Irreversibility and emergent structure in active matter

Emanuele Crosato,[1,2,*] Mikhail Prokopenko,[1] and Richard E. Spinney[1]
[1]*Complex Systems Research Group and Centre for Complex Systems, Faculty of Engineering, The University of Sydney, Sydney NSW 2006, Australia*
[2]*CSIRO Data61, P.O. Box 76, Epping NSW 1710, Australia*

(Received 26 February 2019; revised manuscript received 31 May 2019; published 22 October 2019)

Active matter is rapidly becoming a key paradigm of out-of-equilibrium soft matter exhibiting complex collective phenomena, yet the thermodynamics of such systems remain poorly understood. In this article we study the dynamical irreversibility of large-scale active systems capable of motility-induced phase separation and polar alignment. We use a model with momenta in both translational and rotational degrees of freedom, revealing a hidden component not previously reported in the literature. Steady-state irreversibility is quantified at each point in the phase diagram which exhibits sharp discontinuities at phase transitions. Identification of the irreversibility in individual particles lays the groundwork for discussion of the thermodynamics of microfeatures, such as defects in the emergent structure. The interpretation of the time reversal symmetry in the dynamics of the particles is found to be crucial.

DOI: 10.1103/PhysRevE.100.042613

I. INTRODUCTION

Active matter consists of particles that can utilize stored free-energy reserves to generate directed motion, and as such are characteristically out-of-equilibrium [1–4]. Examples encompass a wide range of systems of self-propelled particles, including self-catalytic colloidal suspensions [5], swimming bacteria [6,7], migrating cells [8], and animal groups [9–11]. Active matter can also refer to passive particles suspended in an active bath, e.g., passive colloids in an aqueous solution set into directed motion by collisions with active swimmers also contained in the solution [12–14]. Such directed motion, in combination with interactions among the particles, can give rise to nontrivial collective dynamics not observed in matter at thermal equilibrium, such as gathering, swarming, and swirling [15].

Widely used models of active matter include active Brownian particles (ABPs) [16] and active Ornstein-Uhlenbeck particles (AOUPs) [17]. Collective motion and kinetic phase transitions can be observed in such models, with the introduction of volume exclusion, e.g., between two-dimensional discs [18]. Indeed, systems of both ABPs and AOUPs have been shown to exhibit motility-induced phase separation (MIPS), where the particles arrange themselves into regions of high and low density [17–26]. In addition, it has been shown that ABPs with alignment interactions can exhibit polar collective motion (or "flocking") [27,28].

Determining the phase diagrams for such behavior has been an active area of research [19–29]; however, there has been less focus on the thermodynamics, especially on the nonequilibrium character of the different kinetic phases, despite some progress in related field-theoretic models [30]. For Brownian particles capable only of an order-disorder transition in polar alignment, Shim *et al.* [31] demonstrated similarly distinct regimes in the entropy production. Later, Fodor *et al.* [17] investigated the entropy production in a system of AOUPs with no alignment interactions, arguing that in a harmonic trap the dynamics respect detailed balance such that the system is in an effective equilibrium. Mandal *et al.* [32] subsequently demonstrated that when a different definition of entropy production is used the nonequilibrium character can be recovered. Recently, Shankar *et al.* [33] investigated the "hidden" components of the entropy production observed free (noninteracting) particles using an under-damped [34] descriptions of the particles' translational dynamics, but only an over-damped description of the rotational degrees of freedom.

In these studies, entropy production of the models is equated with dynamical irreversibility. However, for real active systems this relationship is less clear [35,36]. As such, while one can consider the irreversibility to be the entropy production of the *model*, in physical systems, strictly, one should interpret the reported quantities, and those in this article, as *apparent* entropy productions on the scale of observed trajectories [37]. Nevertheless, this is precisely the scale relevant to emergent structure considered here, as argued in Ref. [38]. For instance, irreversibility of fluctuations in these variables have been related to the formation of emergent structure [39,40].

In contrast to the free particles studies in Refs. [17,32,33], in this article we consider a large system of ABPs interacting via volume exclusion as well as alignment, approximating either a system of self-propelled particles or a system of passive particles in an active bath, and focus on its emergent collective behavior (i.e. polar alignment and MIPS). Further, we use an under-damped description for both translational and rotational degrees of freedom. This reveals an additional component of the irreversibility that had previously remained hidden due to coarse grained rotational dynamics even with under-damped translational dynamics [33]. We show that this coarse graining

*emanuele.crosato@sydney.edu.au

can result in either under or over-estimating the dynamical irreversibility, depending on dynamical interpretations of the model which reflect either self-propelled particles or passive particles in an active bath.

Simulation of the model allows us not only to construct the phase diagram, but also quantify the steady-state thermodynamics/irreversibility at each point in the space. Further, we are able to examine the spatial distributions of irreversibility/entropy production associated with the distinct kinetic phases alongside microfeatures, such as defects, in the emergent structures. We anticipate such methods to be useful for studying the thermodynamic character of collective phenomena that originate from large fluctuations, as in nucleation phenomena, or are strongly spatially heterogeneous as in spinodal decomposition.

