Soft Matter

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Soft Matter*, 2022, **18**, 617

Reinforcement learning reveals fundamental limits on the mixing of active particles†

Dominik Schildknecht, *[a] Anastasia N. Popova,[b] Jack Stellwagen[c] and Matt Thomson[d]

The control of far-from-equilibrium physical systems, including active materials, requires advanced control strategies due to the non-linear dynamics and long-range interactions between particles, preventing explicit solutions to optimal control problems. In such situations, Reinforcement Learning (RL) has emerged as an approach to derive suitable control strategies. However, for active matter systems, it is an important open question how the mathematical structure and the physical properties determine the tractability of RL. In this paper, we demonstrate that RL can only find good mixing strategies for active matter systems that combine attractive and repulsive interactions. Using analytic results from dynamical systems theory, we show that combining both interaction types is indeed necessary for the existence of mixing-inducing hyperbolic dynamics and therefore the ability of RL to find homogeneous mixing strategies. In particular, we show that for drag-dominated translational-invariant particle systems, mixing relies on combined attractive and repulsive interactions. Therefore, our work demonstrates which experimental developments need to be made to make protein-based active matter applicable, and it provides some classification of microscopic interactions based on macroscopic behavior.

Received 29th September 2021,
Accepted 11th December 2021

DOI: 10.1039/d1sm01400e

rsc.li/soft-matter-journal

Open Access Article. Published on 13 December 2021. Downloaded on 4/8/2026 8:50:09 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

1 Introduction

Mixing is crucial in applications ranging from large-scale chemical processes in industries such as petroleum or food[2] down to microscopic scales, including microfluidic applications in medicine.[3,4] However, microscopic mixing is challenging because microscale systems are typically drag-dominated, therefore, efficient mixing strategies are not obvious. While current microfluidic solutions exist for microscopic mixing and material transport, they rely on external pumps, prefabricated device geometries,[5–8] and (active) additives[9–11] so that they are limited in their reconfigurability. In contrast, natural cells use the self-organization of interacting components to make microscopic transport and mixing more compact and energy-efficient.[12]

To better understand nature, and to replicate it for applications, researchers started to study so-called active matter. Here, active matter refers to systems of proteins or other particles that continuously consume energy to achieve non-equilibrium dynamics, which can self-organize to achieve specific tasks. In particular, recent work suggested using active matter to achieve macroscopic tasks, such as generating fluid flows,[13–15] flow rectification,[16] and equilibration of glassy systems.[17]

The unsolved scientific challenge is how to assemble the existing (microscopic) active matter systems to achieve macroscopic results such as mixing and material transport. While recently, machine learning was applied to various topics in active matter (see ref. 18 for a review), the literature on active matter control is limited. While in macroscopic systems, where individual agent properties can be controlled, various goals could be achieved, such as navigation through fluids at various Reynolds numbers[19–25] and self-organization emerging in a multi-agent setting to solve several tasks,[26–31] using microscopic systems to achieving macroscopic behavior is challenging. Indeed, while some limited control was achieved, such as the ability to generate simple fluid flows,[13–16,32] achieve global particle momentum,[33] and accelerated equilibration of otherwise glassy systems,[17] control strategies for more complex scenarios are absent. The reason for this challenging behavior is that active matter systems are interaction-dominated, so strategies have to exert indirect control *via* agent–agent interactions rather than

[a] *Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. E-mail: dominik.schildknecht@gmail.com*
[b] *Applied and Computational Mathematics, California Institute of Technology, Pasadena CA, USA*
[c] *School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA*
[d] *Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. E-mail: mthomson@caltech.edu*

† Electronic supplementary information (ESI) available: Additional details to the simulations, and supplemental figures. The code for the simulation can be found in ref. 1, and the OpenAI gym environment can be found on https://github.com/domischi/SpringBoxDDPG. The data and analysis scripts are accessible. See DOI: 10.1039/d1sm01400e

Open Access Article. Published on 13 December 2021. Downloaded on 4/8/2026 8:50:09 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

single–agent properties. Therefore, finding successful strategies to achieve macroscopic goals requires advanced control methods, including policy-based reinforcement learning.

The key question we address in this paper is finding control strategies for the macroscopic task of mixing using protein-based active matter systems. In particular, we focus on finding mixing strategies of active particles by applying RL to our recent numerical active matter model.[1] If we find good mixing solutions for the active particles, these strategies could then be implemented experimentally to induce mixing in the surrounding medium, for example, *via* fluid–matter interactions. For our model of active particles, we observe that while RL fails to learn good strategies if only attractive or repulsive interactions are available, RL finds good strategies if attractive and repulsive interactions can be combined. We analyze this puzzling behavior using dynamical systems theory, particularly theory to hyperbolic dynamics and Anosov diffeomorphisms,[34] to prove that mixed interactions are indeed necessary to render the problem solvable. Our results, therefore, provide a guideline to make protein-based active matter applicable and answer the question of how different interactions can lead to macroscopic behavior asked in a recent review.[18]

2 Methods

The simulation loop is shown in Fig. 1a The reinforcement learning agent and the simulation interact by exchanging information. In particular, the reinforcement learning system controls the simulation by providing the activations for each bin $\mathscr{A}$. The simulation performs an update and provides the reinforcement learning algorithm with the measurement tensor $\mathscr{M}$, which is additionally used to compute the rewards. In this section, we will describe this process in more detail. In particular, we describe the simulation engine in Section 2.1, the reward function in Section 2.2, and the reinforcement learning framework in Section 2.3.

