring nematic field configurations in experiments requires considerable technical attention.[61,62] Despite these challenges, RL has a proven track record of solving highly complex problems, often surpassing human capabilities,[63,64] making its application to controlling active matter systems quite promising.

Second, the ability of RL programs to successfully manipulate the dynamics of active nematic defects using very low-dimensional state and action spaces suggests that an effective low-dimensional description is sufficiently accurate to capture the defects' dynamics. This finding supports the theoretical arguments presented in ref. 35 and 45. Relatedly, in previous work we demonstrated that dynamic activity fields can be iteratively constructed without RL and without knowledge of the active nematic dynamics using physically motivated yet imperfect local feedback rules.[65] Recent experiments have similarly demonstrated closed-loop control over active nematic flow speeds which matches predictions from a highly coarse-grained model.[18] We have also demonstrated that low-dimensional RL may in principle be used to provide feedback control over other biomimetic mechanochemical protein networks such as the ultrafast Ca^{2+}-powered contractile proteins found in certain protists.[27,28] The sufficiency of imperfect, low-dimensional information to guide active nematics defects has implications for designing experimentally efficient feedback control over living systems and for understanding what information is sufficient to enable biologically plausible control over mechanochemical dynamics during biophysical processes like morphogenesis.[7–15]

Data availability

The code for this paper can be found at **https://github.com/svaikunt/RLNematics/**.

Conflicts of interest

There are no conflicts to declare.

Appendices

Active nematic equations of motion

Active nematic fluids are described by an order parameter $\mathbf{Q}$, which is a symmetric and traceless tensor:

$$\mathbf{Q} = q\left(\hat{\mathbf{n}}\hat{\mathbf{n}} - \frac{1}{d}\mathbf{I}\right), \tag{9}$$

where $\hat{\mathbf{n}}$ is a unit director, q measures the degree of polarization, d is the dimensionality, and $\mathbf{I}$ is the identity tensor. Our treatment follows the experimentally-motivated model in ref. 21, 52, 66, which uses $d = 3$ but projects the field onto a 2D plane to which the nematic is spatially confined, while allowing it to rotate out of this plane. We note that this differs slightly from the strictly 2D treatment used in several theoretical works such as ref. 35, 45.

The tensor $\mathbf{Q}$ couples to a flow field $\mathbf{v}$ and spontaneously seeks to minimize its local free energy. We consider the overdamped limit and the limit of high substrate friction (as in ref. 35, 45, 58), yielding the following equations of motion:

$$\partial_t Q_{ij} = S_{ij}(\mathbf{v}) + \Gamma_H H_{ij}, \tag{10}$$

$$v_i = \gamma_v^{-1}\partial_k(\sigma^a_{ik}(\mathbf{Q}) + \sigma^E_{ik}(\mathbf{Q})). \tag{11}$$

Here, H_{ij} is the symmetric and traceless part of $-\frac{\delta F}{\delta Q_{ij}}$ where F is the free energy functional (see below), S_{ij} is a flow-coupling term, and γ_v is a coefficient of friction. In this overdamped limit, $\mathbf{v}$ is an instantaneous function of $\mathbf{Q}$, so eqn (10) is closed in $\mathbf{Q}$. This simplification allows for computational efficiency which is advantageous in training RL programs. The active stress tensor is given by[67,68]

$$\sigma^a_{ij} = -\alpha Q_{ij}, \tag{12}$$

and the Ericksen stress tensor is given by[69]

$$\begin{aligned}\sigma^E_{ij} &= f\delta_{ij} - \xi H_{ik}\left(Q_{kj} + \frac{1}{3}\delta_{kj}\right) - \xi\left(Q_{ik} + \frac{1}{3}\delta_{ik}\right)H_{kj} \\ &+ 2\xi\left(Q_{ij} + \frac{1}{3}\delta_{ij}\right)H_{kl}Q_{kl} - \partial_j Q_{kl}\frac{\delta F}{\delta\partial_i Q_{kl}} \\ &+ Q_{ik}H_{kj} - H_{ik}Q_{kj}.\end{aligned} \tag{13}$$

The Landau–de Gennes free energy is:

$$F = \int \mathrm{d}\mathbf{r} f(\mathbf{r}), \tag{14}$$

where

$$\begin{aligned}f &= \frac{A_0}{2}\left(1 - \frac{U}{3}\right)\mathrm{Tr}(\mathbf{Q}^2) - \frac{A_0 U}{3}\mathrm{Tr}(\mathbf{Q}^3) \\ &+ \frac{A_0 U}{4}(\mathrm{Tr}(\mathbf{Q}^2))^2 + \frac{L}{2}(\partial_k Q_{lm})^2.\end{aligned} \tag{15}$$