II. IRREVERSIBILITY AND ENTROPY PRODUCTION

In the framework of stochastic thermodynamics, when environmental variables can be treated as relaxing instantaneously such that the system dynamics are Markovian and the medium can be treated as an ideal heat bath, the entropy production (in units where $k_B = 1$) along a single fluctuating path $\vec{\Omega} = \{\Omega(t)|t\in[t_0, \tau]\}$ is given by the log Radon-Nikodym derivative [41,42]

$$\Delta S^{\rm tot}[\vec{\Omega}] = \ln \frac{dP[\vec{\Omega}]}{dP^\dagger[\vec{\Omega}^\dagger]}. \tag{1}$$

Here P is a probability measure assigning probability to sets of paths realized by the system, $P^\dagger$ is a time reversed probability measure and $\vec{\Omega}^\dagger$ is the time reverse of trajectory $\vec{\Omega}$ defined as $\vec{\Omega}^\dagger = \{\Omega^\dagger(t)|t\in[t_0, \tau]\}$ with $\Omega^\dagger(t) = \varepsilon\Omega(\tau+t_0-t)$. The time reversal operator ε reflects variables determined to be odd under time reversal through phase space, while leaving even variables unchanged [43]. For instance, $\varepsilon z = -z$ for an odd translational variable (e.g., velocity), while $\varepsilon z = z+\pi$ for an odd rotational variable.

$\Delta S^{\rm tot}$ obeys an integral fluctuation theorem $\langle\exp[-\Delta S^{\rm tot}[\vec{\Omega}]]\rangle = 1$. Thus the strict inequality $\langle\Delta S^{\rm tot}[\vec{\Omega}]\rangle \geqslant 0$ holds by Jensen's inequality, characterizing the second law. Moreover, assuming appropriate probability densities, p, exist over Ω, this quantity can be decomposed into a system entropy production and medium entropy production contribution viz.

$$\begin{aligned}\Delta S^{\rm tot}[\vec{\Omega}] &= \Delta S^{\rm sys}[\vec{\Omega}] + \Delta S^{\rm med}[\vec{\Omega}] \\ &= \ln \frac{p(\Omega(t_0), t_0)}{p^\dagger(\Omega^\dagger(t_0), t_0)} + \ln \frac{dP\left[\vec{\Omega}_{\setminus t_0}\middle|\Omega(t_0)\right]}{dP^\dagger\left[\vec{\Omega}^\dagger_{\setminus t_0}\middle|\Omega^\dagger(t_0)\right]} \\ &= \ln \frac{p(\Omega(t_0), t_0)}{p^\dagger(\varepsilon\Omega(\tau), t_0)} + \ln \frac{dP\left[\vec{\Omega}_{\setminus t_0}\middle|\Omega(t_0)\right]}{dP^\dagger\left[\vec{\Omega}^\dagger_{\setminus t_0}\middle|\varepsilon\Omega(\tau)\right]}.\end{aligned} \tag{2}$$

$\Delta S^{\rm sys}$ captures changes in the configurational entropy of the system, while $\Delta S^{\rm med}$ quantifies the entropy exported to the environment. Here $\vec{\Omega}_{\setminus t_0} = \{\Omega(t)|t\in(t_0, \tau]\}$ and p indicates one-time probability densities such that

$$P[\vec{\Omega}\in\vec{\mathcal{A}}] = \int_{\mathcal{A}_{t_0}} d\Omega(t_0)\, p(\Omega(t_0)) P\left[\vec{\Omega}_{\setminus t_0}\in\vec{\mathcal{A}}_{\setminus t_0}\middle|\Omega(t_0)\right], \tag{3}$$

where $\vec{\mathcal{A}}$ is a set of paths on $[t_0, \tau]$. $P^\dagger$ represents a probability measure on paths under time-reversal which in the general case may differ from P. Typically this is arrived at by considering a time-reversal operation on protocol variables such that any nonautonomous driving is reversed and odd protocols (e.g., magnetic fields) have their signs reversed. In the model considered subsequently, $P^\dagger[\cdot|\cdot] = P[\cdot|\cdot]$ since all parameters are even with respect to time-reversal and homogeneous in time.

The one-time probability density $p^\dagger$ appearing in the denominator in the *system* entropy production is defined as $p^\dagger(\Omega, t) = p(\varepsilon\Omega, \tau+t_0-t)$. Together with the property that time reversal is involutive ($\varepsilon\varepsilon\Omega = \Omega$), we have $\Delta S^{\rm sys}[\vec{\Omega}] = \ln p(\Omega(t_0), t_0)/p(\Omega(\tau), \tau)$. Crucially the expectation of this quantity vanishes in the steady state, where $p(\cdot, t_0) = p(\cdot, \tau) = p^{\rm st}(\cdot)$, i.e., $\langle\Delta S^{\rm tot}[\vec{\Omega}]\rangle^{\rm st} = \langle\Delta S^{\rm med}[\vec{\Omega}]\rangle^{\rm st}$.

We consider a system $\Omega = \{y_1, \ldots, y_n\}$, described by a set of n coupled stochastic differential equations of the form

$$dy_i = A_{y_i}[\Omega(t), t]dt + B_{y_i}dW_i \tag{4}$$

$$= A^{\rm rev}_{y_i}[\Omega(t), t]dt + A^{\rm ir}_{y_i}[\Omega(t), t]dt + B_{y_i}dW_i \tag{5}$$

where W_i are Wiener processes satisfying $dW_i dW_j = \delta_{ij}dt$ and where

$$A^{\rm rev}_{y_i}[\Omega(t), t] = \frac{A_{y_i}[\Omega(t), t] - \varepsilon_{y_i}A_{y_i}[\varepsilon\Omega(t), t]}{2} \tag{6}$$

$$= -\varepsilon_{y_i}A^{\rm rev}_{y_i}[\varepsilon\Omega(t), t] \tag{7}$$

and

$$A^{\rm ir}_{y_i}[\Omega(t), t] = \frac{A_{y_i}[\Omega(t), t] + \varepsilon_{y_i}A_{y_i}[\varepsilon\Omega(t), t]}{2} \tag{8}$$

$$= \varepsilon_{y_i}A^{\rm ir}_{y_i}[\varepsilon\Omega(t), t], \tag{9}$$

are the reversible and irreversible components of the deterministic dynamics [44,45], where $\varepsilon_{y_i}\in\{-1, +1\}$ for y_i odd and even, respectively. In such a system, the medium entropy production along a path is given by the path integral

$$\Delta S^{\rm med}[\vec{\Omega}] = \int_{t_0}^{\tau} d\Delta S^{\rm med}(t). \tag{10}$$