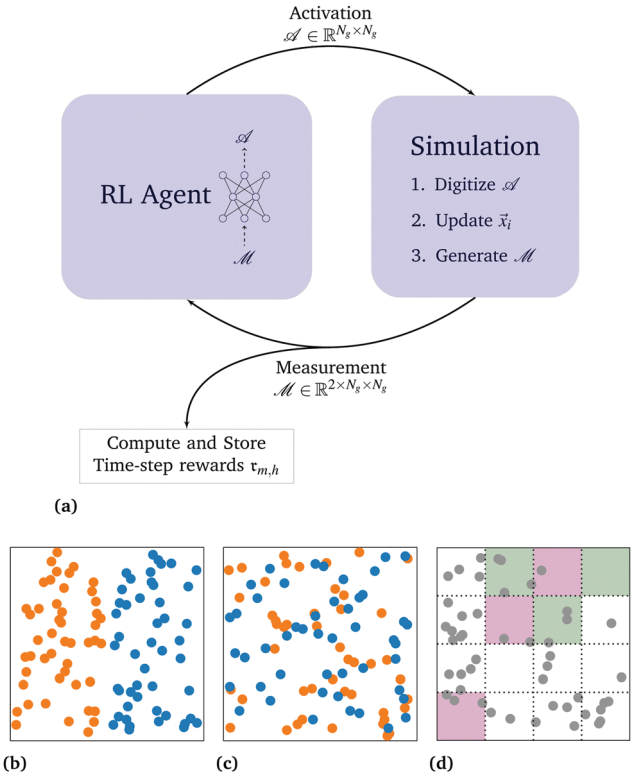


Fig. 1 Sketch of the setup: in (a), we describe the temporal evolution of a single simulation. In particular, we show how the two entities, the RL agent and the simulation, shown by blue boxes, interact with each other by passing information. In particular, the RL agent communicates to the simulation which activation pattern $\mathscr{A}$ should be used in the next time-step, and the simulation communicates the number of particles per tag and bin back to the RL agent using the measurement tensor $\mathscr{M}$. Additionally, this tensor is used in the rewards calculation, to train the RL agents. In (b), the initial condition is depicted with orange "left"-tagged particles and blue "right"-tagged particles. In (c), a good solution that we would like to achieve is depicted using the same color-coding. In (d), we demonstrate the binning and color-code a possible activation pattern with attractive-activation areas depicted in green, repulsive-activation areas in pink and non-activating areas in white.

2.1 Simulation and reinforcement learning environment

In this work, we will start by applying RL to a specific microscopic model of active matter. In particular, we want to select a sufficiently simple model to be amenable to RL while at the same time being suitable to describe various problems central to microscopic material transport and active matter. Microscopic models are often drag-dominated[35] and can be classified into continuum theories[10,11,15,36–40] and particle-based models.[1,41–43] While continuum theories are often preferred due to their analytic properties, we chose a particle-based model because they tend to be quicker to simulate small- and medium-sized systems, and therefore more amenable to RL requiring many simulations. In addition, since we are interested in general properties, a phenomenological model is sufficient, and we used the simulation platform proposed by us in ref. 1. It should be noted that our results will be generalized to a larger array of models in Section 4.

Our model[1] required only attractive interactions to describe the existing experimental platform.[14,15,44] However, this restriction will turn out to limit mixing. Hence, an extended model is introduced as follows: the system consists of N_p particles in two dimensions indexed by i, which can be in three states: attractive-activated ($p_i^a = 1$, $p_i^r = 0$), repulsive-activated ($p_i^a = 0$, $p_i^r = 1$), or inactivated ($p_i^a = 0$, $p_i^r = 0$). Particles can be activated if they were previously inactivated by being in the corresponding activation area (which will form our control input) and deactivate back to the inactivated state with a rate λ. Activated particles interact with other similar-activated particles by a spring potential.‡ The system's dynamics is described by the

‡ *I.e.*, attractive-activated particles only interact with other attractive-interacted particles, and repulsive-activated particles only interact with other repulsive-activated particles.

This journal is © The Royal Society of Chemistry 2022

Open Access Article. Published on 13 December 2021. Downloaded on 4/8/2026 8:50:09 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

equations of motion

$$\vec{x}_i(t+\Delta t) = \vec{x}(t) + \frac{\Delta t^2}{m}\vec{F}_i, \tag{1a}$$

$$\vec{F}_i = \vec{F}_i^a + \vec{F}_i^r, \tag{1b}$$

$$\vec{F}_i^a = -\nabla_i \left[\frac{k}{2} \sum_{\substack{j \neq i \\ r_c < |\vec{x}_i - \vec{x}_j| < R_c}} p_i^a p_j^a |\vec{x}_i - \vec{x}_j|^2 \right], \tag{1c}$$

$$\vec{F}_i^r = -\nabla_i \left[\frac{k}{2} \sum_{\substack{j \neq i \\ r_c < |\vec{x}_i - \vec{x}_j| < R_c}} p_i^r p_j^r (|\vec{x}_i - \vec{x}_j| - R_c)^2 \right], \tag{1d}$$

where $\vec{x}_i(t)$ is the position of particle i at time t, and $\vec{F}_i$ is the total force acting on it. The total force $\vec{F}_i$ is split into attractive and repulsive contributions, and at every times step, at least one of them vanishes because at least one of p_i^a and p_i^r is 0. It should be noted that the attractive and the repulsive force incorporate the same spring constant k, lower truncation r_c, and upper truncation R_c, and differ only by their rest length being 0 or R_c for the attractive and repulsive force, respectively. It should be noted that eqn (1) describes drag-dominated particles, because while these equations were derived from an inertial description, the limit of infinitely strong drag was taken to ensure numerical stability for large time-steps.[1]

Using this model of drag-dominated particle dynamics with controllable pairwise spring interactions, we built an RL environment using OpenAI gym.[45] In particular, $N_p = 96$ particles are initialized randomly in a box with periodic boundary conditions, asserting an equal number of particles on the two halves of the system (for future reference, tagged "left" and "right"). The system is then integrated according to eqn (1) with a time-step of $\Delta t = 0.05$ for $N_t = 100$ time-steps. A sketch of the initial condition can be found in Fig. 1b, and an example of a target state is depicted in Fig. 1c. The detailed simulation parameters can be found in the ESI.† The observation and control spaces are introduced by using an $N_g \times N_g = 4 \times 4$ square grid, as depicted in Fig. 1d. In particular, observations are given to the algorithm by providing a separate count for each tag (corresponding to the orange and blue colors in Fig. 1b and c) and each bin. The system is controlled by associating each square of the binning to either activation area or none at all so that the action space at every time-step is $3^{4\times4} \approx 43 \times 10^6$ dimensional if both interactions are included.§

It should be noted that the individual simulations are relatively small at only 96 particles and a coarse binning in a 4×4 grid. We chose such small system sizes because the individual simulations are run many thousands of times, and as such, they need to run very quickly. Nevertheless, because the scaling of an individual simulation is only $\mathcal{O}(N_p{}^2)$, future work could easily extend to larger particle numbers. In contrast, due to the exponential scaling of the action space with grid resolution, increasing the spatial resolution would slow down learning, and possibly more advanced methods such as curriculum learning would need to be implemented to achieve fine-grained control. However, the main result of this paper, namely the necessity of two types of interactions, will continue to hold independent of the number of particles or the grid resolution, as we will show in Section 4.