The isotropic part of the passive stress in this model depends only on f as we neglect pressure contributions resulting from variations in density of the fluid.[45]

The flow coupling term is

$$S_{ij}(\mathbf{v}) = -v_k\partial_k Q_{ij} + \Phi_{ik}Q_{kj}^{+} + Q_{ik}^{+}\Phi_{kj} - 2\xi Q_{ij}^{+}(Q_{kl}\partial_k v_l),$$

where

$$Q_{ij}^{+} = Q_{ij} + \frac{1}{3}\delta_{ij}, \tag{16}$$

$$\Psi_{ij} = \frac{1}{2}(\partial_i v_j + \partial_j v_i), \tag{17}$$

$$\Omega_{ij} = \frac{1}{2}(\partial_i v_j - \partial_j v_i), \tag{18}$$

and

$$\Phi_{ij} = \xi\Psi_{ij} - \Omega_{ij}. \tag{19}$$

In these equations, ξ, A_0, U, Γ_H, L, γ_v are parameters whose meanings are described in ref. 21, 52, 66, to which we refer the reader for additional details on this model. Throughout the paper we use $\xi = 0.7$, $A_0 = 0.1$, $U = 3.5$, $\Gamma_H = 1.5$, $L = 0.1$, and $\gamma_v = 10$ (all in simulation units). For these choices the equilibrium nematic persistence length $\sqrt{L/A_0}$ is the same as one length unit.[53]

Acknowledgements

We wish to thank Alexandra Lamtyugina, Luca Scharrer, Suraj Shankar, Grant Rotskoff, Michael Hagan, Saptorshi Ghosh, and Aparna Baskaran for helpful discussions. This work was mainly supported by DOE BES Grant DE-SC00197 to SV (theory, design). A. R. D. acknowledges support from National Science Foundation (NSF) award MCB-2201235, and the University of Chicago Materials Research Science and Engineering Center, which is funded by the NSF under award number DMR-2011854. The authors acknowledge the University of Chicago's Research Computing Center for computing resources.

References

1 A. Doostmohammadi, J. Ignés-Mullol, J. M. Yeomans and F. Sagués, Active nematics, *Nat. Commun.*, 2018, **9**(1), 3246.
2 G. Tóth, C. Denniston and J. M. Yeomans, Hydrodynamics of topological defects in nematic liquid crystals, *Phys. Rev. Lett.*, 2002, **88**(10), 105504.
3 M. Serra, L. Lemma, L. Giomi, Z. Dogic and L. Mahadevan, Defect-mediated dynamics of coherent structures in active nematics, *Nat. Phys.*, 2023, **19**(9), 1355–1361.
4 S. Shankar, A. Souslov, M. J. Bowick, M. C. Marchetti and V. Vitelli, Topological active matter, *Nat. Rev. Phys.*, 2022, **4**(6), 380–398.
5 P. Guillamat, C. Blanch-Mercader, G. Pernollet, K. Kruse and A. Roux, Integer topological defects organize stresses