This differential form is given by [45]

$$d\Delta S^{\rm med}(t) = \sum_{y_i\in\Omega} \frac{A^{\rm ir}_{y_i}(\Omega(t))}{D_{y_i}} \circ dy_i(t) - \frac{A^{\rm ir}_{y_i}(\Omega(t))A^{\rm rev}_{y_i}(\Omega(t))}{D_{y_i}}dt \tag{11}$$

where the diffusion coefficients are defined as $D_{y_i} = B^2_{y_i}/2$ and where the $\circ$ notation indicates a Stratonovich integration rule. We also note the convention $0/0 = 0$ for deterministic coordinates.

It is important to note that the identification of the irreversibility of a stochastic dynamical system with entropy production is valid when all dissipative degrees of freedom are captured by the model. A failure to capture all degrees of freedom, through some coarse graining procedure, is well know to underestimate the total entropy production leaving such a quantity more aptly described as an *apparent* entropy production or merely *dynamical irreversibility* denoted $\Xi^{\rm tot}[\vec{\Omega}] = \Xi^{\rm sys}[\vec{\Omega}] + \Xi^{\rm med}[\vec{\Omega}]$. Using such a quantity provides a lower

bound on the physical entropy production,

$$\langle \Xi[\vec{\Omega}]^{\rm tot}\rangle \leqslant \langle \Delta S^{\rm tot}[\vec{\Omega}]\rangle. \tag{12}$$

III. IRREVERSIBILITY OF INTERACTING ABPs

A. ABPs model and irreversibility expressions

We consider a system of N two-dimensional, disk-shaped particles of radius R, mass m, and mobility $m\gamma$ (such that γ is a inverse timescale), undergoing Brownian motion due to a heat bath with inverse temperature β, with, or subject to, active propulsion speed v_0. The translational dynamics of such particles are then augmented by Brownian rotational dynamics, also at inverse temperature β, which determine the direction of the propulsion, designed to reflect (i) the individual particles' orientation (along which they self-propel) and angular velocity, in the case of self-propulsion or (ii) the propulsion direction local to the particle, and its rate of change, of a background active medium/field, in the case of passive particles in an active bath. These rotational dynamics are characterised by an (effective) relaxation timescale $\gamma_{\rm R}$ and moment of inertia I such that the rotational mobility is $I\gamma_{\rm R}$.

The position and velocity of each particle a are denoted as $\mathbf{r}_a = \{r_a^1, r_a^2\}$ and $\mathbf{v}_a = \{v_a^1, v_a^2\}$, while the particles' intrinsic orientation, or the orientation of the local background propulsion at point $\mathbf{r}_a$, is denoted θ_a with corresponding angular momenta/rate of change ω_a. Variables without subscripts or superscripts are to be understood as the total set of such variables in the system, e.g., $\mathbf{r} = \{\mathbf{r}_1, \ldots, \mathbf{r}_N\}$. The propulsion force is modeled as $\mathcal{P}(\theta_a) = \{\mathcal{P}^1(\theta_a), \mathcal{P}^2(\theta_a)\} = m\gamma v_0 \hat{e}(\theta_a)$, where $\hat{e}(\theta_a) = \{\hat{e}^1(\theta_a), \hat{e}^2(\theta_a)\} = \{\cos(\theta_a), \sin(\theta_a)\}$. Excluded volume effects are modelled through a potential $U(\mathbf{r})$, while alignment interactions are modelled directly through a function $\mathcal{T}_a(\mathbf{r}, \theta)$ representing either the torque incident on the particle a or the value of a torque field on the medium orientation at point $\mathbf{r}_a$. These dynamics are described by the following under-damped stochastic differential equations (SDEs):

$$dr_a^j = v_a^j dt, \tag{13}$$

$$dv_a^j = -\gamma v_a^j dt + m^{-1}\mathcal{F}_a^j(\mathbf{r}, \theta)dt + \sqrt{2\gamma/\beta m}dW_{v_a^j}, \tag{14}$$

$$d\theta_a = \omega_a dt, \tag{15}$$

$$d\omega_a = -\gamma_{\rm R}\omega_a dt + I^{-1}\mathcal{T}_a(\mathbf{r}, \theta)dt + \sqrt{2\gamma_{\rm R}/\beta I}dW_{\omega_a}, \tag{16}$$

where $\mathcal{F}_a^j(\mathbf{r}, \theta) = \mathcal{P}^j(\theta_a) - \partial_{r_a^j}U_a(\mathbf{r})$, j is the spatial dimension, β is the inverse temperature (with units $k_B = 1$), and $W_{v_a^j}$ and W_{ω_a} are uncorrelated Wiener processes, such that $\langle dW_{v_a^j}dW_{\omega_a}\rangle = 0$, $\langle dW_{\omega_a}dW_{\omega_b}\rangle = \delta_{ab}dt$, and $\langle dW_{v_a^j}dW_{v_b^k}\rangle = \delta_{jk}\delta_{ab}dt$. Equivalently, the system is described by the Fokker-Planck equation for the instantaneous probability density over all dynamical variables $\Omega = \{\mathbf{r}, \mathbf{v}, \theta, \omega\}$, $p \equiv p(\Omega, t)$:

$$\begin{aligned}\partial_t p = &\sum_{a=1}^{N}\sum_{j=1}^{2}\partial_{v_a^j}\left(\gamma v_a^j - \mathcal{F}_a^j/m\right)p + (\gamma/\beta m)\partial_{v_a^j}^2 p \\ &+ \partial_{\omega_a}\left(\gamma_R\omega_a - \mathcal{T}_a^j/I\right)p + (\gamma_R/\beta I)\partial_{\omega_a}^2 p \\ &- v_a^j\partial_{x_a^j}p - \omega_a\partial_{\theta_a}p.\end{aligned} \tag{17}$$