§ If only one interaction type is included, the action space at every time-step is $2^{4\times4} \approx 65 \times 10^3$ dimensional.

2.2 Reward functions

To apply RL, we need to measure how well mixed a given state is, *i.e.*, we need to define a functional giving a configuration as depicted in Fig. 1b a low score and one as in Fig. 1c a high score. Various approaches could be taken to define such a function: one could use variations of the continuous mix-norm[46–49] or an adversarial neural network approach to distinguish a well-mixed state from a segregated state. However, as we will demonstrate in Section 4, the main results of our paper will be unaffected by the choice of the reward function, and hence, the compute-intensive reward functions can be replaced by a simpler one. In particular, we use the measurement tensor $\mathscr{M}$, which describes the particle count with a specific tag per bin. Hence, $\mathscr{M}$ has the indices (tag, x-bin, y-bin), where tag $\in \{l, r\}$, and x-bin and y-bin $\in \{1, 2, \ldots, N_g\}$. Then, we define the mixing reward at a specific time-step as

$$\tau_m = -\frac{1}{\mathscr{N}_m} \sum_{x=1}^{N_g} \sum_{y=1}^{N_g} \left(\mathscr{M}_{l,xy} - \mathscr{M}_{r,xy}\right)^2 \in \left[-\frac{N_g^2}{2},\ 0\right], \tag{2}$$

where we suppressed the time dependence exhibited of $\mathscr{M}$ and τ_m for easier notation. In eqn (2), the minus sign ensures that the reward is larger for well-mixed states, and $\mathscr{N}_m = (N_p/N_g)^2$ is chosen so that the reward for the initial state is -1 on average. It should be noted that the minimal value of τ_m is much smaller than -1 and is attained if all "left" particles are in one bin and all "right" particles in another.

Since only using eqn (2) will lead to degenerate solutions that collapse all particles to dense clusters, we use an additional (time-step) homogeneity reward, constructed analogously to eqn (2):

$$\tau_h = -\frac{1}{\mathscr{N}_h} \sum_{x=1}^{N_g} \sum_{y=1}^{N_g} \left[\frac{N_p}{N_g^2} - \left(\mathscr{M}_{l,xy} + \mathscr{M}_{r,xy}\right)\right]^2 \in [-1, 0], \tag{3}$$

but rather than comparing the particle counts per tag in each bin, the homogeneity reward compares the number of particles in each bin $(\mathscr{M}_{l,ij} + \mathscr{M}_{r,ij})$ to the average number of particles $\frac{N_p}{N_g^2}$ per bin. Here, $\mathscr{N}_h = N_p{}^2\left(1 - \frac{1}{N_g^2}\right)$ so that the most inhomogeneous solution, namely all particles in a single bin, has a reward of -1. It should be noted that $\mathscr{N}_m$ and $\mathscr{N}_h$ are chosen so that τ_m and τ_h are -1 in the typical situation they should avoid, namely the segregated initial state, and a completely collapsed state, respectively. As such, using these values of $\mathscr{N}_m$ and $\mathscr{N}_h$ provides a natural normalization.

Open Access Article. Published on 13 December 2021. Downloaded on 4/8/2026 8:50:09 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

We introduce the episode rewards as the mean over the time-step rewards for an entire episode:

$$\Re_m = \frac{1}{N_t}\sum_{t=1}^{N_t}\tau_m^{(t)}, \text{ and } \Re_h = \frac{1}{N_t}\sum_{t=1}^{N_t}\tau_h^{(t)}, \quad (4)$$

where here $\tau_m^{(t)}$ and $\tau_h^{(t)}$ are the time-step rewards at time t for mixing and homogeneity, respectively. Finally, the two reward contributions are combined to $\Re = \alpha\Re_m + (1-\alpha)\Re_h$ using the parameter $\alpha \in [0, 1]$ to tune between weighting more homogeneous or more mixed solutions continuously.

2.3 Reinforcement learning framework

Because the action space is large ($3^{4\times4} \approx 43 \times 10^6$ for combined interactions), conventional control theory techniques are intractable due to Bellman's curse of dimensionality.[50] Similarly, the action space is also too large for value-based RL methods, so that we use a policy-based algorithm. In particular, we use the rllib implementation[51] of Proximal Policy Optimization (PPO).[52]

In the RL agents, we use a neural network, which gets as input the tensor $\mathscr{M}$. The neural network itself then consists of three convolutional layers (with 32, 64, and 256 channels, and kernel sizes 3, 3, and 4, respectively), and a dense layer with $N_g \times N_g$ output neurons. The simulation uses the value of these output neurons to determine if a bin should be activated by thresholding.

The hyperparameters for PPO were tuned automatically during training using ray tune's Population Based Training (PBT)[53] with 16 agents. Using PBT, unstable solutions could be overcome, and we observed converged strategies in all training situations. Additional hyperparameters required for PBT can be found in the ESI.†

3 Results

3.1 Attractive interactions

Experimentally, a controllable active matter system with attractive interactions is already available,[14,15,44] so this case is analyzed first. In Fig. 2a, we show the average episode reward of each of the 16 agents as a function of training time as a solid line, with the minimal and maximal episode reward during the batch providing an error band. It should be noted that while the training resulted in some average increase in reward, the span of rewards is still large. Additionally, we can also observe that agents failing due to some instabilities, as seen by a sharp drop in the reward, can recover due to using PBT. Therefore, the data allows for qualitative analysis, but visual inspection needs to confirm the quality of a strategy¶ in a specific instance. In Fig. 2b, an exemplary time series for a converged strategy is shown, where the emergent strategy collapses all particles into a dense cluster. While this strategy is not a good mixer in the common sense, the solution optimizes eqn (2). In particular, if all particles collapse to a cluster in a single bin, then this bin has a balanced number of "left" and "right", so that for this cell $\mathscr{M}_{l,ij} = \mathscr{M}_{r,ij} = N_p/2$, and the difference is 0. All other cells are empty so that for these cells $\mathscr{M}_{l,ij} = \mathscr{M}_{r,ij} = 0$ again leading to a vanishing difference, and hence $\tau = 0\cdot\tau_h + 1\cdot\tau_m = 0$ is optimal.