driving tissue morphogenesis, *Nat. Mater.*, 2022, **21**(5), 588–597.
6 K. Kawaguchi, R. Kageyama and M. Sano, Topological defects control collective dynamics in neural progenitor cell cultures, *Nature*, 2017, **545**(7654), 327–331.
7 Y. Maroudas-Sacks, L. Garion, L. Shani-Zerbib, A. Livshits, E. Braun and K. Keren, Topological defects in the nematic order of actin fibres as organization centres of Hydra morphogenesis, *Nat. Phys.*, 2021, **17**(2), 251–259.
8 Y. Ravichandran, M. Vogg, K. Kruse, D. J. Pearce and A. Roux, Topology changes of Hydra define actin orientation defects as organizers of morphogenesis, *Sci. Adv.*, 2025, **11**(3), eadr9855.
9 Z. Wang, M. C. Marchetti and F. Brauns, Patterning of morphogenetic anisotropy fields, *Proc. Natl. Acad. Sci. U. S. A.*, 2023, **120**(13), e2220167120.
10 L. Metselaar, J. M. Yeomans and A. Doostmohammadi, Topology and morphology of self-deforming active shells, *Phys. Rev. Lett.*, 2019, **123**(20), 208001.
11 L. J. Ruske and J. M. Yeomans, Morphology of active deformable 3D droplets, *Phys. Rev. X*, 2021, **11**(2), 021001.
12 L. A. Hoffmann, L. N. Carenza, J. Eckert and L. Giomi, Theory of defect-mediated morphogenesis, *Sci. Adv.*, 2022, **8**(15), eabk2712.
13 Z. Zhao, H. Li, Y. Yao, Y. Zhao, F. Serra and K. Kawaguchi, *et al.*, Integer topological defects offer a methodology to quantify and classify active cell monolayers, *Nat. Commun.*, 2025, **16**(1), 2452.
14 D. Pearce, C. Thibault, Q. Chaboche and C. Blanch-Mercader, Passive defect driven morphogenesis in nematic membranes, *Phys. Rev. Lett.*, 2025, **134**(1), 018402.
15 D. Krommydas, L. N. Carenza and L. Giomi, Hydrodynamic enhancement of p-atic defect dynamics, *Phys. Rev. Lett.*, 2023, **130**(9), 098101.
16 A. Bailles, E. W. Gehrels and T. Lecuit, Mechanochemical principles of spatial and temporal patterns in cells and tissues, *Annu. Rev. Cell Dev. Biol.*, 2022, **38**(1), 321–347.
17 D. B. Brückner and E. Hannezo, Tissue Active Matter: Integrating Mechanics and Signaling into Dynamical Models, *Cold Spring Harbor Perspect. Biol.*, 2024, a041653.
18 K. Nishiyama, J. Berezney, M. M. Norton, A. Aggarwal, S. Ghosh, M. F. Hagan, *et al.*, Closed-loop control of active nematic flows, *arXiv*, 2024, preprint, arXiv:240814414, DOI: **10.48550/arXiv.2408.14414**.
19 T. D. Schindler, L. Chen, P. Lebel, M. Nakamura and Z. Bryant, Engineering myosins for long-range transport on actin filaments, *Nat. Nanotechnol.*, 2014, **9**(1), 33–38.
20 M. Nakamura, L. Chen, S. C. Howes, T. D. Schindler, E. Nogales and Z. Bryant, Remote control of myosin and kinesin motors using light-activated gearshifting, *Nat. Nanotechnol.*, 2014, **9**(9), 693–697.
21 R. Zhang, S. A. Redford, P. V. Ruijgrok, N. Kumar, A. Mozaffari and S. Zemsky, *et al.*, Spatiotemporal control of liquid crystal structure and dynamics through activity patterning, *Nat. Mater.*, 2021, **20**(6), 875–882.