We now consider the dynamical irreversibility, $\Xi^{\rm tot}[\vec{\Omega}]$ of this system equal, in model, to the entropy production [42] or, in a physical system, to an apparent entropy production in the dynamical degrees of freedom. Critically, the irreversibility/entropy production depends not only on the equations of motion, but on the choice of this time reversal operator ε. While there is no uncertainty in the nature of ε for the translational degrees of freedom, there is disagreement as to the best interpretation for degrees of freedom associated with active forcing. Indeed, the distinct approaches of Fodor [17] and Mandal [32] can be recast in terms of this choice. Dabelow *et al.* [38] argued that when a particle is itself active (e.g., Refs. [19,46,47]), the self-propulsion force should be considered *odd* (i.e., $\varepsilon\theta_a = \theta_a + \pi$) under TRS, while an external force acting on a passive particle (e.g., a passive tracer particle suspended in an active bath [12,14,48–51]) should be considered *even* (i.e., $\varepsilon\theta_a = \theta_a$). In contrast, Shankar *et al.* [33] considered active self-propulsion to be even under TRS, arguing that the direction of motility is the product of a physical asymmetry of the particles.

When the system is intended to model self-propelling ABPs with an intrinsic orientation (such that ω_a are angular momenta, usually understood to have odd time reversal symmetry), we suggest that an even interpretation for θ_a (i.e., $\varepsilon\theta_a = \theta_a$, $\varepsilon\omega_a = -\omega_a$) is more natural. In contrast, when the system is intended to model passive particles, θ_a represents the direction of a local flow or velocity, which may more naturally be treated as odd (i.e., $\varepsilon\theta_a = \theta_a + \pi$, $\varepsilon\omega_a = \omega_a$). We do, however, acknowledge that these interpretations are not necessarily exhaustive or unique, particularly in the case of an passive particle in an active medium for which it is debatable as to whether nonequilibrium medium variables ought to be time reversed at all, again forming the basis for the differences between Fodor [17] and Mandal [32].

We may then compute the mean rate of irreversibility/entropy production using Eq. (11). Importantly, as the Wiener processes are assumed to be uncorrelated (also known as a bipartite, or rather multipartite property [52]), we may associate entropy productions with *individual* particles, with the total being their sum. Under the odd interpretation of θ the expected irreversibility/entropy production rate in the medium for particle a is

$$\left\langle \dot{\Xi}_a^{\rm med}\middle|\Omega\right\rangle = \gamma_{\rm R}\left(\beta I\langle\omega_a^2\rangle - 1\right) + \sum_{j=1}^{2}\gamma\left\{\beta m\left\langle\left[v_a^j - v_0\hat{e}^j(\theta_a)\right]^2\right\rangle - 1\right\}. \tag{18}$$

For the even interpretation we have

$$\left\langle \dot{\Xi}_a^{\rm med}\middle|\Omega\right\rangle = \gamma_{\rm R}\left(\beta I\langle\omega_a^2\rangle - 1\right) + \sum_{j=1}^{2}\gamma\left(\beta m\langle v_a^{j\,2}\rangle - 1\right). \tag{19}$$

The derivation of Eqs. (18) and (19) is provided in Appendix A. The distinction between the parity interpretations is striking: under the even parity interpretation, the irreversibility/entropy production is manifestly a measure of the deviation away from equipartition expected at thermodynamic equilibrium in both the translational and rotational degrees of freedom. In contrast, under the odd parity interpretation the entropy production arising from the translational

variables is modified such that it quantifies deviation from an effective equipartition, relative to the instantaneous heading and propulsion speed.

B. Free particle expressions and a hidden entropy production

The above expressions are quite general, however, in the absence of alignment, external, and exclusion interactions, such that $\mathcal{T}_a(\mathbf{r},\theta)=0$ and $\partial_{r_a^i}U(\mathbf{r})=0$, in the steady state and with vanishing inertia in the relevant variables, we can recover and generalize the results for free ABPs in [33] (see Appendix B). In general, when there is under-damped translational motion, as considered here, the individual free particle irreversibility/entropy production takes the form

$$\langle\dot{\Xi}^{\rm tot}\rangle = m\beta\gamma v_0^2(1-\gamma^2\mathcal{G}) \tag{20}$$

for the odd interpretation of θ and

$$\langle\dot{\Xi}^{\rm tot}\rangle = m\beta\gamma^3 v_0^2\mathcal{G} \tag{21}$$

for the even interpretations of θ [53], where

$$\mathcal{G} = \lim_{t\to\infty} 2\int_0^t dt_1\int_0^t dt_2\, e^{-\gamma(2t-t_1-t_2)}\langle\hat{e}^1(t_1)\hat{e}^1(t_2)\rangle. \tag{22}$$

Shankar *et al.* [33], while considering over-damped rotational motion have $2\langle\hat{e}^1(t_1)\hat{e}^1(t_2)\rangle = e^{-(\beta I\gamma_{\rm R})^{-1}|t_1-t_2|}$ and thus $\mathcal{G}_{\rm over} = \{\gamma[\gamma+(\beta I\gamma_{\rm R})^{-1}]\}^{-1}$. However, a fully under-damped description of the free-particle dynamics yields $2\langle\hat{e}^1(t_1)\hat{e}^1(t_2)\rangle = e^{(\beta I\gamma_R^2)^{-1}(1-\gamma_{\rm R}|t_1-t_2|-\exp[-\gamma_{\rm R}|t_1-t_2|])}$. Except for specific choices of parameters (e.g., all free parameters set to 1), the integral has no closed form solution, but strictly satisfies $\mathcal{G}_{\rm under}\geqslant\mathcal{G}_{\rm over}$ indicating an additional hidden component in the entropy production—see details in Appendix B.

