



Cite this: DOI: 10.1039/d5sm00655d

# Learning general pair interactions between self-propelled particles

Jérôme Hem, <sup>ab</sup> Alexis Poncet, <sup>c</sup> Pierre Ronceray, <sup>d</sup> Daiki Nishiguchi <sup>ef</sup> and Vincent Démery <sup>\*ac</sup>

Synthetic active matter systems, such as active colloids, often have complex interactions, which can be of hydrodynamic, chemical or electrostatic origin and cannot be computed from first principles. Here, we use stochastic force inference to learn general pair interactions, including transverse forces and torques, between self-propelled Janus particles from experimental trajectories. We use data from two experiments: one where the particles flock, and one where the system remains disordered. The learned interactions are then fed to numerical simulations, which reproduce all the experimental observables and could be extrapolated to different densities. Overall, we find that the radial interaction is mostly repulsive and isotropic, while the angular interaction has a richer angular dependence, which controls the behavior of the system; the transverse interaction is negligible. Finally, testing the symmetry relations obeyed by the inferred interactions allows us to show that they cannot come from electrostatics only, so that they must have a hydrodynamic component.

Received 25th June 2025,  
Accepted 16th August 2025

DOI: 10.1039/d5sm00655d

[rsc.li/soft-matter-journal](https://rsc.li/soft-matter-journal)

## 1 Introduction

The field of active matter has emerged as a framework to describe living systems as diverse as bacterial colonies,<sup>1,2</sup> schools of animals<sup>3–5</sup> or human crowds.<sup>6–8</sup> To help the development of new models, it was soon realized that synthetic active matter systems were needed to perform better controlled experiments.<sup>9</sup> Focusing on self-propelled polar particles, experimental realizations have already used vibrated disks,<sup>10,11</sup> swimming droplets,<sup>12–14</sup> as well as rolling and suspended colloids.<sup>15,16</sup> In particular, Janus particles propelled by a vertical electric field through induced-charge electrophoresis<sup>17–19</sup> are very attractive model systems, with rich collective behaviors that depend on tunable pair interactions.<sup>20–26</sup> However, despite being well controlled, computing the interactions between the particles in these systems is very challenging. For example, in the case of Janus particles propelled by an electric field, hydrodynamic interactions were sometimes neglected and only electrostatic interactions were taken into account,<sup>20,22,26</sup> but these can only be computed under several approximations, among which is the hypothesis that the

two colloids are far from any boundary. Considering that the particles are actually in the vicinity of an electrode, this is a rather crude assumption. Similarly, the hydrodynamic flow around a particle can be computed from the Stokes equation, assuming that the particle is far from any other particle and far from the electrodes.<sup>19</sup> If the interactions cannot be computed but their knowledge is essential to understand the behavior of the system, they need to be measured in experiments. This approach is routinely used for the self-propulsion speed  $U$ , which is always measured in experiments in the dilute regime but never computed from first principles. Although the methods to extract one particle parameters such as  $U$  are usually quite straightforward, a mean value obtained from a distribution, they cannot be readily applied to measure the interactions because of the many supplementary degrees of freedom. Instead, one needs to develop more sophisticated methods that have the ability to measure the interactions in the presence of multiple source of noises, which are always present in experiments on colloidal systems.

To measure the interactions from experimental trajectories, many approaches can be followed. One, which gained in popularity over the last years, is neural networks based machine learning (NN) techniques, where the interactions are encoded in a neural network.<sup>27</sup> Already several attempts have been made for colloidal systems: the radial pair interaction between colloids has been measured in ref. 28, whereas the coarse-grained hydrodynamical model has been inferred directly in ref. 29 and 30. One can also find model-free NN techniques whose goal is to forecast the dynamics. For example the authors in ref. 31 successfully predicted the dynamics of

<sup>a</sup> Gulliver, UMR CNRS 7083, ESPCI Paris, Université PSL, 75005 Paris, France

<sup>b</sup> Institut Jean Lamour, CNRS, Université de Lorraine, 54011 Nancy, France

<sup>c</sup> CNRS, ENS de Lyon, LPENSL, UMR5672, 69342, Lyon cedex 07, France

<sup>d</sup> Aix Marseille Univ, CNRS, CINaM, Turing Center for Living Systems, Marseille, France

<sup>e</sup> Department of Physics, Institute of Science Tokyo, Ookayama 2-12-1, Tokyo 152-8551, Japan

<sup>f</sup> Department of Physics, The University of Tokyo, Hongo 7-3-1, Tokyo 113-0033, Japan

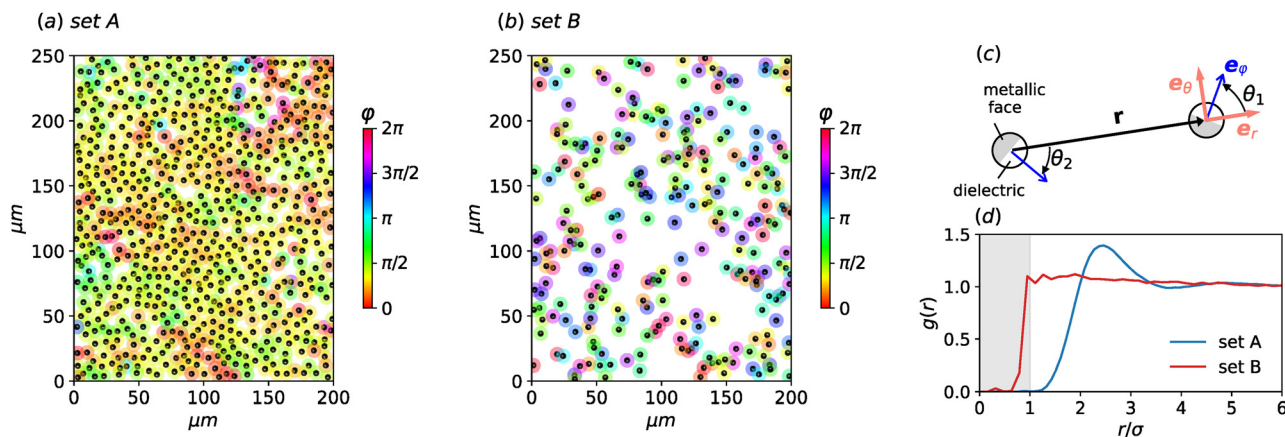
complex particle assemblies up to few seconds. Besides, it is worth mentioning the existence of other promising frameworks, based on NN techniques or not, which focus on the inference of specific dynamical quantities, such as memory kernels for the generalized Langevin equation<sup>32</sup> or specific potentials.<sup>33,34</sup>

In this article, we choose another approach called stochastic force inference (SFI).<sup>35</sup> This framework relies on an explicit and robust denoising process, based on Ito–Stratonovich conversion, which is designed to disentangle the deterministic force and the intrinsic and measurement noises. It is then followed by a standard fitting procedure where the displacements and the diffusion are decomposed on a basis. This robustness against different source of noises makes it specifically relevant for the study of experimental colloidal systems. To date, in the field of soft matter, this approach has been used only to infer the underdamped dynamic of a single particle<sup>36</sup> although, in principle, it can be applied to large assemblies. For example, ref. 35 demonstrates the inference of isotropic radial pair interactions between simulated active polar particles.

Here, we use SFI to determine the most general pair interactions between self-propelled Janus particles from the experimental trajectories. These particles have metallic and dielectric hemispheres, and the difference in their electric properties generates horizontal self-propulsion when an AC vertical electric field is applied. By tuning the frequency of the electric field and the ion concentrations of the suspending fluid, we can fine-control the self-propulsion direction and the interactions among the particles, which can realize various collective behaviors.<sup>20,25</sup> Typically, at low frequencies  $\lesssim 30$  kHz, the Janus particles propel with their dielectric side forward due to induced-charge electrophoresis (ICEP). In contrast, for high-frequencies  $\gtrsim 30$  kHz, the particles move with their metal side forward due to self-dielectrophoresis (sDEP). We consider two different experiments.<sup>23,24</sup> The first one, denoted set A in what follows, contains a rather high density of particles (area fraction  $\phi_A = 0.097$ ) at high frequency, which form a flock with true long

range polar order and giant number fluctuations<sup>24</sup> (Fig. 1(a)). In this set, the high density of particles makes the inference challenging, as it requires to disentangle the forces from the different sources of noise and between many particles. In contrast, the second dataset, denoted set B, contains a lower density of particles ( $\phi_B = 0.04$ ), at low frequency, with the addition of ions to realize weaker interactions, and the system remains in a disordered state<sup>23</sup> (Fig. 1(b)). This is evidenced either in a movie, where many collisions between the particles can be observed, or in the radial pair correlation function  $g(r)$ , where particles are observed in close contact (Fig. 1(d)). Here, the inference is challenging because of the weak interactions that take place at short distances between the particles.

This article is structured as follows. In Section 2, we define the model and the interactions to consider for the self-propelled particles. Then, we recall the core mechanism of SFI and the functional basis we used to decompose the forces. In Sections 3 and 4, we run SFI and discuss the inferred pair interactions, respectively for the flocking and the disordered experiments. In both cases, the quality of the inference is assessed by running numerical simulations with the inferred interactions, so as to reproduce quantitatively the experimental observables. Specific to the flocking system, we show in addition that it is possible to extrapolate the dynamics at different densities, which could be useful for example to predict the transition between the disordered and the ordered states. Finally, in Section 5, we discuss the possible origins of the interactions. The basis used to describe the interactions is chosen to satisfy general symmetry relations. If the interactions were purely electrostatic, they would derive from a potential and several additional symmetry relations would be satisfied. In the flocking experiment, we show that the first order harmonic terms of the interactions can be explained by a dominant electrostatic interaction but that the higher order terms cannot, which proves that another contribution, probably of hydrodynamic origin, should be present, supporting the flocking mechanisms discussed in ref. 24. For the disordered experiment,



**Fig. 1** (a) and (b) Snapshots of size  $200 \times 250 \mu\text{m}^2$  for the two sets A and B. The colored rings surrounding the particles indicate the value of the polar angle  $\varphi \in [0, 2\pi]$ . The direction  $\varphi = 0$  is along the horizontal axis. Note that for set A, the particles move with the metallic (dark) face ahead whereas for set B, they move with the dielectric (white) face ahead. (c) Notations for the relative distance and orientations between particles 1 and 2 moving with the white face ahead; the blue arrows indicate the direction of the self-propulsion. (d) Radial pair correlation function  $g(r)$  for the two sets A and B. The gray area indicates the range  $r/\sigma < 1$  with the particle diameter  $\sigma = 3.17 \mu\text{m}$  in both cases.

the lack of symmetry relation at any order suggests instead a strong hydrodynamic component compared to the electrostatic one.