22 L. M. Lemma, M. Varghese, T. D. Ross, M. Thomson, A. Baskaran and Z. Dogic, Spatio-temporal patterning of extensile active stresses in microtubule-based active fluids, *PNAS Nexus*, 2023, **2**(5), pgad130.
23 S. Chandrasekar, J. R. Beach and P. W. Oakes, Shining a light on RhoA: Optical control of cell contractility, *Int. J. Biochem. Cell Biol.*, 2023, **161**, 106442.
24 F. Yang, S. Liu, H. J. Lee, R. Phillips and M. Thomson, Dynamic flow control through active matter programming language, *Nat. Mater.*, 2025, 1–11.
25 R. Sakamoto and M. P. Murrell, Mechanical power is maximized during contractile ring-like formation in a biomimetic dividing cell model, *Nat. Commun.*, 2024, **15**(1), 9731.
26 I. Linsmeier, S. Banerjee, P. W. Oakes, W. Jung, T. Kim and M. P. Murrell, Disordered actomyosin networks are sufficient to produce cooperative and telescopic contractility, *Nat. Commun.*, 2016, **7**(1), 12615.
27 X. Lei, C. Floyd, L. C. Ferrer, T. Chakrabortty, N. Chandrasekharan and A. Dinner, Light-induced reversible assembly and actuation in ultrafast Ca^{2+}-driven chemomechanical protein networks, *bioRxiv*, 2025, preprint, 2025-03, DOI: **10.1101/2025.03.03.641304**.
28 C. Floyd, A. T. Molines, X. Lei, J. E. Honts, F. Chang and M. W. Elting, *et al.*, A unified model for the dynamics of ATP-independent ultrafast contraction, *Proc. Natl. Acad. Sci. U. S. A.*, 2023, **120**(25), e2217737120.
29 T. Litschel, D. Vavylonis and D. A. Weitz, 3D printing cytoskeletal networks: ROS-induced filament severing leads to surge in actin polymerization, *bioRxiv*, 2025, preprint, 2025-03, DOI: **10.1101/2025.03.19.644260**.
30 M. M. Norton, P. Grover, M. F. Hagan and S. Fraden, Optimal control of active nematics, *Phys. Rev. Lett.*, 2020, **125**(17), 178005.
31 S. Ghosh, C. Joshi, A. Baskaran and M. F. Hagan, Spatiotemporal control of structure and dynamics in a polar active fluid, *Soft Matter*, 2024, **20**, 7059–7071.
32 S. Ghosh, A. Baskaran and M. F. Hagan, Achieving designed texture and flows in bulk active nematics using optimal control theory, *J. Chem. Phys.*, 2025, **162**, 134902.
33 J. Alvarado, E. Teich, D. Sivak and J. Bechhoefer, Optimal Control in Soft and Active Matter, *arXiv*, 2025, preprint, arXiv:250408676, DOI: **10.48550/arXiv.2504.08676**.
34 C. G. Wagner, M. M. Norton, J. S. Park and P. Grover, Exact coherent structures and phase space geometry of preturbulent 2d active nematic channel flow, *Phys. Rev. Lett.*, 2022, **128**(2), 028003.
35 S. Shankar, L. V. Scharrer, M. J. Bowick and M. C. Marchetti, Design rules for controlling active topological defects, *Proc. Natl. Acad. Sci. U. S. A.*, 2024, **121**(21), e2400933121.
36 W. T. Irvine, A. D. Hollingsworth, D. G. Grier and P. M. Chaikin, Dislocation reactions, grain boundaries, and irreversibility in two-dimensional lattices using topological tweezers, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**(39), 15544–15548.
37 R. S. Sutton, *Reinforcement learning: an introduction*, A Bradford Book, 2018.
38 J. Bechhoefer, *Control theory for physicists*, Cambridge University Press, 2021.