This is reminiscent of the well known property that coarse graining procedures lead to underestimates of the entropy production [54]. Such absent terms have been referred to as "anomalous" [55] or "hidden" and have been previously implicated in heat transfer where under-damped models are crucial to observe physically plausible entropy productions [45]. However, here we observe something more nuanced. Depending on the choice of time reversal symmetry, the free particle entropy can either decrease or *increase* when moving from an under to over-damped description, usually associated with a coarse-graining procedure. Explicitly, for odd θ we have

$$\langle\dot{\Xi}^{\rm tot}\rangle_{\rm under} \leqslant \langle\dot{\Xi}^{\rm tot}\rangle_{\rm over} \tag{23}$$

and for even θ

$$\langle\dot{\Xi}^{\rm tot}\rangle_{\rm under} \geqslant \langle\dot{\Xi}^{\rm tot}\rangle_{\rm over} \tag{24}$$

for stationary free ABP dynamics with under-damped translational motion.

On first inspection this appears to be at odds with the usual results and intuition that coarse-graining must reduce entropy productions. However, there are two features which make such a phenomenon possible and characterise it as a fundamentally different category of hidden entropy production as compared to, for example, Refs. [33,55]. First, we recognize that such a result does not violate the usual second law-like inequalities as found in [54,56] since the under and over-damped dynamics are not equivalent, but have differing timescales. Second, and most crucially, separating it from hidden entropy phenomena such as Refs. [33,55] is the fact that the degree of freedom which is integrated out is not itself dissipative, but merely controls the timescale on which the rotational degree of freedom θ relaxes as the dissipative variable v evolves. This is analogous to how under-damped translational models, while bounding over-damped models from above in the stationary state, can produce less entropy production transiently, under driving and relaxation [45]. In this case, despite being in the steady state, the continuous relaxation of the rotational degrees of freedom *relative* to the translation degrees of freedom yields an equivalent result.

The quantitative discrepancies are nontrivial. For instance, setting $\gamma=\gamma_{\rm R}=\beta I=1$ yields $\mathcal{G}_{\rm under}=(2e-4)\mathcal{G}_{\rm over}\simeq 1.44\,\mathcal{G}_{\rm over}$, with commensurate over and underestimates in the entropy production for odd and even θ, respectively.

IV. NUMERICAL RESULTS

A. Three distinct kinetic phases

To investigate the relationship between irreversibility and emergent structure, the system is simulated by integrating Eqs. (13)–(16), using a stochastic velocity Verlet algorithm (see Appendix D). Excluded volume interactions are modeled using a truncated and shifted Lennard-Jones potential by choosing $U(\mathbf{r})=\sum_a U_a(\mathbf{r})$ with $U_a(\mathbf{r})=\sum_{b\neq a}\epsilon[(2R/\mathbf{r}_{ab})^{12}-(4R/\mathbf{r}_{ab})^6]+\epsilon$ if $|\mathbf{r}_{ab}|\leqslant R$ and $U_a(\mathbf{r})=0$ if $|\mathbf{r}_{ab}|>R$, where ϵ is the depth of the potential well and $\mathbf{r}_{ab}=\mathbf{r}_a-\mathbf{r}_b$. In addition, informed by the Kuramoto model [57], alignment interactions are modeled as $\mathcal{T}_a(\mathbf{r},\theta)=-K\sum_{b\neq a}g(\mathbf{r}_{ab})\sin(\theta_a-\theta_b)$ where K is the coupling strength and $g(\mathbf{r}_{ab})=1$ if $|\mathbf{r}_{ab}|\leqslant 2R$ and zero otherwise. This torque function is integrable except for a single discontinuity at $2R$ which allows for the significant simplification of being able to ignore terms in $\gamma_{\rm R}$ in Eqs. (18-19), i.e those deriving from the rotational degrees of freedom. Further details can be found in the Appendix C.

We explore the model's behavior over $\gamma_{\rm R}$, K and the particles density ϕ. These variables were chosen specifically to investigate the thermodynamic character of the emergent structures, rather than those which derive from the strength of the self-propulsion force and external heat bath, which would together entirely determine the entropy production of a free particle without the rotational degree of freedom. For instance, MIPS is typically controlled using the Péclet number $\mathrm{Pe}\propto v_0\beta\sqrt{mI\gamma\gamma_{\rm R}}$ [3] by varying the propulsion force, environmental temperature and relative timescales. Instead, we restrict ourselves to varying only the relative timescales through $\gamma_{\rm R}$. Consequently, we hold all other variables constant, setting $N=10\,000$, $R=0.5$, $v_0=3$, $m=I=\gamma=1$, $\beta=50$, and $\epsilon=1$, and also utilize periodic boundary conditions.