## 2 Method

### 2.1 Model

Modeling and determining general interactions between the particles is challenging, mainly because of the many degrees of freedom. However, the range of possibilities can be drastically reduced if we assume the particles identical and consider only pairwise interactions. Considering particles moving in two dimensions, these assumptions reduce the effect of the particle 2 on the particle 1 to the presence of an additional velocity  $\mathbf{v}$  and an angular velocity  $\omega$ . These velocities enter in the overdamped equations of motion for the position  $\mathbf{r}_i(t)$  and orientation  $\varphi_i(t)$  of the particle  $i$  as:

$$\dot{\mathbf{r}}_i(t) = U\mathbf{e}_{\varphi_i(t)} + \sum_{j \neq i} \mathbf{v}(\mathbf{r}_i - \mathbf{r}_j, \varphi_i, \varphi_j) + \sqrt{2D}\eta_i(t), \quad (1)$$

$$\dot{\varphi}_i(t) = \sum_{j \neq i} \omega(\mathbf{r}_i - \mathbf{r}_j, \varphi_i, \varphi_j) + \sqrt{2D_r}\xi_i(t). \quad (2)$$

In eqn (1), the first term represents the self-propulsion,  $U$  being the self-propulsion speed and  $\mathbf{e}_{\varphi}$  the unit vector pointing in the direction  $\varphi$  (Fig. 1(c)), the second term represents the translational effect of the interactions, and the third term represents the translational noise, with diffusion coefficient  $D$ ,  $(\eta_i(t))_{i=1, \dots, N}$  being Gaussian white noises with unit variance. In eqn (2), the first term represents the rotational effect of the interactions and the second term represents the rotational noise, with diffusion coefficient  $D_r$ ,  $(\xi_i(t))_{i=1, \dots, N}$  being Gaussian white noises with unit variance. Note that the term “velocity”, which is used for  $U$ ,  $\mathbf{v}$  and  $\omega$ , corresponds to the “drift” in the framework of stochastic differential equations. The interaction velocity  $\mathbf{v}$  and angular velocity  $\omega$  are the products of the translational and angular mobilities with the interaction force and torque; the velocities are the natural outcome of the analysis of the trajectories and we refer to them as the “interaction velocities” or as the “interactions”. Note that we assume that the diffusion coefficients do not depend on the presence of neighboring particles; this is a common assumption and assessing its validity is beyond the scope of this work.

Further assuming the rotational invariance of the system, these quantities depend on the distance  $r = |\mathbf{r}|$  between the particles, with  $\mathbf{r} = \mathbf{r}_i - \mathbf{r}_j$ , and on their orientations  $\theta_i$  and  $\theta_j$  with respect to  $\mathbf{r}$  (Fig. 1(c)). Last, the interaction velocity  $\mathbf{v}$  can be decomposed in a component along  $\mathbf{r}$ ,  $v_r$ , and a component perpendicular to it,  $v_{\theta}$ :

$$\mathbf{v}(r, \theta_1, \theta_2) = v_r(r, \theta_1, \theta_2)\mathbf{e}_r + v_{\theta}(r, \theta_1, \theta_2)\mathbf{e}_{\theta}; \quad (3)$$

the unit vectors  $\mathbf{e}_r$  and  $\mathbf{e}_{\theta}$  are defined in Fig. 1(c). Altogether, the interactions are encoded in three scalar functions of three scalar variables,  $v_r(r, \theta_1, \theta_2)$ ,  $v_{\theta}(r, \theta_1, \theta_2)$  and  $\omega(r, \theta_1, \theta_2)$ .

### 2.2 Stochastic force inference

The stochastic inference formalism of ref. 35 and 37 proposes several estimators to reconstruct the drift velocities as well as the diffusion part of eqn (1) and (2). In particular, ref. 35 provides a proof of concept of model inference in the case of self-propelled particles with isotropic interactions, which we expand here to anisotropic interactions and apply to real data. We now briefly recapitulate the main features of SFI in this context.

The input of SFI is twofold: on the one hand, the set of trajectories, *i.e.* of state variables  $\mathbf{x}_i(t) = (\mathbf{r}_i(t), \varphi_i(t))$  of all particles  $i$ , recorded at successive times  $t_0, t_1, t_2, \dots$  with total step number  $N_t$  and time interval  $\delta t = t_{n+1} - t_n$  between frames. We denote here  $\mathbf{X}(t) = \{\mathbf{x}_i(t)\}_{i=1 \dots N}$  the state of the full system at time  $t$ . On the other hand, it requires a set of basis functions<sup>†</sup>  $\{\mathbf{b}_{\alpha}(\mathbf{X})\}_{\alpha=1 \dots N_b}$  to fit the drift with, which we split here into single-particle terms (the self-propulsion) and two-particle terms (the interactions). We discuss the choice of these functions in Section 2.3. The main output of SFI are fit coefficients  $\hat{F}_{\alpha}$  for the drift field, which can then be reconstructed as  $\hat{\mathbf{F}}(\mathbf{X}) = \sum_{\alpha} \hat{F}_{\alpha} \mathbf{b}_{\alpha}(\mathbf{X})$ .

SFI provides several estimators for the coefficients  $\hat{F}_{\alpha}$ , which require the evaluation of a stochastic integral. We have used here the Stratonovich estimator, whose robustness against measurement errors yielded the best results on our data. Specifically, it consists in:

$$\hat{F}_{\alpha} = \sum_{\beta=1}^{N_b} (G^{-1})_{\alpha\beta} (v_{\beta} + w_{\beta}) \quad (4)$$

where we have used the Gram matrix  $G_{\alpha\beta} = N_t^{-1} \sum_{t=1}^{N_t} \mathbf{b}_{\alpha}(\mathbf{X}(t)) \cdot \mathbf{b}_{\beta}(\mathbf{X}(t))$ , the nonequilibrium velocity fit coefficient  $v_{\alpha} = (2N_t\delta t)^{-1} \sum_{t=1}^{N_t} [\mathbf{X}(t+\delta t) - \mathbf{X}(t)] [\mathbf{b}_{\alpha}(\mathbf{X}(t)) + \mathbf{b}_{\alpha}(\mathbf{X}(t+\delta t))]$  (which is odd under time reversal and thus captures irreversibility) and the “equilibrium” term  $w_{\alpha} = N_t^{-1} \sum_{t=1}^{N_t} \hat{\mathbf{D}}(t) \cdot \nabla \mathbf{b}_{\alpha}(\mathbf{X}(t))$  (which is even under time reversal). Here  $\hat{\mathbf{D}}(t) = [\delta \mathbf{X}(t)^2 + \delta \mathbf{X}(t + \delta t)^2 + 4\delta \mathbf{X}(t)\delta \mathbf{X}(t + \delta t)]/(4\delta t)$ , with  $\delta \mathbf{X}(t) = \mathbf{X}(t + \delta t) - \mathbf{X}(t)$ , is the noise-corrected instantaneous diffusion estimator introduced by Vestergaard *et al.*<sup>38</sup> In practice, the diffusion on the particle positions is too small to be measured with the experimental temporal resolution as the displacement due to diffusion between two frames,  $\sqrt{D\delta t}$ , is much smaller than the displacement due to self-propulsion,  $U\delta t$ . As a consequence, we set  $D = 0$  and retain only a constant rotational diffusion coefficient  $D_r$ . This greatly reduces the algorithmic complexity, leading to a scaling as  $\mathcal{O}(N_t N_b)$  fit function evaluations, each of which costs  $\mathcal{O}(N^2)$ . Assuming that the interactions and the fit functions vanish for an interparticle distance  $r > r_{\max}$ , we use a cell-list

<sup>†</sup> Note that we adopt here the formalism of ref. 37 where basis functions are vectors and fitting coefficients are scalar, which allows more control over symmetries compared to ref. 35.

type algorithm, as in many-particle simulations, to reduce this last cost to  $\mathcal{O}(N)$ .

Finally, SFI provides guidance regarding the amount and quality of data needed to perform well. Specifically, the normalized mean squared error due to sampling noise is self-consistently estimated to be  $\approx 2N_b/I$ , with the information being  $I = N_t^{-1} \sum_{t=1}^{N_t} \hat{\mathbf{F}}(\mathbf{X}(t)) \cdot \mathbf{D}^{-1} \cdot \hat{\mathbf{F}}(\mathbf{X}(t)) \delta t$ , which is expected to scale linearly with the system size. As a consequence, the larger the number of fit functions  $N_b$ , the more data is needed to achieve a given error level. This estimator allows to control for overfitting. Moreover, some relative positions of pairs of particles can be unexplored, which reduces the accuracy of the inference in these regions. With the Stratonovich estimator discussed above, the error induced by measurement error is a blur of the drift field of similar magnitude – (e.g. an angular error of  $10^\circ$  over orientations will roughly result in a Gaussian blurring of angle-dependent forces of similar width). This is to be compared to much more dramatic biases scaling as  $\mathcal{O}(1/\delta t)$  induced by more naive estimators.<sup>35</sup> Lastly, biases induced by time discretization are of the order of the typical drift variation over  $\delta t$ —that is, if drifts typically vary by 10% over a time step, the bias on inferred forces will be of  $\sim 10\%$ . For this reason, the typical distance travelled between two frames should remain small with respect to the diameter of the particles  $\sigma$ :  $U\delta t/\sigma \lesssim 0.1$ . We note that this can be significantly improved by turning to an alternative trapezoid estimator<sup>37,39</sup> at the cost of robustness to measurement error.

### 2.3 Fit functions

We now define the basis on which to decompose the particle velocities. For one particle forces, the choice is straightforward. As a Janus particle is self-propelled along its orientation  $\varphi$ , the most natural fit function is the unitary vector  $\mathbf{e}_\varphi = (\cos(\varphi), \sin(\varphi))$ , which provides the self-propulsion speed  $U$ . In contrast, the choice for the pair interactions is still broad. However, it is possible to further reduce the basis by setting additional constraints based on the symmetry of the problem.