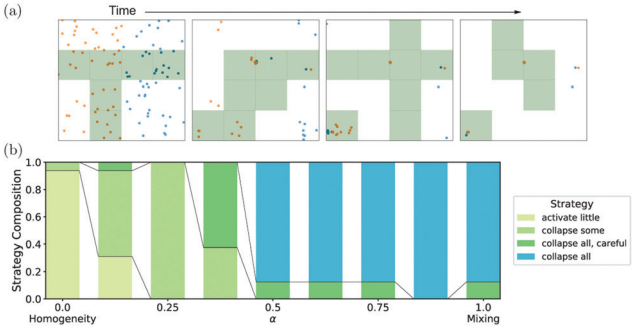


Fig. 2 Training with only attractive interactions achieves only collapsed clusters: In (a), the mean episode reward of the 16 agents training at $\alpha = 1$ is shown in solid lines, with minimal and maximal episode rewards as an error band. It can be observed that while training converged, reward variations were still substantial, and several dips in reward were recovered using PBT. In (b), an exemplary time series for only mixing-focused training ($\alpha = 1$) is shown, in which a collapse of the particle density can be observed. The panels are enumerated from 1 to 4, corresponding to the time-steps 0, 20, 40, and 60, respectively. (A curated list of time series for all attractive-interaction strategies can be found in Fig. S4, ESI†). In (c), the composition of emergent strategies at various values of α is shown, where all the intermediate emergent strategies are taking over continuously. However, no emergent strategy displays a homogeneous mixing.

To overcome the trivial "collapse all"-solution, α can be reduced so that homogeneity is weighted in. For the extreme case of $\alpha = 0$, where the algorithm optimizes for homogeneity only, almost no particles were activated as the initial condition already has a high homogeneity. We called this emerging strategy "activate little", which also did not lead to homogeneous mixing. We analyzed the behavior for intermediate values of α by performing additional simulations at 7 additional values of α (to a total of 9), and we hand-labeled the last validation video for each of the 16 PBT agents. It should be noted, that hand-labeling produced interpretable strategies, but that a quantitative analysis of the activation dynamics (*cf.* Fig. S7 and S8, ESI†) shows that the activation patterns can be roughly classified into groups consistent with our hand-labeling procedure. The results of the emergent strategy composition are plotted as a function of α in Fig. 2c. There we can observe that for intermediate values of α, we observed two additional strategies. Firstly, for some simulations at $0 \leq \alpha < 0.5$, a "collapse some"-strategy emerged, which focuses on collecting dense particle clusters of opposite tags to remedy the worst bins, but leaves most of the other parts of the system inactivated. Secondly, for various values of α, we observed a "collapse all, careful"-strategy similar to the "collapse all"-strategy in that it tries to collect all particles into a single bin. However, it does so over longer periods of time (typically taking more than half of the simulation time) in order to avoid

¶ It should be noted that we define a strategy to summarize all actions of an agent throughout an entire simulation. In contrast, one could consider motifs, *i.e.*, actions an agent performs during a short time period. However, most strategies observed in this paper can be described by a single motif which is used continuously, or at most two that alternate.

This journal is © The Royal Society of Chemistry 2022

Open Access Article. Published on 13 December 2021. Downloaded on 4/8/2026 8:50:09 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

intermediate dense clusters, particularly if these dense clusters are of a single color. As an additional feature of this strategy, due to giving the system longer times to react, the "collapse all, careful"-strategy tends to be more thorough in collapsing all particles in a single bin compared to the "collapse all"-strategy, that mainly focuses on speed. While the exact details, *i.e.*, which strategy emerges for which value of α, will depend on a variety of factors such as the number of bins $N_g \times N_g$, the initial condition, and the subjective assessment during the hand-labeling (*cf.* ESI† Section SC), we can observe two main results: firstly, the strategies continuously transform from one to another by tuning α. Secondly, none of the observed strategies achieves a homogeneous mixed state.

Therefore, using only attractive interactions, no strategy emerged that provides a solution similar to our target, as depicted in Fig. 1c, even when varying α. Indeed, we will show in Section 4 that no reward-shaping could overcome this problem, as attractive interactions are insufficient to arrive at a homogeneous and mixed state.

3.2 Repulsive interactions

While not yet implemented in the experimental platform, a repulsive–interactions system could be engineered by introducing another motor protein type. Using repulsive interactions only, a more homogeneous mixing strategy exists than in the attractive-only case: a typical time series for the mixing focused $\alpha = 1$ case is shown in Fig. 3a, where we observe that the emergent strategy is the "activate one side"-strategy. For $N_g = 4$ and with periodic boundary conditions, every bin is a boundary bin between the two sides of the initial configuration. As such, every particle expelled from one side enters the other without contracting to a cluster. Two features of this mixing solution should be highlighted: first, this strategy again maximizes $\Re_m$ by reducing the homogeneity of the system. Second, mixing is only facilitated by exploiting the boundary conditions.

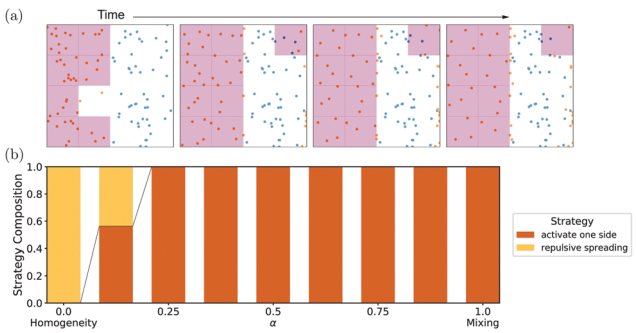


Fig. 3 Training with repulsive interactions has a dominant strategy using the periodic boundary conditions to achieve mixing: in (a), an exemplary time series for only mixing-focused training ($\alpha = 1$) is shown, in which activation of one side of the system can be observed, using the periodic boundary conditions to facilitate mixing. (A curated list of time series of all repulsive-interaction strategies can be found in Fig. S5, ESI†). In (b), the different emergent strategies are shown, demonstrating that the "activate one side"-strategy is dominant over most of the α-parameter space.