To characterize the configurational change associated with MIPS we utilize the local (per particle) sixfold bond-orientational order: $|q_6(a)|=|\frac{1}{6}\sum_{b\in\mathcal{N}_a}e^{i6\alpha_{ab}}|$, where α_{ab} is the angle between $\mathbf{r}_{ab}$ and an arbitrary axis and $\mathcal{N}_a$ are the closest 6 neighboring particles of a. An order parameter for the phase separation is therefore provided by the average bond-orientational order $\langle|q_6(a)|\rangle$. This can be complemented by statistics of the local density $\mathcal{X}_d(\mathbf{x})$, defined as

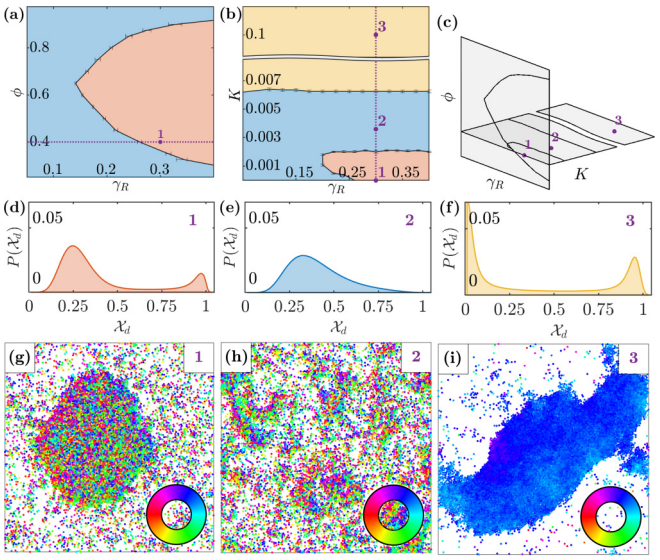

FIG. 1. Summary of the kinetic phases. Panel (a) shows the phase diagram of the system with respect to ϕ and γ_R, when $K = 0$. Panel (b) shows the phase diagram with respect to γ_R and K, at $\phi = 0.4$. Panel (c) illustrates the two sections through the ϕ–K–γ_R space. In both diagrams the error bars indicate the intervals within which the phase transitions are observed to occur, based on the simulations. The black lines are approximations of the critical lines, given the error bars. The purple lines represent trajectories across the phase diagrams over which the dynamical irreversibility is shown in Fig. 2. Three representative points along these lines, corresponding to the three phases, are labeled with numbers. For each of them, panels (d–f) show the distribution of the local density $\mathcal{X}_d$ (with $d = 4.5$), while panels (g–i) show a typical configuration observed during the simulations (color represents the particles' heading).

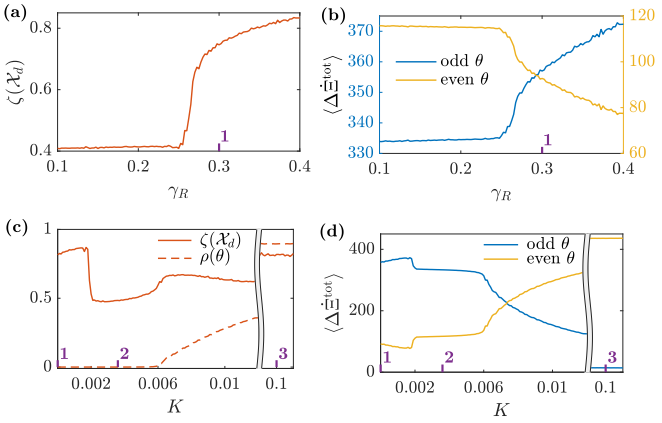

FIG. 2. Expected steady-state dynamical irreversibility over the three kinetic phases. In panels (a) and (b) $\phi = 0.4$ and $K = 0$, while γ_R is varied [cf. purple line in Fig. 1(a)]. Panel (a) shows the average bimodality coefficient $\zeta(\mathcal{X}_d)$ (with $d = 4.5$) at steady state, while panel (b) shows the expected irreversibility for both the odd and even interpretation of θ. In panels (c) and (d) $\phi = 0.4$ and $\gamma_R = 0.3$, while K is varied [cf. purple line in Fig. 1(b)]. Panel (c) shows the average $\zeta(\mathcal{X}_d)$ and the average alignment coefficient $\rho(\theta_a)$ at steady state, while panel (d) shows the irreversibility for the odd and even interpretation of θ. In all figures, the purple ticks indicate the representative points (cf. Fig. 1).

the empirical density within a radius d centered on $\mathbf{x}$ [58], since we expect a bimodal distribution under MIPS. We consider the bimodality coefficient $\zeta(\mathcal{X}_d) = [\lambda(\mathcal{X}_d) + 1]/\kappa(\mathcal{X}_d)$, where $\lambda(\mathcal{X}_d)$ and $\kappa(\mathcal{X}_d)$ are, respectively, the third and and the fourth standardized moments of $\mathcal{X}_d$. The alignment within the system is instead quantified as $\rho(\theta) = \langle 2\cos^2(\theta_a - \bar{\theta}) - 1 \rangle$, where $\bar{\theta}$ is the mean heading across all particles. We also introduce a measure of per particle alignment $\tilde{\rho}_a(\mathbf{r}, \theta) = \langle 2\cos^2(\theta_a - \bar{\theta}_{\mathcal{N}_a}) - 1 \rangle$, where $\bar{\theta}_{\mathcal{N}_a}$ is the mean heading within $\mathcal{N}_a$.

When only excluded volume interactions are considered (i.e., $K = 0$) as expected we observe two distinct phases: a phase with MIPS and a phase without MIPS, separated by a single critical value of γ_R for any given ϕ [see Fig. 1(a)]. Analogous behavior was observed in Refs. [25,26]. A third kinetic phase is possible when alignment interactions are included, characterized by both polar order and MIPS [see Fig. 1(b)]. At density $\phi = 0.4$, for example, this third phase is observed for values of $K \gtrsim 0.006$. For lower values of K the system does not exhibit polar order; however, the alignment interactions affect MIPS, which occurs only at values of $K \lesssim 0.002$. A typical configuration of the system at three values of K corresponding to the three different phases [cf. purple points in Figs. 1(a)–1(c)] is shown Figs. 1(g)–1(i), while Figs. 1(d)–1(f) show the distribution of the local density $\mathcal{X}_d$ at such values of K. Importantly, the two MIPS phases with and without polar order are emergent via two distinct and incompatible mechanisms. Explicitly, phase separation without polar order arises due to long rotational correlation times which induces jamming-like behavior, while phase separation with polar order arises due to flocking behavior. At intermediate K there is enough alignment to reduce the correlation times of the single particle rotational dynamics but not enough to cause global rotational correlations necessary for flocking.