Assuming that the system is invariant under rotations and parity inversion, the interactions should be symmetric with respect to the axis defined by the particles 1 and 2:

$$v_r(r, \theta_1, \theta_2) = v_r(r, -\theta_1, -\theta_2), \quad (5)$$

$$v_\theta(r, \theta_1, \theta_2) = -v_\theta(r, -\theta_1, -\theta_2), \quad (6)$$

$$\omega(r, \theta_1, \theta_2) = -\omega(r, -\theta_1, -\theta_2). \quad (7)$$

We choose a basis that satisfies these relations and separates the angular and radial dependencies:

$$v_r(r, \theta_1, \theta_2) = \sum_{l=0}^{N_r-1} g_l(r) \sum_{n=0}^{N_a} \sum_{k=-n+1}^n a_{n,k,l}^r \cos([n - |k|]\theta_1 + k\theta_2), \quad (8)$$

$$v_\theta(r, \theta_1, \theta_2) = \sum_{l=0}^{N_r-1} g_l(r) \sum_{n=1}^{N_a} \sum_{k=-n+1}^n a_{n,k,l}^\theta \sin([n - |k|]\theta_1 + k\theta_2), \quad (9)$$

$$\omega(r, \theta_1, \theta_2) = \sum_{l=0}^{N_r-1} g_l(r) \sum_{n=1}^{N_a} \sum_{k=-n+1}^n a_{n,k,l}^\omega \sin([n - |k|]\theta_1 + k\theta_2). \quad (10)$$

We have chosen a Fourier basis of order  $N_a$  for the angular dependence: the angular functions are of the form  $\cos(k_1\theta_1 + k_2\theta_2)$ , with  $|k_1| + |k_2| \leq N_a$  (cos is replaced by sin for  $v_\theta$  and  $\omega$ ). Finally, we use smooth localized functions for the radial dependence:

$$g_l(r) = \frac{1}{l!} \left( \frac{r}{r_0} \right)^l e^{-r/r_0}. \quad (11)$$

The function  $g_l(r)$  has a maximum at  $lr_0$  and a width  $r_0$ ; as it decays exponentially, it can be assumed to vanish beyond a distance  $r_{\max}$ . Thus, the number of basis functions is  $N_b = N_r \times [1 + 3N_a(N_a + 1)]$ . We emphasize that this decomposition is general and does not make any assumption on the form of the interactions, beyond the symmetry relations that they satisfy. From the coefficients  $a_{n,k,l}$ , we also define the derived quantity  $a_{n,k}(r) = \sum_l a_{n,k,l} g_l(r)$ , the amplitude of the harmonic function indexed by  $n, k$  as a function of the distance  $r$ .

Note that, although the coefficients  $a_{n,k,l}$ , or equivalently  $a_{n,k}(r)$ , represent the amplitude of harmonic terms contributing to the inferred velocity, their value may change with the order of the angular basis  $N_a$ . This comes from the fitting procedure that adjusts the coefficients in order to minimize the error of the overall functional basis. If the chosen order  $N_a$  is large enough to capture the dynamics, the underlying coefficients  $a_{n,k}(r)$  could give, indeed, information on the harmonic composition of the velocities. On the contrary, their physical meaning as well as that of the inferred velocity are compromised if the dynamics is not captured properly. It is therefore important to assess the quality of the inference before discussing the meaning of the coefficients  $a_{n,k,l}$  or  $a_{n,k}(r)$ .

## 3 Interactions in the flocking system

### 3.1 Results

We turn to the experiments of ref. 24 and especially the dataset A, for which the ordered phase is observed. For this, we run SFI with the basis functions defined above at second order ( $N_a = 2$ ), which provides first the inferred self-propulsion speed  $U \simeq 9.7 \mu\text{m s}^{-1}$  and the rotational diffusion  $D_r \simeq 0.092 \text{ s}^{-1}$  (details given in the Appendix). These values are found in good agreement with those measured in ref. 24.

Then, we assess the reliability of the inference by analyzing the statistical distribution of the particles with respect to each other. Although our approach provides the interactions for all distances and orientations, the system does not explore uniformly all the configuration space, so that the interactions may

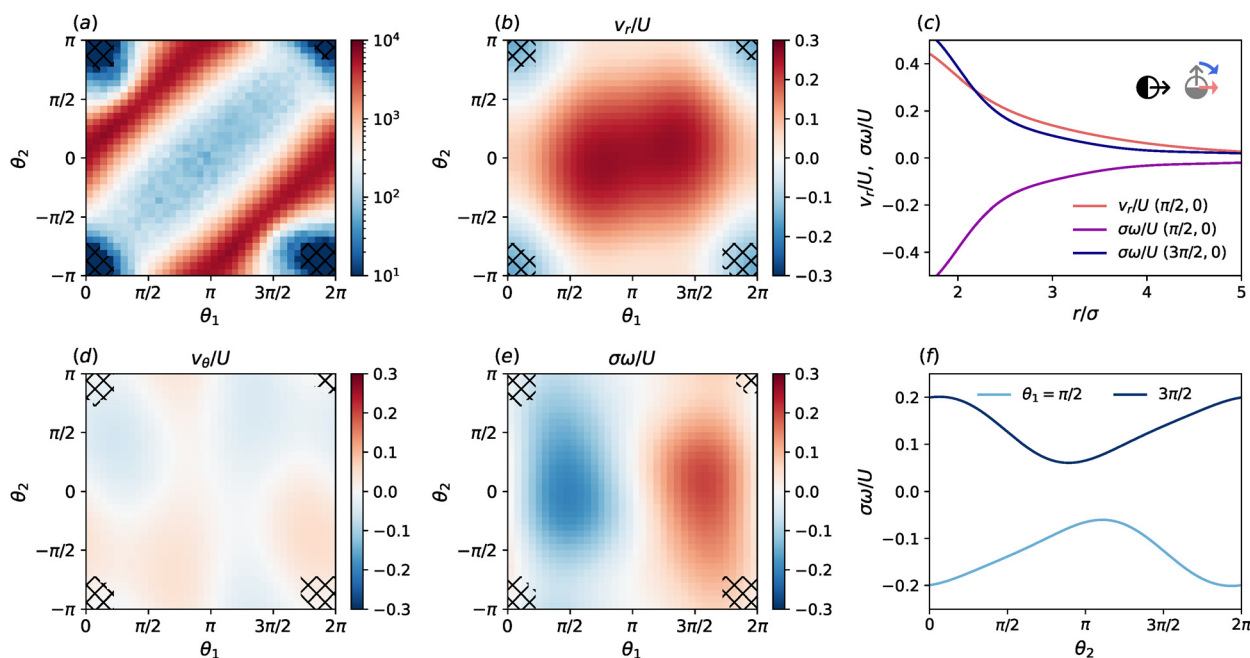


not be inferred properly in unexplored regions. Indeed, the radial pair correlation (Fig. 1(d)) indicates the absence of neighbor particles at short distance  $r/\sigma < 2$ , with the particle diameter  $\sigma = 3.17 \mu\text{m}$ , an accumulation around  $r/\sigma = 2.5$  and a uniform presence beyond  $r/\sigma > 3$ . Since the distance  $r/\sigma = 2.5$  is akin to a mean distance in the flock, we choose it and discuss the corresponding histogram of the relative orientations in the plane  $(\theta_1, \theta_2)$  (Fig. 2(a)). Similar to the distance  $r$ , the distribution of neighbors is also not uniform. There is a region of high density, defined by  $\theta_1 \simeq \theta_2$ , which corresponds to particles moving in the same direction, revealing the flocking behavior. There is also a region of intermediate density and an unexplored one, located around  $\theta_1 \simeq 0, \theta_2 \simeq \pi$ , corresponding to the situation where the two particles are back to back. The presence or even the excess of particles consolidate the inference in these regions. On the contrary, the inference in the unexplored region should be taken with care, although it can be qualitatively correct, for example in term of sign. For clarity, we hatch the questionable area if the bin count in the histogram falls below the arbitrary threshold of 10 occurrences in a pixel.

The inferred interactions  $v_r, v_\theta$ , and  $\omega$  exerted by a particle 2 on a neighbor particle 1 at a distance  $r/\sigma = 2.5$  are presented in Fig. 2(b), (d) and (e). The interactions and distances are rescaled by  $U$  and  $\sigma$  to allow the comparison of the interaction components between each other and with the self-propulsion. Note that in the plane  $(\theta_1, \theta_2)$ , a neighbor particle 1 is located in front of the particle 2 for  $-\pi/2 < \theta_2 < \pi/2$  and behind it for  $-\pi < \theta_2 < -\pi/2$  or  $\pi/2 < \theta_2 < \pi$ .

The radial interaction  $v_r$  is mainly repulsive for a large range of orientations Fig. 2(b), so that the interaction prevents self-contact. In particular, there are two maxima located at  $(\theta_1, \theta_2) \simeq (\pi/2, 0)$  and  $(\theta_1, \theta_2) \simeq (3\pi/2, 0)$ , that is, when particle 1 is located in front of the particle 2 and oriented sideways to it. Although the statistics suggest caution, it is probable that the interaction when particles are back to back is negligible or weakly negative. The azimuthal interaction  $v_\theta$  has a non trivial shape but with an amplitude much smaller than the radial or the angular interactions (Fig. 2(d)).

In contrast, the amplitude of the torque (Fig. 2(e)) is, in dimensionless units, as strong as the radial interaction, but with a more complex dependence on the orientations. It is negative for  $0 \leq \theta_1 \leq \pi$  and positive for  $\pi \leq \theta_1 \leq 2\pi$ , which means that the particle 1 always turns away from the particle 2. Such “turning away” interaction, in addition to an isotropic radial repulsion, has been identified as being responsible for the flocking in ref. 26. In this work, Das *et al.* retained the lowest order term in a multipole expansion of the electrostatic interaction between Janus particles, which has the slowest decay, leaving  $\omega(r, \theta_1, \theta_2) \propto -\sin(\theta_1)/r^4$ . Here, we find that the interaction is strongly modulated by the orientation of the particle 2: it is maximal when the particle 2 faces the particle 1 ( $\theta_2 \simeq 0$ ) and almost vanishes when the particle 2 turns its back on particle 1 ( $\theta_2 \simeq \pi$ ). This modulation could appear at higher order in the multipole expansion of ref. 26. Another representation is proposed in Fig. 2(f), where we see that the amplitude of the torque can significantly change with  $\theta_2$  for fixed  $\theta_1$ . Due to its strong magnitude, we hypothesize that it



**Fig. 2** SFI results for set A (flocking system). (a) Pair correlation: histogram of the relative orientations at a distance  $r/\sigma = 2.5$ . (b), (d) and (e) Inferred pair interactions at a distance  $r/\sigma = 2.5$  as a function of the orientations  $\theta_1$  and  $\theta_2$ : (b) radial, (d) azimuthal and (e) angular velocities. The hatched areas correspond to areas of the histogram with lower statistics (bin count below 10). (c) Radial and angular velocities as function of the distance  $r$  for specific orientations:  $(\theta_1, \theta_2) = (\pi/2, 0)$ ,  $(3\pi/2, 0)$ . Inset: Particle 1 (grey) and particle 2 (black) with the orientations  $(\theta_1, \theta_2) = (\pi/2, 0)$ , with the red and blue arrows indicating the directions of the radial and angular velocities, respectively. (f) Angular velocity at a distance  $r/\sigma = 2.5$  as a function of the orientation  $\theta_2$ , for two values of the orientation  $\theta_1$ .

could play an important role in the flocking. We come back to this question in the next section.