For $\alpha = 0$, *i.e.*, maximizing the homogeneity of the system, the network uses repulsive interactions in more densely populated bins to achieve homogenization beyond the initial condition ("repulsive spreading"-strategy). However, as becomes clear from Fig. 3b, this spreading strategy only emerges very close to $\alpha = 0$ and is quickly overtaken by the "activate one side" strategy (*cf.* Fig. S7 and S8, ESI†). Indeed, the homogeneity reward strongly punishes dense clusters as emergent in the "collapse all"-strategy but only weakly punishes a situation where only half of the cells become empty. Hence, the "activate one side"-strategy is dominant over large parts of the α-parameter-space. While the "activate one side"-strategy achieves some mixing with the repulsive-only interactions, the strategy is not tunable, and it heavily relies on the periodic boundary conditions. Indeed, we will demonstrate in Section 4 that phase-space-limiting boundary conditions (such as periodic boundary conditions) are required to facilitate mixing with repulsive interactions only.

3.3 Combined interactions

In contrast to the previous cases, using both interaction types, we will achieve a tunable mixing strategy that can achieve both homogeneity and mixing without exploiting the boundary conditions. Starting with the $\alpha = 1$ case, we observe the "collapse all"-strategy emerging once more. However, already for $\alpha = 1$, the network sometimes uses repulsive interactions to spread dense particle clusters, only to attract them again. Because this cycle typically occurs multiple times, we call these strategies oscillatory, and they emerge in different flavors. For large α, the emergent strategy still collapses particles to a dense cluster (see Fig. 4a) in a strategy we term "oscillation w/collapse". For smaller α, the strategies still use oscillations but avoid very dense clusters in a strategy we call "oscillations w/o collapse", depicted in Fig. 4b. Finally, for $\alpha = 0$, the RL algorithm finds an "attractive-repulsive spreading"-strategy which homogenizes the system using both interaction types to spread dense cells and collapse in less populated cells.

Overall, an interesting strategy evolution as a function of α emerges using both interactions, as shown in Fig. 4c. In particular, a smooth transition from one strategy to the next can be observed (maybe with exception to the "attractive-repulsive spreading"-strategy), and various intermediate strategies produce well-mixed states without relying on boundary conditions. It should be noted that the actual emergent strategies depend on various factors and are to a degree subjective due to the hand-labeling procedure. Indeed, both oscillatory strategies and the "collapse all"-strategy are similar (*cf.* the more quantitative analysis presented in Fig. S7 and S8, ESI†), and their differences are can mainly be attributed to when a spread of a dense cluster occurs. As such, this classification relies on the hand-labeling procedure to correlate the particle configuration with the activation pattern. However, we can observe some general features: namely, the strategy evolution is mostly continuous, and the existence of a homogeneous mixing strategy seemed to depend mainly on the interaction set. This result is surprising as attractive or repulsive interactions alone are insufficient to produce efficient mixing, but

Open Access Article. Published on 13 December 2021. Downloaded on 4/8/2026 8:50:09 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

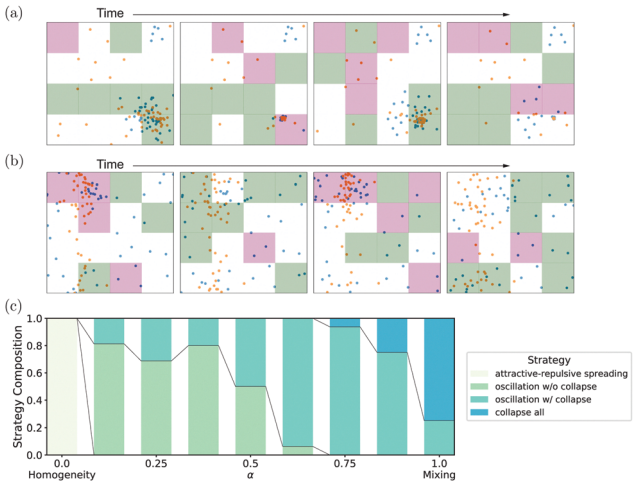


Fig. 4 Training on a system with both interactions achieves homogeneous mixing by alternating attractive and repulsive interactions: in (a and b), two exemplary time series for oscillation strategies are shown, with and without collapse, respectively. (A curated list of time series for all combined-interaction strategies can be found in Fig. S6, ESI†). In (c), the fraction of runs with a particular strategy at various values of α is shown, with various emergent strategies continuously transforming into each other.

the combination of both is sufficient. We can now analyze the success-failure pattern of RL to find efficient strategies using dynamical systems theory to analyze the problem more closely. Doing so will reveal that the necessity of combined interactions is not a specific feature of the model in eqn (1) but a much more general statement about different models.

4 Theoretical analysis

In the numerical results presented in Section 3, we observed that only the combination of attractive and repulsive interactions led to homogeneously mixed states and that tweaking the reward function did not find a suitable solution for restricted interaction sets. While these results are limited in their applicability to the model described by eqn (1), here, we will demonstrate how these results can be generalized using dynamical system theory and ergodic theory. Indeed, mixing is studied mathematically as a subdiscipline of ergodic theory, where (weak) mixing is a stronger form of ergodicity, *i.e.*, ergodicity is a necessary but not sufficient condition for mixing (see textbooks[54–56]). In particular, we will use that if a map is an Anosov diffeomorphism, it will induce (weak) mixing,[34,57] to show that homogeneous mixing in drag-dominated translational-invariant particle systems can only occur in systems with attractive and repulsive interactions.

We start by reconsidering the model in eqn (1). Without truncating the sums, the equations of motion are

$$\vec{x}_i(t+\Delta t) = \vec{x}_i(t) - \frac{kp_i\Delta t^2}{m} \sum_{\substack{i\neq j \\ p_i=1=p_j}} \frac{|\vec{d}_{ij}(t)| - r_0}{|\vec{d}_{ij}(t)|}\,\vec{d}_{ij}(t), \qquad (5)$$

where $r_0 = 0$ in the case of attractive interactions, and $r_0 = R_c$ for repulsive interactions. Here, the difference vector is abbreviated as $\vec{d}_{ij}(t) = \vec{x}_i(t) - \vec{x}_j(t)$. Hence, the updates without periodic boundary conditions can be written as a matrix-vector multiplication: $\vec{X}(t+\Delta t) = M\vec{X}(t)$, where $\vec{x} \in \mathbb{R}^{2N_p}$ is the collection of all particle positions. It should be noted that M depends on which particles are activated (hence making it time-dependent), and on $\vec{X}$ (making the update non-linear). However, both dependencies are small since $M[\vec{X}, t] = 1 + \Delta t^2 \tilde{M}[\vec{X}, t]$, where $\Delta t \ll 1$, and $\tilde{M}$ describes the particle–particle interactions.