B. Steady-state irreversibility in the three phases

The steady-state irreversibility of the three kinetic phases is illustrated by considering two representative trajectories through the phase diagram. The first follows the onset of MIPS in the absence of alignment interactions (i.e., $K = 0$) at fixed density $\phi = 0.4$ by varying γ_R indicated in Fig. 1(a). The structural and thermodynamic character along the trajectory is then illustrated in Figs. 2(a) and 2(b): Increasing γ_R up to the critical value $\sim$0.26 has little effect before an abrupt increase in the bimodality coefficient at the critical point indicating the onset of MIPS. This is accompanied by a decrease in mean particle velocity through jamming causing an equally abrupt change in the expected steady-state entropy production, calculated as the sum of all individual contributions [see Eqs. (18) and (19)]. Crucially, odd and even TRS imply completely opposite variation in the entropy production rates with an even interpretation implying lower dissipation under MIPS and vice versa. If the parity interpretation in Ref. [38] is considered (odd θ), our results would suggest that the work performed by active particles undergoing MIPS is relatively high and vice versa under the interpretation favored in Ref. [33] (even θ).

The second trajectory is indicated in Fig. 1(b) for $\phi = 0.4$ and $\gamma_R = 0.3$ as MIPS without polar order is first interrupted

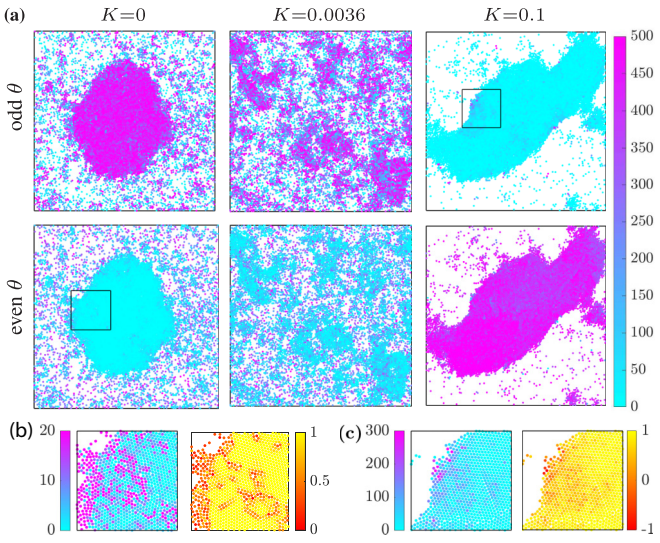

FIG. 3. Expected dynamical irreversibility associated with individual particles. Panel (a) shows the same three configurations in Figs. 1(d)–1(f) with color representing each particles' expected irreversibility, distinguishing between odd [see Eq. (18)] and even [see Eq. (19)] interpretations of θ under TRS. The left-hand side of panel (b) magnifies the box in panel (a) corresponding to $K = 0$ and even θ, using a higher resolution for the irreversibility that can capture small differences between low values. The right-hand side of panel (b) shows the local sixfold bond-orientational order $|q_6|$, highlighting spatial defects across the emergent structure. The left-hand side of panel (c) magnifies the box in panel (a) corresponding to $K = 0.1$ and odd θ, again using a different resolution for the irreversibility. The right-hand side of panel (c) shows the local alignment $\tilde{\rho}$, highlighting orientational defects.

and then reintroduced with polar order by increasing the alignment interactions through K. The relevant structural and thermodynamic consequences are then illustrated in Figs. 2(c) and 2(d). Polar order, measured through ρ, emerges beyond a critical $K \sim 0.006$. However, spatial order is more complicated with a large and increasing bimodality coefficient abruptly dropping when the jamming mechanism is interrupted, before distinctly rising at the onset of polar order due to flocking. The bimodality coefficient then slowly increases, although not monotonically, as MIPS with polar order dominates. Below the onset of polar order the dynamical irreversibility follows the spatial order as in the $K = 0$ trajectory with mean velocity controlled by jamming. However, beyond this point the irreversibility follows the polar order as the increased alignment allows for higher velocities. Once again, odd and even TRS interpretations of θ implicate opposite variation in the nonequilibrium behavior. Relatively little work is performed on/by the flocking particles under the odd interpretation [38], while more work is performed under the even interpretation [33].

C. Per particle irreversibility and defects

The spatial distribution of the entropy production can be investigated by considering the dissipation associated with individual particles [cf. Eqs. (18) and (19)]. This is exemplified in Fig. 3, for the three configurations of the system previously seen in Figs. 1(g)–1(i). In the absence of polar order the dissipative contribution from each particle closely follows the local density [Fig. 3(a), $K = 0$ and $K = 0.0036$]. When polar order is high [Fig. 3(a), $K = 0.1$], this trend with density is reversed, reflecting the distinct phase separation mechanism since dense regions occur due to flocking. Under an opposite odd/even interpretation, the trend is once again reversed due to the equilibrium being characterized by equipartition centered the laboratory frame ($\mathbf{v}_a^j = 0$) or the particle frame [$\mathbf{v}_a^j = v_0 \hat{e}^j(\theta_a)$] [cf. Eqs. (18) and (19)].