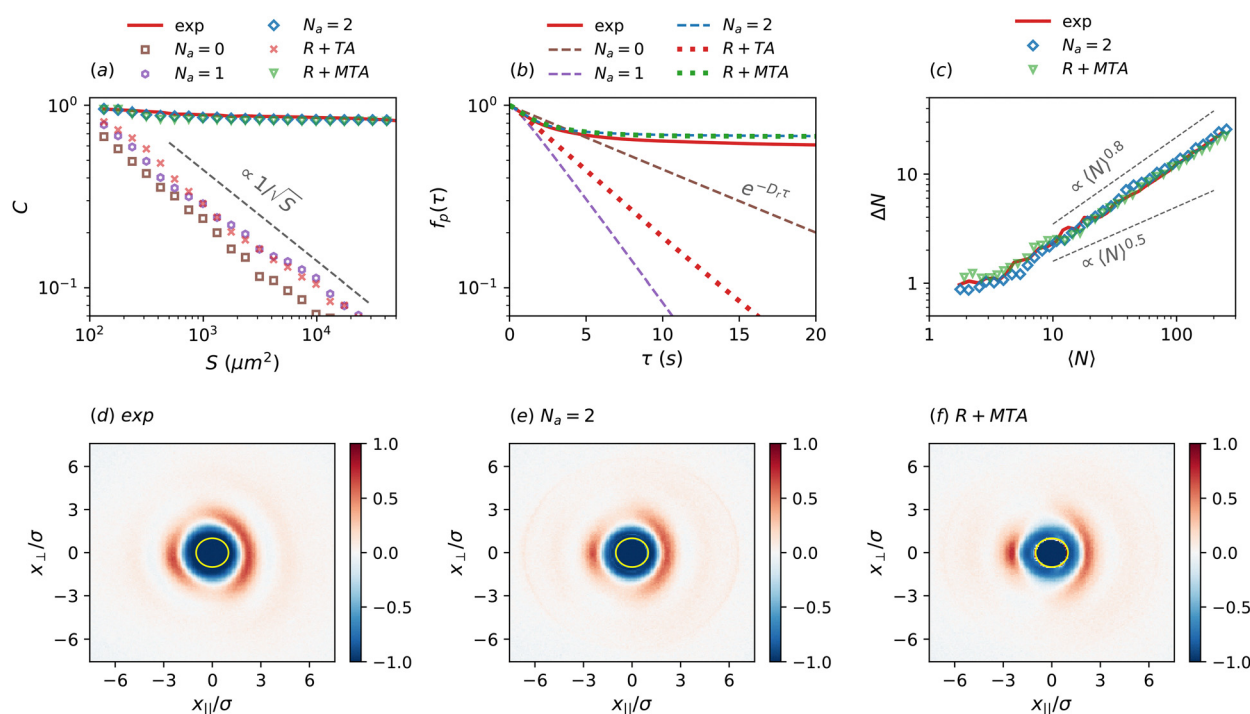
Finally, Fig. 2(c) presents the evolution of the radial and angular interactions as a function of the distance  $r$ , for orientations  $(\theta_1, \theta_2)$  that maximize the interactions. We observe that, in both cases, the interactions decrease monotonically and vanish above  $r/\sigma \gtrsim 5$ . These behaviors of the interactions are robust features supported by sufficient statistics.

### 3.2 Numerical simulations

To further assess the reliability the inferred interactions, we run numerical simulations with these interactions and compare the observables with those measured in experiment (details given in the Appendix). The results are presented in Fig. 3 for different orders  $N_a = 0, 1, 2$  of the angular basis. Overall, we find that the synthetic ordered phase, obtained at the second order, presents the same statistical properties than the experimental one, and that this order is also the minimal one required to describe faithfully the dynamics.

First, we consider the polar order parameter  $C = \langle |\langle e^{i\varphi} \rangle_S| \rangle_t$ , where the inner average runs over the particles contained in rectangular regions of area  $S$  (regions of interest, ROI), averaged over temporally uncorrelated frames (outer average). The dependence on the area  $S$  can generally be fitted with  $C(S) = C_\infty + kS^{-\gamma/2}$ ,<sup>24,40,41</sup> a finite value of  $C_\infty$  indicating the existence

of a long range order. The order parameter, for both the experiment and the simulation, levels off towards the same finite value  $C_\infty \simeq 0.8$  for  $S > 10^4 \mu\text{m}^2$  (Fig. 3(a)). Then, we look at the polarity autocorrelation function  $f_p(\tau) = \langle \mathbf{e}_\varphi(t) \cdot \mathbf{e}_\varphi(t + \tau) \rangle_t$  where  $\mathbf{e}_\varphi(t)$  stands for the polarity of the particle at time  $t$ . This quantity decreases exponentially in a disordered system but decays to a finite plateau in a flock.<sup>24</sup> Again, a plateau is observed in both cases, indicating a high level of correlation maintained even after a long period  $\tau > 10$  s (Fig. 3(b)). Another interesting quantity is the number fluctuations,  $\Delta N$ , which is defined as the standard deviation of the number of particles in a ROI of a given size.<sup>24,42</sup> Compared to  $\langle N \rangle$ , the average number of particles in a ROI of the same size, it describes the density fluctuations in the system. We find in the numerical simulations the same behavior as in the experiments,  $\Delta N \sim \langle N \rangle^{0.8}$  (Fig. 3(c)), indicating giant density fluctuations, which are the hallmark of an ordered Toner-Tu phase.<sup>25,43</sup> Finally, the pair correlation  $B(\mathbf{r})$  integrated over the angle of particle 2, has the same form in the experiment and the simulation (Fig. 3(d) and (e)).  $B(\mathbf{r})$  indicates the variation of the density of particles at a position  $\mathbf{r}$  with respect to a particle moving in the direction  $\mathbf{e}_\varphi$ .<sup>23</sup> It is constructed from the polar pair correlation  $C(\mathbf{r}, \theta, \theta')$ ,<sup>23,44</sup> which from rotational invariance is contained in  $C(\mathbf{r}, 0, \theta')$ :  $B(\mathbf{r}) = (1/2\pi) \int_0^{2\pi} C(\mathbf{r}, 0, \theta') d\theta'$ . In particular, note in both cases the accumulation of particles around  $r/\sigma = 2.5$  (the red ring in Fig. 3(d) and (e)).



**Fig. 3** Comparison of the observables between: the experimental flocking system (set A), the simulations with the inferred interactions at different angular orders  $N_a = 0, 1, 2$  and the simulations denoted R + TA and R + MTA including only certain terms at  $N_a = 2$ . In all simulations, the frame size and the particle number are identical to the experiment. (a) Polar order parameter  $C$  vs. the area  $S$  of the ROI. (b) Polarity autocorrelation  $f_p(\tau)$ . (c) Number fluctuation  $\Delta N$  vs. average number of particle  $N$ . (d)–(f) Pair correlation function  $B(\mathbf{r})$ , with  $x_{||}$  and  $x_{\perp}$  the coordinates in the reference frame of an active particle pointing to the right, respectively for (d) the experiment, (e) the simulation at  $N_a = 2$  and (f) the simulation R + MTA. The size of the particle is represented by the yellow circle in the center. See main text for the definition of the interactions in the simulations R + TA, R + MTA and the definition of the observables.

In comparison, the interactions inferred at zero or first order lead to incorrect predictions. The particles do not flock and the system remains disordered. This is evidenced in Fig. 3(a) where the polar order function  $C(S)$  decreases as  $\propto 1/\sqrt{S}$ , as expected for randomly orientated particles, which is characteristic of uncorrelated movements, or in Fig. 3(b) where the polarity function  $f_p(\tau)$  decreases as  $e^{-D'\tau}$  with the effective diffusion constant  $D'$ . Note that, at zeroth order, the constant  $D'$  is identical to the rotation diffusion  $D_r$ , whereas at first order  $D'$  increases, so the diffusion is enhanced in this case. The pair correlation function  $B(r)$  at zeroth or first order is also very different compared to the one observed in the experiment (not shown).

Besides, we find instructive to search for the dominant terms that can explain the main features of the experimental dynamics. From the inferred interactions at second order, Fig. 2(b), (d) and (e), it directly appears that some terms can be neglected, such as the azimuthal interaction  $v_\theta$  due to its low amplitude. For the radial and angular interactions, the choice of the important terms is less obvious. However, we can identify them by turning on and off certain terms of the interaction in the simulations and see the effect on the dynamics. First, we choose to keep only the isotropic radial repulsion  $a_{0,0}^r$  and the  $\theta_2$ -independent “turning away” interaction  $a_{1,0}^\omega$  (denoted R + TA):

$$v_r(r, \theta_1, \theta_2) = a_{0,0}^r(r), \quad (12)$$

$$\omega(r, \theta_1, \theta_2) = a_{1,0}^\omega(r)\sin(\theta_1), \quad (13)$$

with  $v_\theta = 0$ . Thus, the angular dependence of the interaction is identical to the one studied in ref. 26. We see that the system remains disordered (Fig. 3(a) and (b)); note that it does not contradict the conclusions of ref. 26, as flocking could be recovered at higher density. Second, in contrast, we choose to keep both the radial repulsion  $a_{0,0}^r$  and the full form of the angular interaction, which thus includes the modulation as a function of  $\theta_2$  (denoted R + MTA, for “modulated turning away” in Fig. 3):

$$v_r(r, \theta_1, \theta_2) = a_{0,0}^r(r), \quad (14)$$

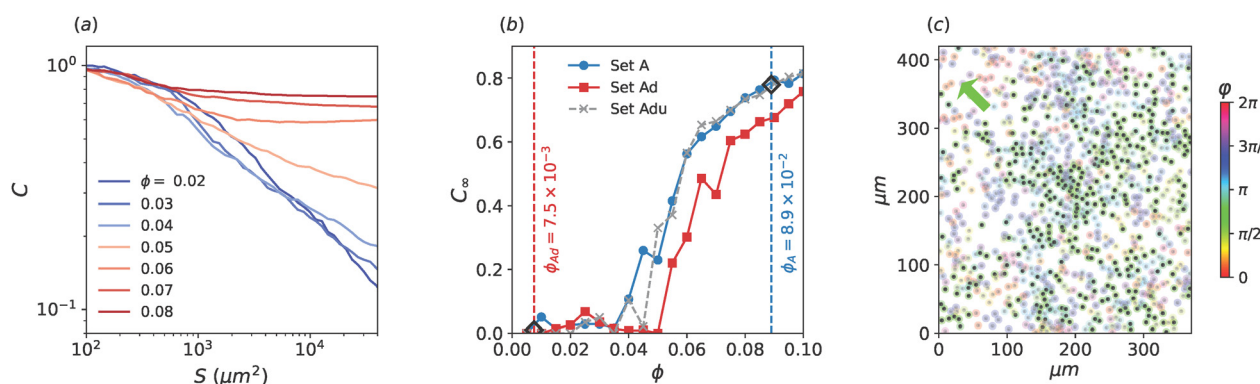
$$\omega(r, \theta_1, \theta_2) = \sum_{n=1}^2 \sum_{k=-n+1}^n a_{n,k}^\omega(r) \sin([n - |k|]\theta_1 + k\theta_2), \quad (15)$$

with  $v_\theta = 0$ . With these interactions, the flocking is recovered (Fig. 3(a) and (b)), as well as the various observables of the experimental system (Fig. 3). From the radically different behavior observed in numerical simulations with the R + TA and R + MTA interactions, we conclude that the full form of the angular interaction, in addition to an isotropic radial repulsion, is important and contributes to the flocking behavior of the experimental system.