The long-term evolution of the system can be analyzed by considering the eigenvalues λ_i of M at every time-step. Namely, ergodicity (and therefore mixing) requires the update to be measure-preserving in phase space,[56] meaning that the determinant $\det M = \prod_i \lambda_i$ has to be 1. Suppose the update map M is sufficiently (namely C^2) differentiable, has eigenvalues smaller and larger than 1, but maintains a determinant of 1. Then, the map is so-called Anosov, is guaranteed to mix the phase space (*i.e.*, the $2 \times N_p$-dimensional space),[34,57] and as such it is expected to mix the projection in the 2-dimensional real space.

Therefore, if we understand the spectrum of M (*i.e.*, the set of all eigenvalues), strong statements about mixing can be achieved. The update matrix M can be expressed as

$$M = \begin{pmatrix} 1-\sum_j c_{1j} & c_{12} & \cdots & c_{1N_a} \\ c_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & c_{N_a-1,N_a} \\ c_{N_a1} & \cdots & c_{N_a,N_a-1} & 1-\sum_j c_{N_aj}, \end{pmatrix}, \qquad (6a)$$

where

$$c_{ij} = \frac{k\Delta t^2}{m}\left[\underbrace{p_i^a p_j^a \frac{|\vec{d}_{ij}(t)| - 0}{|\vec{d}_{ij}(t)|}}_{=1>0} + \underbrace{p_i^r p_j^r \frac{|\vec{d}_{ij}(t)| - R_c}{|\vec{d}_{ij}(t)|}}_{<0}\right]\chi\left[r_c < |\vec{d}_{ij}| < R_c\right], \qquad (6b)$$

for $i \neq j$, with the characteristic function $\chi\left[r_c < |\vec{d}_{ij}| < R_c\right] = 1$ if the condition in the brackets is met and 0 otherwise, ensuring that only particles in the interaction range interact. While eqn (6b) is specific to the force model described by eqn (1), it should be noted that the form of M shown in eqn (6a) is quite general. Namely, M having a row sum 1 and small off-diagonal elements will generally arise for particle-based drag-dominated translational-invariant systems.[35] Indeed, only the general form of M shown in eqn (6a) is required to derive the necessary results about the spectrum, and hence all the following results are valid for a large class of models.

This journal is © The Royal Society of Chemistry 2022

Open Access Article. Published on 13 December 2021. Downloaded on 4/8/2026 8:50:09 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Because M is symmetric, all eigenvalues are real.‖ Because of Gershgorin's circle theorem[58], the spectrum of M

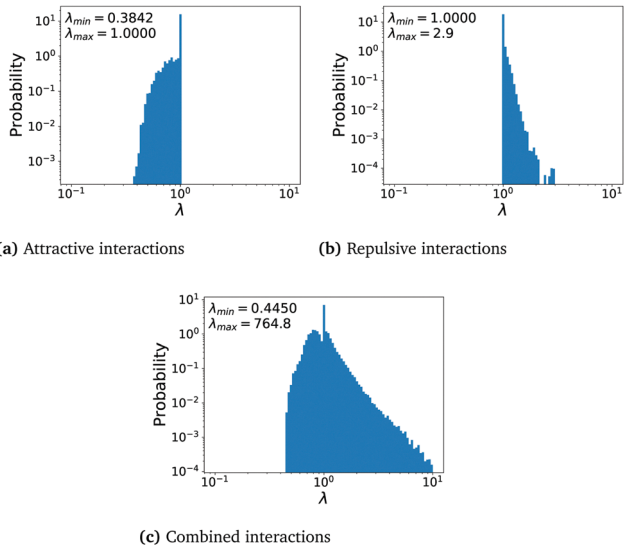
$$\lambda(M) \subset \bigcup_i \left[1 - \sum_{j\neq i} c_{ij} - \sum_{j\neq i} |c_{ij}|,\ 1 - \sum_{j\neq i} c_{ij} + \sum_{j\neq i} |c_{ij}|\right]. \quad (7)$$

Using eqn (7) for attractive interactions where $c_{ij} \geq 0$: $\lambda(M) \subset \left[1 - 2\max_i \sum_{j\neq i} c_{ij}, 1\right]$ so that the largest possible eigenvalue is 1, while all other eigenvalues are less than 1. Therefore, $\det M \leq 1$ and iterating M on any initial vector will either collapse the phase space direction (associated eigenvalue $\lambda < 1$) or leave it invariant (associated eigenvalue $\lambda = 1$). Hence, with only attractive interactions, the system will only be able to collapse, never leading to homogeneous mixing.

Analogously, for repulsive interactions, where $c_{ij} \leq 0$, $\lambda(M) \subset \left[1, 1 + 2\max_i \sum_{j\neq i} |c_{ij}|\right]$ so that all eigenvalues are larger or equal to 1, and $\det M \geq 1$. Therefore, without phase-space-limiting boundary conditions, the phase space is expanding, never leading to homogeneous mixing. However, with periodic boundary conditions, phase space is limited and can be folded into itself. Thus, the successful mixing with only repulsive interactions presented in Section 3.2 was only possible due to the periodic boundary conditions.

Finally, for attractive and repulsive interactions, the off-diagonal elements are around 0 but can have either sign** so that the eigenvalues of M lie around 1 (more precisely $\lambda(M) \subset \left[1 - 2\max_i \sum_{j\neq i} |c_{ij}|, 1 + 2\max_i \sum_{j\neq i} |c_{ij}|\right]$). While it is not guaranteed that the determinant will be 1, the updates now can induce hyperbolic dynamics, *i.e.*, phase space stretching in some directions and compressing in others, bringing the dynamics closer to an Anosov diffeomorphism, hence being able to induce mixing in agreement with our successful mixing simulations in Section 3.3. Furthermore, our results indicate that in order to optimize mixing, the determinant should be close to 1. This optimal mixing could be achieved in future work by tuning α dynamically during training by monitoring the average determinant.