39 M. J. Falk, V. Alizadehyazdi, H. Jaeger and A. Murugan, Learning to control active matter, *Phys. Rev. Res.*, 2021, **3**(3), 033291.
40 S. Chennakesavalu and G. M. Rotskoff, Probing the theoretical and computational limits of dissipative design, *J. Chem. Phys.*, 2021, **155**(19), 194114.
41 S. Chennakesavalu, S. K. Manikandan, F. Hu and G. M. Rotskoff, Adaptive nonequilibrium design of actin-based metamaterials: Fundamental and practical limits of control, *Proc. Natl. Acad. Sci. U. S. A.*, 2024, **121**(8), e2310238121.
42 U. Khadka, V. Holubec, H. Yang and F. Cichos, Active particles bound by information flows, *Nat. Commun.*, 2018, **9**(1), 3864.
43 T. Bäuerle, A. Fischer, T. Speck and C. Bechinger, Self-organization of active particles by quorum sensing rules, *Nat. Commun.*, 2018, **9**(1), 3232.
44 X. Tang and J. V. Selinger, Orientation of topological defects in 2D nematic liquid crystals, *Soft Matter*, 2017, **13**(32), 5481–5490.
45 S. Shankar, S. Ramaswamy, M. C. Marchetti and M. J. Bowick, Defect unbinding in active nematics, *Phys. Rev. Lett.*, 2018, **121**(10), 108002.
46 C. D. Schimming, C. Reichhardt and C. Reichhardt, Analytical model for the motion and interaction of two-dimensional active nematic defects, *Soft Matter*, 2025, **21**(1), 122–136.
47 C. Joshi, S. Ray, L. M. Lemma, M. Varghese, G. Sharp and Z. Dogic, *et al.*, Data-driven discovery of active nematic hydrodynamics, *Phys. Rev. Lett.*, 2022, **129**(25), 258001.
48 D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra and M. Riedmiller, *Deterministic policy gradient algorithms*, In: International Conference on Machine Learning. Pmlr, 2014, pp. 387–395.
49 T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, *et al.*, Continuous control with deep reinforcement learning, *arXiv*, 2015, preprint, arXiv:150902971, DOI: **10.48550/arXiv.1509.02971**.
50 C. Floyd, S. Vaikuntanathan and A. R. Dinner, Simulating structured fluids with tensorial viscoelasticity, *J. Chem. Phys.*, 2023, **158**(5), 054906.
51 C. Floyd, A. R. Dinner and S. Vaikuntanathan, Pattern formation in odd viscoelastic fluids, *Phys. Rev. Res.*, 2024, **6**(3), 033100.
52 S. A. Redford, J. Colen, J. L. Shivers, S. Zemsky, M. Molaei and C. Floyd, *et al.*, Motor crosslinking augments elasticity in active nematics, *Soft Matter*, 2024, **20**(11), 2480–2490.
53 E. J. Hemingway, P. Mishra, M. C. Marchetti and S. M. Fielding, Correlation lengths in hydrodynamic models of active nematics, *Soft Matter*, 2016, **12**(38), 7943–7952.
54 J. Tian, other contributors. ReinforcementLearning.jl: A Reinforcement Learning Package for the Julia Programming Language; 2020. Available from: **https://github.com/JuliaReinforcementLearning/ReinforcementLearning.jl**.
55 L. Giomi, M. J. Bowick, P. Mishra, R. Sknepnek and M. Cristina Marchetti, Defect dynamics in active nematics, *Philos. Trans. R. Soc., A*, 2014, **372**(2029), 20130365.
56 Y. Song, A. Romero, M. Müller, V. Koltun and D. Scaramuzza, Reaching the limit in autonomous racing: Optimal control *versus* reinforcement learning, *Sci. Rob.*, 2023, **8**(82), eadg1462.
57 R. Lowe, Y. I. Wu, A. Tamar, J. Harb, O. Pieter Abbeel and I. Mordatch, Multi-agent actor-critic for mixed cooperative-competitive environments, *Adv. Neural Inf. Process. Syst.*, 2017, **30**, 6382–6393.
58 A. Doostmohammadi, M. F. Adamer, S. P. Thampi and J. M. Yeomans, Stabilization of active matter by flow-vortex lattices and defect ordering, *Nat. Commun.*, 2016, **7**(1), 10557.
59 R. C. Coelho, N. A. Araújo and M. M. T. da Gama, Dispersion of activity at an active-passive nematic interface, *Soft Matter*, 2022, **18**(39), 7642–7653.
60 T. E. Bate, M. E. Varney, E. H. Taylor, J. H. Dickie, C. C. Chueh and M. M. Norton, *et al.*, Self-mixing in microtubule-kinesin active fluid from nonuniform to uniform distribution of activity, *Nat. Commun.*, 2022, **13**(1), 6573.
61 Y. Li, Z. Zarei, P. N. Tran, Y. Wang, A. Baskaran and S. Fraden, *et al.*, A machine learning approach to robustly determine director fields and analyze defects in active nematics, *Soft Matter*, 2024, **20**(8), 1869–1883.
62 P. N. Tran, S. Ray, L. Lemma, Y. Li, R. Sweeney and A. Baskaran, Deep-learning optical flow for measuring velocity fields from experimental data, *Soft Matter*, 2024, **20**(36), 7246–7257.
63 D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre and G. Van Den Driessche, *et al.*, Mastering the game of Go with deep neural networks and tree search, *Nature*, 2016, **529**(7587), 484–489.
64 G. Reddy, J. Wong-Ng, A. Celani, T. J. Sejnowski and M. Vergassola, Glider soaring *via* reinforcement learning in the field, *Nature*, 2018, **562**(7726), 236–239.
65 C. Floyd, A. R. Dinner and S. Vaikuntanathan, Learning to control non-equilibrium dynamics using local imperfect gradients, *arXiv*, 2024, preprint, arXiv:240403798, DOI: **10.48550/arXiv.2404.03798**.
66 J. Colen, M. Han, R. Zhang, S. A. Redford, L. M. Lemma and L. Morgan, *et al.*, Machine learning active-nematic hydrodynamics, *Proc. Natl. Acad. Sci. U. S. A.*, 2021, **118**(10), e2016708118.
67 Y. Hatwalne, S. Ramaswamy, M. Rao and R. A. Simha, Rheology of active-particle suspensions, *Phys. Rev. Lett.*, 2004, **92**(11), 118101.
68 M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost and M. Rao, *et al.*, Hydrodynamics of soft active matter, *Rev. Mod. Phys.*, 2013, **85**(3), 1143–1189.
69 A. N. Beris and B. J. Edwards, *Thermodynamics of flowing systems: with internal microstructure. 36*, Oxford University Press, USA, 1994.