### 3.3 Extrapolation to different densities

Another advantage of inferring the pair interactions between the particles is to be able to predict a behavioural change in the system as a function of parameters such as the particle density. To illustrate this possibility, we run numerical simulations with the inferred parameters obtained previously ( $U$ ,  $D_r$  and the pair interactions at second order for set A), and then vary the area fraction from a low density  $\phi = 5 \times 10^{-3}$  to a high density  $\phi = 0.1$ . The behavioural change is quantified by the evolution of the polar order function  $C(S)$  and the parameter  $C_\infty$ , extracted from the relation  $C(S) = C_\infty + kS^{-\gamma/2}$  (Fig. 4). We observe two regimes. For small area fractions,  $\phi < 0.04$ , the system is disordered,  $C \propto S^{-1/2}$  and  $C_\infty \simeq 0$ . In contrast, for higher values  $\phi > 0.06$ , we retrieve the flocking behavior observed in the experimental set A (Fig. 1(c))  $C$  converging to  $C_\infty \simeq 0.8$  for  $S > 10^3 \mu\text{m}^2$ . At the threshold  $0.04 < \phi < 0.06$ , the two phases coexist. For example, Fig. 4(c) shows the emergence of a swarm of particles moving along the same direction for  $\phi = 0.05$ .

The existence of the two dynamical regimes can be verified experimentally: ref. 24 reports another dataset at lower area fraction  $\phi_{Ad} = 7.5 \times 10^{-3}$ , which we denote set Ad, where the system remains disordered. Since only the density is varied between the two datasets, one expects to find identical self-propulsion speeds, rotational diffusion coefficients, and pair interactions in both cases. However, SFI finds an increase of the self-propulsion speed ( $U_{Ad} = 14.1 \mu\text{m s}^{-1}$ ) compared to the



**Fig. 4** Simulations with the inferred parameters, with varying area fraction  $\phi$ . (a) Evolution of the function  $C(S)$  with  $\phi$  on a log–log scale, with the inferred parameters extracted from set A. (b) Evolution of the parameter  $C_\infty$  with  $\phi$ , with the inferred parameters of set A, Ad and Adu. The value  $C_\infty$  fits the relation:  $C = C_\infty + kS^{-\gamma/2}$ . The parameters Adu are the same as for Ad except the value  $U$  equal to that of set A. The vertical dashed lines indicate the area fractions  $\phi_{Ad}$ ,  $\phi_{Ad}$  of the experimental sets A and Ad, the black diamonds indicate the related experimental values  $C_\infty$ . (c) Snapshot of the data set taken close to the transition ( $\phi = 5 \times 10^{-2}$ ), with the inferred parameters of set A.

ordered phase ( $U_A = 9.7 \mu\text{m s}^{-1}$ ); this difference could be attributed to a variation in some experimental parameters, such as the distance between the electrodes, or to a “screening effect”, particles in a flock feeling a weaker field than lonely particles. We run SFI for the set Ad with the same parameters as before, and run simulations with the inferred interactions for different area fractions  $\phi$ . We see in Fig. 4(b) that there is almost no difference between sets A and Ad; in particular the two dynamical regimes are retrieved. The shift between the two curves is due to the change in  $U$ : if we set  $U = U_A$  in the simulations with the interactions inferred from the set Ad, then the two curves become identical (compare sets A and Adu). Thus, the pair interactions can be considered as identical between the two sets with respect to their flocking propensity. The inference, applied to a reduced set of particles or applied to a single dataset at given area fraction, can thus help to predict the behavior at any other density and to reveal possible phase transitions, which could be useful to guide experiments.

## 4 Interactions in the disordered system

### 4.1 Results

We turn now to set B provided by ref. 23, where the system remains in a disordered state. One goal of this study was to provide a pure experimental active Brownian particles (ABP) system, that is with only radial pair interactions ( $v_r(r, \theta_1, \theta_2) = v_r(r)$ ,  $v_\theta = 0$ ,  $\omega = 0$ ),<sup>45</sup> in order to compare with the theoretical predictions; in absence of torque, ABP do not align, although their activity may trigger a motility-induced phase separation.<sup>46</sup> To achieve this goal, the experimental parameters (electrical field, concentration of sodium chloride) were tuned so as to decrease the overall hydrodynamic and electrostatic interactions. Indeed, we see the particles moving in a disordered way with no polar order measured, and, thanks to the weak interactions, the particles collide with one another. The dynamics looks qualitatively very similar to the one expected for ABP. However, if one takes a finer observable, such as the pair correlation  $B(\mathbf{r})$ , the experimental one shows the formation of a depletion zone just behind the particle which is absent for ABP (Fig. 6(a) and (c)). It is thus possible that residual interactions are still present in the system, which could be characterized.

In contrast to the first dataset, the detection of the force is more challenging, because the interactions are much weaker and take place at smaller distances between particles, within  $1 < r/\sigma < 3$ . At such distance, the inference may be blurred by the collisions, especially if the time step is a bit long, as is the case here where  $U\delta t/\sigma \simeq 0.2$ . We also note that the error on the orientation detection is higher ( $\varphi_{\text{err}} \approx 9^\circ$ ), which also decreases the measurement precision. Despite these difficulties, we shall see that it is possible to obtain the interactions in this system.

Once again, we run SFI with the basis functions defined at second order ( $N_a = 2$ ). The inference provides the self-propulsion speed  $U \simeq 7.0 \mu\text{m s}^{-1}$  and the rotational diffusion

$D_r \simeq 0.085 \text{ s}^{-1}$ , which are similar to those reported in ref. 23. Then, we choose the distance  $r/\sigma = 1.5$  and discuss the histogram of the relative orientations presented in Fig. 5(a). Since the system is disordered, the distribution of neighbors is quite homogeneous along  $\theta_1$ ,  $\theta_2$  (Fig. 5(a)). The greater densities are observed around  $(\theta_1, \theta_2) = (\pi/2, -\pi/2)$  or  $(3\pi/2, \pi/2)$ . Here, the unexplored region where the particles are back to back appears larger than compared to set A, because of the closer distance  $r/\sigma = 1.5$ . Again, the questionable areas are hatched and reported on top of the inferred interactions.

The inferred interactions  $v_r$ ,  $v_\theta$ ,  $\omega$  at distance  $r/\sigma = 1.5$  are presented in Fig. 5(b), (d) and (e). First, the radial interaction Fig. 5(b) looks similar to the previous experiment. It is still positive when the two particles face each other, and is maximally repulsive for two angular positions:  $(\theta_1, \theta_2) \simeq (2\pi/3, -\pi/3)$  or  $(4\pi/3, \pi/3)$ , that is, when the particles fly by each other. In contrast, the interaction is weakly attractive when the particles are back to back  $(\theta_1, \theta_2) \simeq (0, \pi)$ , although this last region should be taken with care because of the poor statistics. As in the previous experiment, the azimuthal interaction is very weak (Fig. 5(d)).

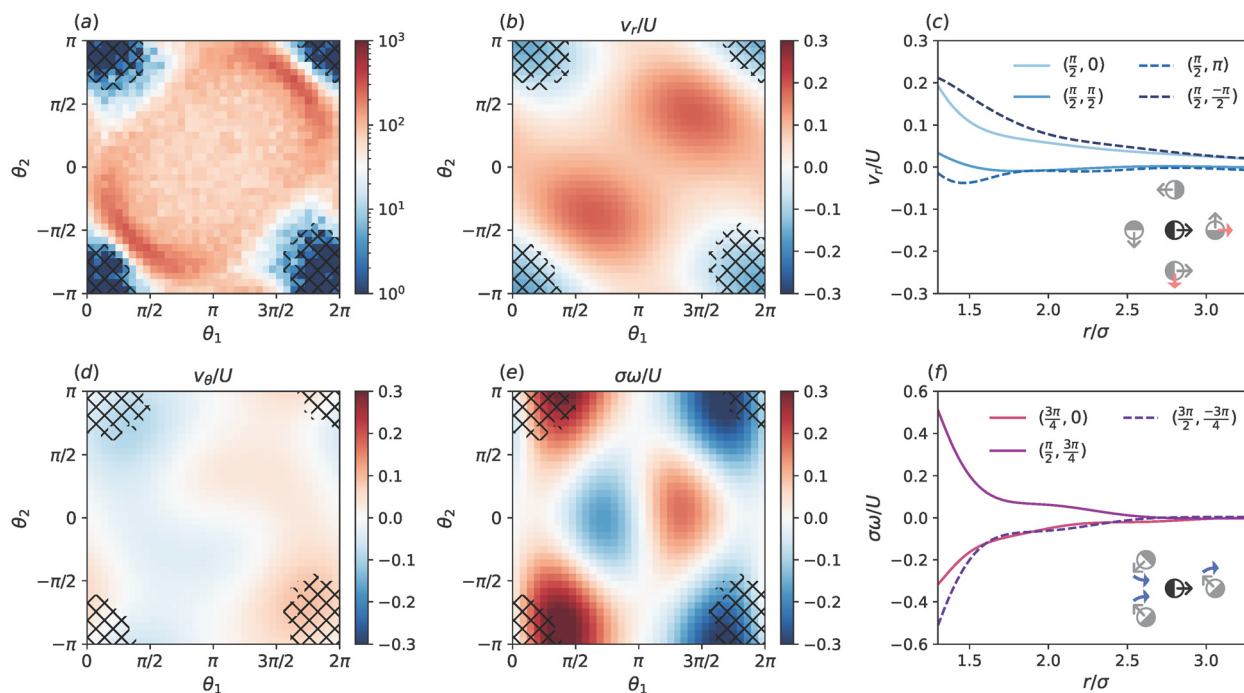
The most surprising feature is the presence of a relatively strong angular interaction, which was not considered in ref. 23 (Fig. 5(e)). The particle 1 turns away from particle 2 when particle 2 is oriented towards particle 1 ( $\theta_2 \simeq 0$ ) and it turns towards particle 2 when particle 2 turns its back on particle 1 ( $\theta_2 \simeq \pi$ ). Fig. 5(c) and (f) show the evolution of the radial and angular velocities with the distance  $r$  for specific positions. In both cases, the interactions monotonically decrease and vanish above  $r/\sigma \simeq 3$ .