The results on the eigenvalue spectra can be verified numerically: additional simulations at $\alpha = \frac{1}{2}$ were performed for all three interaction cases. We analyzed the update matrices of the last validation runs for each of the agents at every time point of the simulation and plotted their eigenvalue frequency as histograms in Fig. 5. Indeed, we observe the bounded spectra for limited interaction sets while the system with combined interactions exhibits a more balanced eigenvalue spectrum. While the spectrum for combined interactions in Fig. 5c is not entirely balanced around 1, and the determinant is larger than 1, the matrix is closer to describe an Anosov diffeomorphism and can describe one if necessary. In particular, the oscillatory strategies discussed in Section 3.3 alternate attractive and repulsive interactions to achieve eigenvalues larger and smaller than 1 over multiple time-steps (*cf.* Fig. S9, ESI†). In conclusion, we indeed observe the predicted eigenvalue spectra throughout all simulations, clearly demonstrating why limited interaction sets cannot provide mixing, whereas the combined interaction set can.



(a) Attractive interactions (b) Repulsive interactions

(c) Combined interactions

Fig. 5 Histogram of the eigenvalue spectra of M for the case of only attractive, only repulsive, and combined interaction sets in (a) to (c), respectively. It can be observed that the spectra in (a) and (b) are bounded, whereas the spectrum for combined interactions does not exhibit such a limitation.

5 Conclusion

This paper analyzed a challenging control problem arising for active matter components, where control cannot be exerted over individual agents or particles but merely over their pairwise interaction. In particular, we attempted to find efficient microscopic mixing strategies for active particles using policy-based RL, where we observed that RL only finds efficient mixing strategies for models with combined attractive and repulsive interactions but not for those with only one type of interaction. This peculiar success/failure pattern was then analyzed using dynamical systems theory, particularly theory to hyperbolic dynamics, on the update matrix M to prove that both types of interactions are required to induce mixing in arbitrary drag-dominated translational-invariant particle models. Finally, the mathematical results on M were confirmed in additional training runs, demonstrating why RL could only solve the problem with both interaction types.

The results of this paper suggest further work in two areas: first, our results could be transferred to the experimental system[14,15,43] to verify that mixing of active particles can induce

‖ For non-symmetric matrices, the argument would be analogous, limiting the discussion to the absolute value of the eigenvalue.

** Per row, each off-diagonal will have the same sign, as only the same activation states interact with each other. However, due to the possibility of changing the interaction state, the same row can have an opposite sign at different times.

Open Access Article. Published on 13 December 2021. Downloaded on 4/8/2026 8:50:09 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

mixing in the surrounding medium *via* fluid–matter interactions. While the experimental platform is not yet ready, since it lacks an additional repulsive light-controllable motor protein, our paper provides a clear and feasible guideline for necessary developments to the platform to enable active–matter-based mixing. As such, we hope to provide the necessary motivation for the field to develop a dual-controlled active–matter system.

Second, the strategies we found should be further refined for actual applications. While some computational effort could be saved by tuning α dynamically to match a unit determinant, future simulations should use a higher grid resolution, a larger number of particles, better boundary conditions, and more realistic particle–particle interactions to produce realistic strategies. While such refined simulations are highly interesting, their computational demand will undoubtedly be more extensive, so that we expect more advanced machine learning approaches such as transfer learning or curriculum learning to be necessary. Nevertheless, we are confident that our work paves the way for future applications of RL to this area of study, namely control of complex interaction-dominated active matter systems.

Author contributions

DS and MT conceived the project. DS, ANP, and JS wrote the simulation environment. DS performed the simulations and analyzed the results. DS and MT wrote the manuscript with input from all authors.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank Arjuna Subramanian, Jerry Wang, Pranav Bhamidipati, Ivan Jimenez, Jeremy Bernstein, Dr Guannan Qu, Dr Christopher Miles, and Dr Shahriar Shadkhoo for scientific discussions and feedback on the manuscript. We thank Inna-Marie Strazhnik for help with the figures. We acknowledge funding through the Foundational Questions Institute and Fetzer Franklin Fund through FQXi 1816, the Packard Foundation (2019-69662), and the Heritage medical research institute. ANP acknowledges additional funding through the SFP SURF program.