### 4.2 Numerical simulations

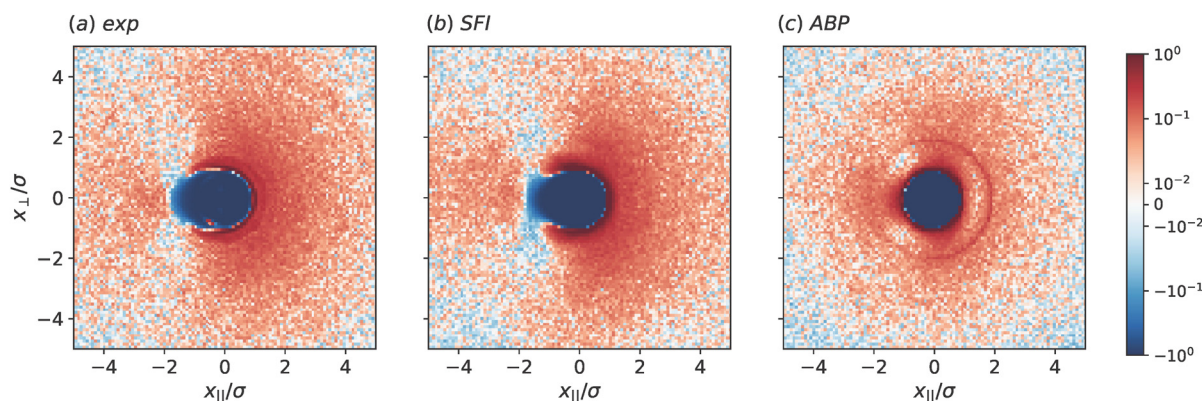
To assess the reliability of the inference, we run numerical simulations with the inferred interactions and focus on the pair correlation function  $B(\mathbf{r})$ . The experimental result is recalled in Fig. 6(a). We observe an accumulation zone in front of the particle, and a depletion zone in the back. Two depletion wings emerge in the back of the particle. To compare, we plot in Fig. 6(b) the pair correlation function of the simulation with the inferred interactions. All the features described previously are retrieved: the accumulation and depletion zones, as well as the depletion wings. In contrast, Fig. 6(c) shows the pair correlation for ABP with a soft sphere potential interaction. In this case, we see two accumulations zones in the front and in the back of the particle with two weak depletion wings emerging on the sides of the particle,<sup>23</sup> which is significantly different than in the previous figures.

Therefore, we see that despite the overall decrease of the interactions, the ABP assumption remains too crude to explain all the features of the dynamics. The pair correlation function  $B(\mathbf{r})$  is correctly recovered if one takes into account also the angular interactions present at short distance around the particle. Furthermore, this example is interesting from the viewpoint of the method, as it shows that, in spite of the multiple sources of noise, SFI is sensitive enough to capture weak interactions.





**Fig. 5** SFI results for set B (disordered system). (a) Histogram of the relative orientations at a distance  $r/\sigma = 1.5$ . (b), (d) and (e) Inferred pair interaction at a distance  $r/\sigma = 1.5$  as a function of the orientations  $\theta_1$  and  $\theta_2$ . (b) Radial, (d) azimuthal and (e) angular velocities. As before, the areas with lower statistics are hatched (bin count below 10). (c) Radial velocity as function of the distance  $r$  for specific orientations  $(\theta_1, \theta_2)$ ; inset: particle 2 (black) and particle 1 (grey) at different positions, with the red arrows indicating the direction of the radial velocity. (f) Angular velocity as function of the distance  $r$  for specific orientations; inset: particle 2 (black) and particle 1 (grey) at different positions, with the blue arrows indicating the direction of the angular velocity.



**Fig. 6** Evaluation of the pair correlation function  $B(r)$  for set B for a similar level of statistics (frame number  $8 \times 10^3$ , particle number  $\sim 500$ ). (a) Experimental measure.<sup>23</sup> (b) Simulated data with the inferred interactions. (c) Simulated data for active Brownian particles (ABP).

## 5 Symmetry considerations

SFI allowed us to determine precisely the interactions between the particles, but it does not give any information about their origin. As outlined in the introduction, none of the electrostatic or hydrodynamic interactions can be determined analytically. However, these interactions have specific properties, which can help to discriminate them. For hydrodynamic interactions, the effect of the particle 2 on the particle 1 can be expected to be dominated by the advection of the particle 1 by the flow created by the particle 2. As such, it should not depend on the orientation of the particle 1, so that  $\theta_1$ -dependent terms can

mainly be attributed to electrostatics. Since the interactions are strongly dependent on  $\theta_1$ , we conclude that the electrostatic interaction is at least present, if not dominant, in both sets A and B.

In turn, the electrostatic interaction imposes that the forces derive from an effective potential  $V(r, \theta_1, \theta_2)$ :<sup>20</sup>

$$v_r = -\mu \frac{\partial V}{\partial r}, \quad (16)$$

$$v_\theta = \frac{\mu}{r} \left[ \frac{\partial V}{\partial \theta_1} + \frac{\partial V}{\partial \theta_2} \right], \quad (17)$$

$$\omega = -\mu_r \frac{\partial V}{\partial \theta_1}, \quad (18)$$

where  $\mu$  and  $\mu_r$  and the translational and rotational mobilities. As a consequence, the interactions should follow the relations:

$$\frac{\partial v_r}{\partial \theta_1} \propto \frac{\partial \omega}{\partial r}, \quad (19)$$

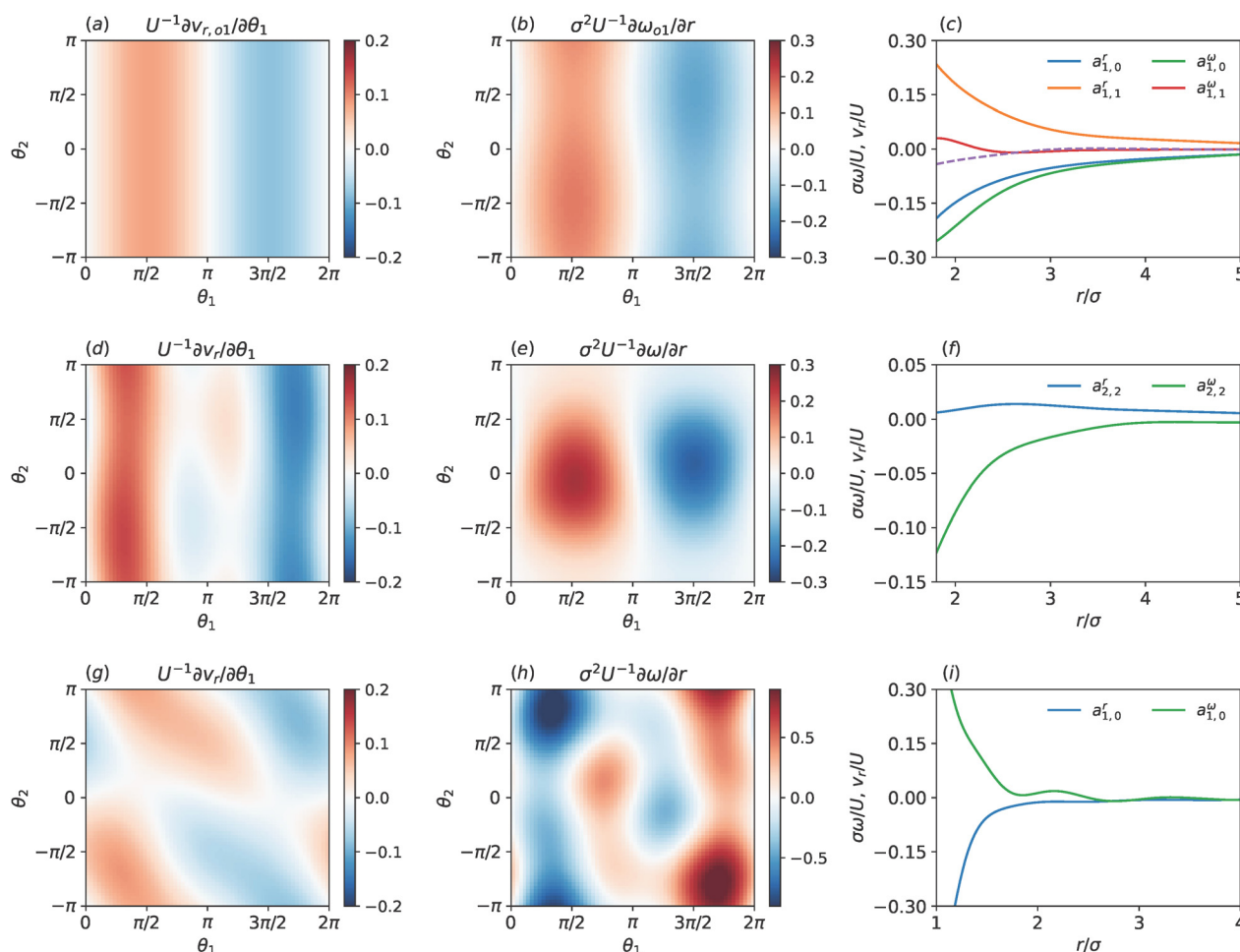
$$\frac{\partial(rv_\theta)}{\partial r} = -\frac{\partial v_r}{\partial \theta_1} - \frac{\partial v_r}{\partial \theta_2}, \quad (20)$$

where the proportionality involves the unknown ratio of translational and rotational mobilities. Since both  $v_r(r, \theta_1, \theta_2)$  and  $\omega(r, \theta_1, \theta_2)$  are significant in sets A and B, eqn (19) becomes an interesting criteria to test the presence of the electrostatic interaction in those systems.

Fig. 7(a) and (b) shows the relation (19) in case of set A and for only the first order terms of the angular basis. Overall, we see a good agreement which implies a specific behavior for each harmonic terms, see Fig. 7(c). For example, the term  $a_{1,1}^\omega$  is

almost zero as expected since  $\partial\omega/\partial\theta_2 \simeq 0$  (Fig. 2(e)). The terms  $a_{1,0}^\omega$  and  $a_{1,0}^r$  are of same sign, which can be explained as follows. Suppose  $a_{1,0}^\omega \propto r^{-n}$ , then  $a_{1,0}^r \propto n/r^{n+1}$  and the ratio is necessarily  $a_{1,0}^\omega/a_{1,0}^r > 0$ . In particular, it justifies the electrostatic origin of the “turning away” interaction observed for the angular velocity since the amplitude  $a_{1,0}^\omega$  has the expected sign. Besides, in case of electrostatic interactions, the potential should be invariant under reflection with respect to the axis defined by the particles 1 and 2,  $V(r, \theta_1, \theta_2) = V(r, -\theta_1, -\theta_2)$ , but also and under the exchange of the particles,  $V(r, \theta_1, \theta_2) = V(r, \theta_2 + \pi, \theta_1 + \pi)$ . This additional symmetry implies the equality  $a_{1,0}^r = -a_{1,1}^r$  which can be verified as  $|a_{1,0}^r| - |a_{1,1}^r| \simeq 0$  (dashed line in Fig. 7(c)). Therefore, we find at first order all the hallmarks of a dominant electrostatic interaction.

However, Fig. 7(d)–(f) show that the symmetry relations do not hold anymore if the second order terms are included. For example, the terms  $a_{2,2}^r$  and  $a_{2,2}^\omega$  behave very differently with the distance  $r$ , which is not compatible with the existence of an effective potential. For set A, we must conclude that there is



**Fig. 7** Evaluation of the symmetry relation eqn (19): (a) and (b) for set A including only the first order term of the angular basis at a distance  $r/\sigma = 2.5$ , (c) evolution of the coefficients  $a_{n,k}(r)$  with the distance. The dashed line represents the difference  $|a_{1,0}^r| - |a_{1,1}^r|$ . (d) and (e) For set A including all the terms up to the second order at  $r/\sigma = 2.5$ , (f) evolution of  $a_{2,2}^r(r)$ ,  $a_{2,2}^\omega(r)$  with the distance  $r$ . (g) and (h) For set B including all the terms up to the second order at  $r/\sigma = 1.5$ , (i) evolution of  $a_{1,0}^r(r)$ ,  $a_{1,0}^\omega(r)$  with the distance. The function  $a_{n,k}(r)$  stands for the coefficient of the harmonic term  $\cos((n - |k|)\theta_1 + k\theta_2)$  or  $\sin((n - |k|)\theta_1 + k\theta_2)$ , see notations in eqn (8)–(10).

another component present in the system, probably of hydrodynamic origin, which does not derive from a potential.

In a similar way, we conclude that the interactions in set B are a mix of electrostatic and hydrodynamic as the symmetry relations are also not satisfied (Fig. 7(g) and (h)). Note that in this case, the violation is even stronger since it is not verified at first order; for instance the terms  $a_{1,0}^r$  and  $a_{1,0}^\omega$  are of opposite sign (Fig. 7(i)). In set B, it is thus likely that the electrostatic and hydrodynamic interactions are of the same order of magnitude.

## Conclusions

We have used stochastic force inference (SFI) to learn the pair interactions between self-propelled Janus colloids from experimental trajectories. In particular, we have studied two datasets: a dense set of flocking particles and another one where the system remains disordered. Although in both cases the experimental conditions were not specifically optimized for this inference task, we successfully measured the pair interactions, as evidenced by the ability to reproduce all the statistical properties of the experimental systems in numerical simulations. Moreover, the precise determination of the interactions allowed us to gain information about the nature of the interactions. Based on the symmetry relations that an electrostatic potential should satisfy, we find that in the flocking system the interactions are mainly electrostatic, with the presence of an additional hydrodynamic component that contributes sensitively to the dynamics. In contrast, the relative contributions between the electrostatic and hydrodynamic interactions are more balanced in the second experiment.

Such information can be crucial, especially for the study of Janus colloids propelled by AC electric field. The behavior of these systems under different experimental parameters can be predicted,<sup>20,22,47</sup> but these theoretical predictions have been obtained by focusing on electrostatics only and neglecting the other influences, such as the hydrodynamic flow. Our results confirm the global shape of the interactions but also unveil unpredicted and quantitatively significant dependencies. To further characterize the interactions in these systems, one could infer them in experiments that use different parameters such as the amplitude and the frequency of the electric field.

Applied to experiments where the interactions are unknown, such as Janus colloids self-propelled through phoretic effects,<sup>48</sup> the inference of interactions could be a decisive step towards understanding and modeling the system. Whether a preliminary qualitative knowledge of the interactions exists or not, their precise determination may allow one to use numerical simulations to explore the effect of changing parameters such as the density or the system size.

An important difficulty that arises when learning the interactions is the orientation detection, albeit the orientation is clearly visible on experimental images. In other cases, such as rolling colloids,<sup>15</sup> the orientation is not even visible on the images, preventing the application of the current framework. To infer the interactions in such case, a possible route is to treat

the hidden orientation of the particles as a memory, and to extend SFI to use multiple time steps to infer the interactions. This extension would allow to deal with systems where memory is embedded in the surrounding fluid, such as for light-activated colloidal microswimmers in a viscoelastic fluid<sup>49</sup> or self-propelling droplets that modify their environment by leaving chemical footprints.<sup>13</sup>

## Conflicts of interest

There are no conflicts to declare.

## Data availability

The program codes are available in the public repository: [https://github.com/jhem272/sfiabp\\_py](https://github.com/jhem272/sfiabp_py).

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

## Appendices

### Experimental parameters

The experimental setup for both set A and B is briefly described here; more information can be found in the original articles.<sup>23,24</sup> In the two experiments, the particles under study are silica based colloids with a diameter of  $\sigma = 3.17 \mu\text{m}$ , whose half hemisphere is coated with a titanium layer. The particles are suspended in a NaCl solution, that is sandwiched between two ITO electrodes with a spacer of thickness  $h$ . Then, a perpendicular AC electric field of frequency  $f$  and amplitude  $E$  is applied, which ensure the self-propulsion of the particles.

For the video recording, a CMOS camera (Baumer LXG-80 3000 × 2400 pixels, 8 bit grayscale) is mounted on an inverted microscope (Olympus IX70 or Nikon EXLIPSE TE2000-U). To increase the contrast of the hemispheres, a green filter is inserted between the sample and the halogen lamp. Hence, in the images, the metallic hemisphere appears in dark gray and the dielectric side in light gray. In addition, several routines based on Python libraries are used for the tracking of the particles. In particular, the position of the particle is provided by the Hough transformation whereas the orientation is given by the center of mass of the pixels. These allows to extract for each particle, its position  $\mathbf{r}_i(t) = (x_i(t), y_i(t))$  and its orientation  $\varphi_i(t)$  over time.

Parameters for set A.<sup>24</sup> Video clip of  $3.1 \times 10^3$  frames with time interval  $\delta t = 0.067 \text{ s}$ , area fraction  $\phi_A = 9.7 \times 10^{-2}$  with around  $1.7 \times 10^3$  particles in a frame of size  $420 \times 336 \mu\text{m}^2$ . The particles move with the metallic (dark) face ahead in deionized water with an electrode separation  $h = 130 \mu\text{m}$  and an applied AC electric field of frequency  $f = 1 \text{ MHz}$  and amplitude  $E = 1.23 \times 10^5 \text{ V}_{\text{pp}} \text{ m}^{-1}$ . Standard error for the detection of the orientation:  $\varphi_{\text{err}} \simeq 2^\circ$ ; this value is obtained by fitting the mean squared angular displacement (MSAD) with  $\langle [\varphi(t + \Delta t) - \varphi(t)]^2 \rangle = 2D_r\Delta t + 2\varphi_{\text{err}}^2$ .<sup>23</sup> The standard error for the detection of the position is below one pixel ( $0.14 \mu\text{m}$ ) and is thus negligible.

Parameters for set Ad.<sup>24</sup> Video clip of  $3.1 \times 10^3$  frames with time interval  $\delta t = 0.1$  s, area fraction  $\phi_{\text{Ad}} = 7.5 \times 10^{-3}$ ,  $\sim 400$  particles in a frame of size  $735 \times 588 \mu\text{m}^2$ . Same experimental condition as for set A. Standard error for the orientation:  $\varphi_{\text{err}} \simeq 12^\circ$ .

Parameters for set B.<sup>23</sup> Video clip of  $8.4 \times 10^3$  frames with time interval  $\delta t = 0.1$  s, area fraction  $\phi_{\text{B}} = 4 \times 10^{-2}$  with around  $5 \times 10^2$  particles in a frame of size  $360 \times 288 \mu\text{m}^2$ . Here, the particles move with the dielectric (white) face ahead for  $h = 50 \mu\text{m}$  in a sodium chloride solution of concentration  $1 \times 10^{-4} \text{ mol L}^{-1}$  and an applied AC electric field of frequency  $f = 5 \text{ kHz}$  and amplitude  $E = 2 \times 10^5 \text{ V}_{\text{pp}} \text{ m}^{-1}$ . Standard error for the orientation:  $\varphi_{\text{err}} \simeq 9^\circ$ .

### SFI parameters

For both experiments, the Stratonovitch and Vestergaard estimators have been used for the velocities (also called “drift”) and for the rotational diffusion coefficient, respectively. The rectangular approximation has been used for the integrals.

Set A, Fig. 2. Parameters for the basis:  $N_{\text{a}} = 2$ ,  $N_{\text{r}} = 8$  and  $N_{\text{b}} = 152$ . The index  $l$  for the radial functions  $g_l(r)$  spans  $l \in \{0, 1, \dots, 7\}$  with  $r_0 = 2 \mu\text{m}$ . Cell dimension for the SFI algorithm:  $r_{\text{max}} = 21 \mu\text{m} \simeq 6.6\sigma$ . Full video analyzed ( $3.1 \times 10^3$  frames). Inferred active velocity  $U \simeq 9.7 \mu\text{m s}^{-1}$  and rotational diffusion  $D_{\text{r}} \simeq 0.092 \text{ s}^{-1}$ . With these parameters, running SFI took 2.4 h on 30 cores.

Set B, Fig. 5. Parameters for the basis:  $N_{\text{a}} = 2$ ,  $N_{\text{r}} = 16$  and  $N_{\text{b}} = 285$ . The index  $l$  for the radial functions  $g_l(r)$  spans  $l \in \{0, 1, 2, \dots, 15\}$  with  $r_0 = 1 \mu\text{m}$ . Cell dimension for the SFI algorithm:  $r_{\text{max}} = 24 \mu\text{m} \simeq 7.5\sigma$ .  $4 \times 10^3$  frames analyzed. Inferred active velocity  $U \simeq 7.0 \mu\text{m s}^{-1}$  and rotational diffusion  $D_{\text{r}} \simeq 0.085 \text{ s}^{-1}$ . With these parameters, running SFI took 2.8 h on 30 cores.