References

1 D. Schildknecht and M. Thomson, *New J. Phys.*, 2021, **23**, 083001.
2 E. L. Paul, V. Atiemo-Obeng and S. Kresta, *Handbook of Industrial Mixing*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2003.
3 P. M. Valencia, O. C. Farokhzad, R. Karnik and R. Langer, *Nat. Nanotechnol.*, 2012, **7**, 623–629.
4 P. Neužil, S. Giselbrecht, K. Länge, T. J. Huang and A. Manz, *Nat. Rev. Drug Discovery*, 2012, **11**, 620–632.
5 A. D. Stroock, S. K. W. Dertinger, A. Ajdari, I. Mezic, H. Stone and G. M. Whitesides, *Science*, 2002, **295**, 647–651.
6 H. Y. Gan, Y. C. Lam, N. T. Nguyen, K. C. Tam and C. Yang, *Microfluid. Nanofluid.*, 2006, **3**, 101–108.
7 J. S. Kuo and D. T. Chiu, *Annu. Rev. Anal. Chem.*, 2011, **4**, 275–296.
8 J. Ortega-Casanova, *J. Fluids Struct.*, 2016, **65**, 1–20.
9 A. Groisman and V. Steinberg, *Nature*, 2001, **410**, 905–908.
10 D. Saintillan and M. J. Shelley, *Phys. Rev. Lett.*, 2008, **100**, 1–4.
11 D. Saintillan and M. J. Shelley, *Phys. Fluids*, 2008, **20**, 123304.
12 M. Guo, A. J. Ehrlicher, M. H. Jensen, M. Renz, J. R. Moore, R. D. Goldman, J. Lippincott-Schwartz, F. C. Mackintosh and D. A. Weitz, *Cell*, 2014, **158**, 822–832.
13 D. Marenduzzo and E. Orlandini, *Soft Matter*, 2010, **6**, 774.
14 Z. Qu, D. Schildknecht, S. Shadkhoo, E. Amaya, J. Jiang, H. J. Lee, D. Larios, F. Yang, R. Phillips and M. Thomson, *Commun. Phys.*, 2021, **4**, 198.
15 Z. Qu, J. Jiang, H. J. Lee, R. Phillips, S. Shadkhoo and M. Thomson, 2021, arXiv: 2101.08464v1, pp. 1–8.
16 J. Stenhammar, R. Wittkowski, D. Marenduzzo and M. E. Cates, *Sci. Adv.*, 2016, **2**, e1501850.
17 A. K. Omar, Y. Wu, Z.-G. Wang and J. F. Brady, *ACS Nano*, 2019, **13**, 560–572.
18 F. Cichos, K. Gustavsson, B. Mehlig and G. Volpe, *Nat. Mach. Intell.*, 2020, **2**, 94–103.
19 G. Reddy, A. Celani, T. J. Sejnowski and M. Vergassola, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, E4877–E4884.
20 S. Colabrese, K. Gustavsson, A. Celani and L. Biferale, *Phys. Rev. Lett.*, 2017, **118**, 158004.
21 S. Muiños-Landin, A. Fischer, V. Holubec and F. Cichos, *Sci. Robot.*, 2021, **6**, eabd9285.
22 G. Reddy, J. Wong-Ng, A. Celani, T. J. Sejnowski and M. Vergassola, *Nature*, 2018, **562**, 236–239.
23 S. Verma, G. Novati and P. Koumoutsakos, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 5849–5854.
24 L. Biferale, F. Bonaccorso, M. Buzzicotti, P. C. D. Leoni and K. Gustavsson, *Chaos*, 2019, **29**, 103138.
25 E. Schneider and H. Stark, *EPL*, 2019, **127**, 64003.
26 C.-H. Yu and R. Nagpal, Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI), 2010.
27 M. Rubenstein, A. Cornejo and R. Nagpal, *Science*, 2014, **345**, 795–799.
28 Y. Yang, R. Luo, M. Li, M. Zhou, W. Zhang and J. Wang, 35th International Conference on Machine Learning, ICML 2018, 2018, pp. 8869-8886.
29 M. Durve, F. Peruani and A. Celani, *Phys. Rev. E*, 2020, **102**, 012601.
30 K. Zhang, Z. Yang and T. Baar, in *Handbook of Reinforcement Learning and Control*, ed. K. G. Vamvoudakis, Y. Wan, F. L. Lewis and D. Cansever, Springer International Publishing, Cham, 2021, pp. 321–384.
31 S. Chennakesavalu and G. M. Rotskoff, *J. Chem. Phys.*, 2021, **155**, 194114.
32 M. M. Norton, P. Grover, M. F. Hagan and S. Fraden, *Phys. Rev. Lett.*, 2020, **125**, 178005.

 This journal is © The Royal Society of Chemistry 2022

Open Access Article. Published on 13 December 2021. Downloaded on 4/8/2026 8:50:09 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

33 M. J. Falk, V. Alizadehyazdi, H. Jaeger and A. Murugan, 2021, arXiv: 2105.04641, pp. 1–11.
34 D. V. Anosov, *Proc. Steklov Inst. Math.*, 1967, **90**, 235.
35 P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter*, Cambridge University Press, Cambridge, 1995, ch. 7.
36 P. G. de Gennes and J. Prost, *The Physics of Liquid Crystals*, Clarendon Press, 1993.
37 J. F. Joanny, F. Jülicher, K. Kruse and J. Prost, *New J. Phys.*, 2007, **9**, 422.
38 L. Giomi, M. C. Marchetti and T. B. Liverpool, *Phys. Rev. Lett.*, 2008, **101**, 198101.
39 L. Giomi, *Phys. Rev. X*, 2015, **5**, 031003.
40 P. J. Foster, S. Fürthauer, M. J. Shelley and D. J. Needleman, *eLife*, 2015, **4**, 1–21.
41 H. Berendsen, D. van der Spoel and R. van Drunen, *Comput. Phys. Commun.*, 1995, **91**, 43–56.
42 S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1–19.
43 F. Nédélec and D. Foethke, *New J. Phys.*, 2007, **9**, 427.
44 T. D. Ross, H. J. Lee, Z. Qu, R. A. Banks, R. Phillips and M. Thomson, *Nature*, 2019, **572**, 224–229.
45 G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang and W. Zaremba, *arXiv*, 2016, 1–4.
46 C. R. Doering and J. L. Thiffeault, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2006, **74**, 1–4.
47 J.-L. Thiffeault, *Nonlinearity*, 2012, **25**, R1–R44.
48 C. J. Miles and C. R. Doering, *Nonlinearity*, 2018, **31**, 2346–2359.
49 C. J. Miles and C. R. Doering, *J. Nonlinear Sci.*, 2018, **28**, 2153–2186.
50 R. Bellman, *Dynamic Programming*, Princeton University Press, 1957.
51 E. Liang, R. Liaw, P. Moritz, R. Nishihara, R. Fox, K. Goldberg, J. E. Gonzalez, M. I. Jordan and I. Stoica, 35th International Conference on Machine Learning, ICML 2018, 2018, pp. 3053–3062.
52 J. Schulman, F. Wolski, P. Dhariwal, A. Radford and O. Klimov, *arXiv*, 2017, 1–12.
53 R. Liaw, E. Liang, R. Nishihara, P. Moritz, J. E. Gonzalez and I. Stoica, *arXiv*, 2018, https://arxiv.org/abs/1807.05118.
54 P. Walters, *An Introduction to Ergodic Theory*, Springer-Verlag, New York, 1982.
55 D. Kerr and H. Li, *Ergodic Theory. Independence and Dichotomies*, Springer International Publishing, 2017.
56 J. Hawkins, *Ergodic Dynamics*, Springer International Publishing, Cham, 2021, vol. 289.
57 M. Brin and G. Stuck, *Introduction to Dynamical Systems*, Cambridge University Press, 2010, ch. 6, pp. 141–152.
58 S. Gershgorin, *Bull. Acad. Sci. URSS, Cl. Sci. Phys.-Math.*, 1931, 749–754.