### Simulation parameters

To simulate the dynamics of the particles, we integrate numerically the equations of motion (1) and (2) with the inferred parameters, with the Euler–Maruyama scheme, with an integration time  $t_{\text{s}} = 5 \mu\text{s}$ . We use the cell-list algorithm with periodic boundary condition, the dimension of the cell being set to the value of  $r_{\text{max}}$  used in the SFI analysis. The inferred forces are activated only for a specific range  $r_{\text{min}} < r < r_{\text{max}}$  to circumvent the bad estimation at lower distance. In addition to them, to model the steric repulsion and avoid unphysical moves at small distances, we add a repulsive interaction deriving from the potential:  $V(r) = (\epsilon/2)(1 - r/\sigma)^2 H(\sigma - r)$  with strength  $\epsilon = 800$  and diameter  $\sigma = 3.17 \mu\text{m}$ , where  $H(x)$  is the Heaviside function. This short range steric interaction is not found by SFI because (i) the particles do not come into close contact in the set A (see Fig. 1(d)) and (ii) this interaction is blurred by the displacement between two frames in both sets. We also ran numerical simulations without this interaction and found no difference in the observables. The number of particles is given by  $\phi/(S_{\text{box}}\pi\sigma^2/4)$  with  $\phi$  the area fraction and  $S_{\text{box}}$  the area of the box.

Comparison of the observables, Fig. 3. For all simulations, we use the same particle number and box size as for the experiment set A. Parameters:  $r_{\text{max}} = 21 \mu\text{m}$ ,  $r_{\text{min}} = 4 \mu\text{m}$ , time interval  $\delta t = 0.1$  s between saved configurations and frame number  $4 \times 10^3$ . The initial distribution of the particles is the same as for the first experimental frame.

Inferred simulations with varying area fraction  $\phi$ , Fig. 4. Parameters for all simulations:  $r_{\text{max}} = 20 \mu\text{m}$ ,  $r_{\text{min}} = 4 \mu\text{m}$ , time interval  $\delta t = 0.1$  s, frame number  $5 \times 10^3$  and periodic box of size  $420 \times 420 \mu\text{m}^2$ . The number of particles depends on the area fraction  $\phi$ . The initial positions of the particle are uniformly distributed in the box.

Evaluation of the pair correlation function  $B(\mathbf{r})$ , Fig. 6. Spatial bin resolution  $0.3 \mu\text{m}$ . For the experiment Fig. 6(a),  $8 \times 10^3$  frames processed with time interval  $\delta t = 0.1$  s and average number of particles  $\sim 500$ . For all simulations, 500 particles constrained within a periodic box of size  $360 \times 288 \mu\text{m}^2$ ,  $r_{\text{max}} = 24 \mu\text{m}$ ,  $r_{\text{min}} = 3 \mu\text{m}$ ,  $8 \times 10^3$  frames processed with time interval  $\delta t = 0.1$  s. For the ABP simulation Fig. 6(c), the radial interaction is given only by the soft sphere interaction.

## Acknowledgements

We thank J. Iwasawa for sharing the experimental data of sets A and Ad. J. H. and V. D. acknowledge financial support from CNRS through the Emergence@INC program for the project CORRELACT. P. R. acknowledges funding from the French National Research Agency (ANR-16-CONV-0001), from Excellence Initiative of Aix-Marseille University A\*MIDEX and from the European Union (ERC-SuperStoc-101117322). D. N. is supported in part by JSPS KAKENHI Grant Numbers JP23K25838 and JP25K22005, and JST PRESTO Grant Number JPMJPR2108, Japan, and the JSPS Core-to-Core Program “Advanced core-to-core network for the physics of self-organizing active matter (JPJSCCA20230002).”

## References

- 1 I. S. Aranson, *Rep. Prog. Phys.*, 2022, **85**, 076601.
- 2 H. Xu, Y. Huang, R. Zhang and Y. Wu, *Nat. Phys.*, 2023, **19**, 46–51.
- 3 T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen and O. Shochet, *Phys. Rev. Lett.*, 1995, **75**, 1226–1229.
- 4 J. Gautrais, F. Ginelli, R. Fournier, S. Blanco, M. Soria, H. Chaté and G. Theraulaz, *PLoS Comput. Biol.*, 2012, **8**, e1002678.
- 5 X. Chen, M. Winiarski, A. Puścian, E. Knapska, A. M. Walczak and T. Mora, *Phys. Rev. X*, 2023, **13**, 041053.
- 6 N. Bain and D. Bartolo, *Science*, 2019, **363**, 46–49.
- 7 J. Cordes, A. Schadschneider and A. Nicolas, *PNAS Nexus*, 2024, **3**, pgae120.
- 8 F. Gu, B. Guiselin, N. Bain, I. Zuriguel and D. Bartolo, *Nature*, 2025, **638**, 112–119.
- 9 C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe and G. Volpe, *Rev. Mod. Phys.*, 2016, **88**, 045006.
- 10 J. Deseigne, O. Dauchot and H. Chaté, *Phys. Rev. Lett.*, 2010, **105**, 098001.



- 11 C. A. Weber, T. Hanke, J. Deseigne, S. Léonard, O. Dauchot, E. Frey and H. Chaté, *Phys. Rev. Lett.*, 2013, **110**, 208001.
- 12 Z. Izri, M. N. van der Linden, S. Michelin and O. Dauchot, *Phys. Rev. Lett.*, 2014, **113**, 248302.
- 13 B. V. Hokmabad, J. Agudo-Canalejo, S. Saha, R. Golestanian and C. C. Maass, *Proc. Natl. Acad. Sci. U. S. A.*, 2022, **119**, e2122269119.
- 14 M. Lin, P. Kim, S. Arunachalam, R. Hardian, S. Adera, J. Aizenberg, X. Yao and D. Daniel, *Phys. Rev. Lett.*, 2024, **132**, 058203.
- 15 A. Bricard, J.-B. Caussin, N. Desreumaux, O. Dauchot and D. Bartolo, *Nature*, 2013, **503**, 95–98.
- 16 J. Palacci, C. Cottin-Bizonne, C. Ybert and L. Bocquet, *Phys. Rev. Lett.*, 2010, **105**, 088304.
- 17 M. Z. Bazant and T. M. Squires, *Curr. Opin. Colloid Interface Sci.*, 2010, **15**, 203–213.
- 18 A. Boymelgreen, G. Yossifon and T. Miloh, *Langmuir*, 2016, **32**, 9540–9547.
- 19 D. Nishiguchi and M. Sano, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2015, **92**, 052309.
- 20 J. Yan, M. Han, J. Zhang, C. Xu, E. Luijten and S. Granick, *Nat. Mater.*, 2016, **15**, 1095.
- 21 D. Nishiguchi, J. Iwasawa, H.-R. Jiang and M. Sano, *New J. Phys.*, 2018, **20**, 015002.
- 22 J. Zhang, R. Alert, J. Yan, N. S. Wingreen and S. Granick, *Nat. Phys.*, 2021, **17**, 961–967.
- 23 A. Poncet, O. Bénichou, V. Démery and D. Nishiguchi, *Phys. Rev. E*, 2021, **103**, 012605.
- 24 J. Iwasawa, D. Nishiguchi and M. Sano, *Phys. Rev. Res.*, 2021, **3**, 043104.
- 25 D. Nishiguchi, *J. Phys. Soc. Jpn.*, 2023, **92**, 121007.
- 26 S. Das, M. Ciarchi, Z. Zhou, J. Yan, J. Zhang and R. Alert, *Phys. Rev. X*, 2024, **14**, 031008.
- 27 F. Cichos, K. Gustavsson, B. Mehlig and G. Volpe, *Nat. Mach. Intell.*, 2020, **2**, 94–103.
- 28 M. Ruiz-Garcia, G. C. M. Barriuso, L. C. Alexander, D. G. A. L. Aarts, L. M. Ghiringhelli and C. Valeriani, *Phys. Rev. E*, 2024, **109**, 064611.
- 29 C. Joshi, S. Ray, L. M. Lemma, M. Varghese, G. Sharp, Z. Dogic, A. Baskaran and M. F. Hagan, *Phys. Rev. Lett.*, 2022, **129**, 258001.
- 30 R. Supekar, B. Song, A. Hastewell, G. P. T. Choi, A. Mietke and J. Dunkel, *Proc. Natl. Acad. Sci. U. S. A.*, 2023, **120**, e2206994120.
- 31 S. Ha and H. Jeong, *Sci. Rep.*, 2021, **11**, 12804.
- 32 G. Jung, M. Hanke and F. Schmid, *J. Chem. Theory Comput.*, 2017, **13**, 2481–2488.
- 33 A. Malpica-Morales, P. Yatsyshin, M. A. Duran-Olivencia and S. Kalliadasis, *J. Chem. Phys.*, 2023, **159**, 104109.
- 34 S. Torquato and H. Wang, *Phys. Rev. E*, 2022, **106**, 044122.
- 35 A. Frishman and P. Ronceray, *Phys. Rev. X*, 2020, **10**, 021009.
- 36 D. B. Brückner, P. Ronceray and C. P. Broedersz, *Phys. Rev. Lett.*, 2020, **125**, 058103.
- 37 A. Gerardos and P. Ronceray, *arXiv*, 2025, preprint, arXiv: 2501.10339, DOI: [10.48550/arXiv.2501.10339](https://doi.org/10.48550/arXiv.2501.10339).
- 38 C. L. Vestergaard, P. C. Blainey and H. Flyvbjerg, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2014, **89**, 022726.
- 39 S. Amiri, Y. Zhang, A. Gerardos, C. Sykes and P. Ronceray, *Phys. Rev. Res.*, 2024, **6**, 043030.
- 40 D. Nishiguchi, K. H. Nagai, H. Chaté and M. Sano, *Phys. Rev. E*, 2017, **95**, 020601.
- 41 H. Chaté, *Annu. Rev. Condens. Matter Phys.*, 2020, **11**, 189–212.
- 42 V. Narayan, S. Ramaswamy and N. Menon, *Science*, 2007, **317**, 105–108.
- 43 M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao and R. A. Simha, *Rev. Mod. Phys.*, 2013, **85**, 1143–1189.
- 44 J. K. G. Dhont, G. W. Park and W. J. Briels, *Soft Matter*, 2021, **17**, 5613–5632.
- 45 A. Zöttl and H. Stark, *Annu. Rev. Condens. Matter Phys.*, 2023, **14**, 109–127.
- 46 M. E. Cates and J. Tailleur, *Annu. Rev. Condens. Matter Phys.*, 2015, **6**, 219–244.
- 47 J. Zhang, B. A. Grzybowski and S. Granick, *Langmuir*, 2017, **33**, 6964–6977.
- 48 F. Ginot, I. Theurkauff, D. Levis, C. Ybert, L. Bocquet, L. Berthier and C. Cottin-Bizonne, *Phys. Rev. X*, 2015, **5**, 011004.
- 49 N. Narinder, C. Bechinger and J. R. Gomez-Solano, *Phys. Rev. Lett.*, 2018, **121**, 078003.