The ability to quantify the thermodynamic effects of specific local spatial configurations allows consideration of defects in the emergent structures. Specifically, we show that defects allow for either increases or decreases in the entropy production of individual particles depending on the phase and TRS interpretation—reminiscent of how crystal defects are associated with local maxima in local configurational entropy. For example, Fig. 3(b) contrasts the expected entropy production rates of individual particles with their local sixfold bond-orientational order $|q_6|$ under MIPS without polar order. In this phase, particles along the spatial defects are characterized by higher (lower) entropy production rates compared to the particles in highly ordered regions for even (odd) θ. Similarly, Fig. 3(c) contrasts the expected entropy production rates with the local alignment $\tilde{\rho}$ under MIPS with polar order. In this phase, for suitably high K, polar defects (as measured by $\tilde{\rho}$) are characterized by lower (higher) entropy production for even (odd) θ.

Further work may investigate whether any collective behavior are dependent on the existence of such defects, or fluctuations in general, and in turn their thermodynamic character. For instance, it would be of interest to capture the thermodynamics of the rare fluctuations responsible for initiating nucleation phenomena or the spatial distribution of dissipation under spinodal decomposition. Similar questions are starting to be investigated, for example through the use of large-deviation theory [39,40]. How do parity interpretations influence this description? If a nonintegrable torque function $\mathcal{T}$ is used, are the thermodynamic properties of rotational defects substantially changed?

V. CONCLUSIONS

In recent years, nonequilibrium thermodynamics and active matter have both received much attention. Lately, studies concerning their connection suggest they may have much to gain from each other [17,30–33,38–40,59]. In this study we have explored the dynamical irreversibility associated with the rich collective behavior exhibited by active matter. Our results suggest that the richness, commonly associated with the phase structure of active matter, is mirrored in its irreversibility, opening up a new tool to study collective phenomena on both a micro- and macroscopic scale. For example, the dynamical irreversibility was shown to capture the diversity between the two distinct mechanisms underlying MIPS with polar order and MIPS without polar order. Additionally, the ability to quantify dynamical irreversibility associated with specific local spatial configurations has revealed nontrivial relationships with the formation of defects in the emergent structures.

Our study has highlighted a hidden entropy production where rotational timescales impact dissipation in the

translational degrees of freedom. This is a fundamentally different finding as compared to the well known phenomenon that coarse graining procedures lead to underestimates of the entropy production [54] and, therefore, it is distinct result from hidden entropy phenomena such as [33,55]. In our under-damped model the degree of freedom which is integrated out is not itself dissipative, but merely controls the timescale on which the rotational degree of freedom relaxes as the dissipative translational degree of freedom evolves. Crucially, we observe that depending on the choice of time reversal symmetry (intimately, albeit complicatedly, related to the use of such a model to describe self-propelled particles or passive particles in an active bath) the entropy production and irreversibility can increase or decrease when moving from an under to over-damped description of the rotational dynamics.

We hope that the work will contribute to a deeper understanding of active systems and, more broadly, the dynamics that can lead to emergent structures. Important questions remain, including the delicate issue of TRS which we have shown to dramatically influence any thermodynamic interpretation. Future work will extend the analysis of the dynamical irreversibility associated to defects in the emergent structure. For example, we will aim to characterize the thermodynamics of the rare fluctuations that trigger clusters formation and the spatial distribution of irreversibility under spinodal decomposition.

ACKNOWLEDGMENTS

E.C. was supported by the University of Sydney's "Postgraduate Scholarship in the field of Complex Systems" from Faculty of Engineering & IT and by a CSIRO top-up scholarship. The authors acknowledge the University of Sydney HPC service at The University of Sydney for providing HPC resources that have contributed to the research results reported within this paper.

APPENDIX A: DERIVATION OF EXPRESSIONS FOR THE IRREVERSIBILITY

We consider the dynamics described the SDEs (13)–(16). Regardless of the odd or even interpretation of θ under TRS, we have $A^{\rm ir}_{\omega_a} = -\gamma_{\rm R}\omega_a, A^{\rm rev}_{\omega_a} = \mathcal{T}_a(\mathbf{r},\theta)/I, A^{\rm ir}_{\theta_a} = 0, A^{\rm rev}_{\theta_a} = \omega_a$, $A^{\rm ir}_{r_a^j} = 0, A^{\rm rev}_{r_a^j} = v_a^j, D_{\omega_a} = \gamma_{\rm R}/\beta I, D_{\theta_a} = 0, D_{v_a^j} = \gamma/\beta m$, and $D_{r_a^j} = 0$.

1. Even propulsion

For the even interpretation of θ (i.e., $\varepsilon\theta_a = \theta_a, \varepsilon\omega_a = -\omega_a$) we have $A^{\rm ir}_{v_a^j} = -\gamma v_a^j$ and $A^{\rm rev}_{v_a^j} = \gamma v_0 \hat{e}^j(\theta_a) - \partial_{r_a^j} U_a(\mathbf{r})/m$. Applying Eq. (11) gives

$$d\Delta\Xi^{\rm med} = -\sum_{a=1}^{N} \beta\omega_a \circ d\omega_a + \beta\omega_a \mathcal{T}_a(\mathbf{r},\theta)dt - \sum_{j=1}^{2} \beta m v_a^j \circ dv_a^j - \beta v_a^j\big[\partial_{r_a^j} U_a(\mathbf{r}) - v_0\hat{e}^j(\theta_a)\big]dt. \tag{A1}$$

Converting to Itō form, inserting the stochastic differentials dv_a^j and $d\omega_a$ and taking expectations yields

$$\frac{d\Delta\Xi^{\rm med}}{dt} = \sum_{a=1}^{N}\left[\gamma_{\rm R}\big(\beta I\langle\omega_a^2\rangle - 1\big) + \sum_{j=1}^{2}\gamma\big(\beta m\langle(v_a^j)^2\rangle - 1\big)\right], \tag{A2}$$

which may be straight forwardly decomposed into its per particle contributions. However, we may alternatively recognize that $\sum_{a,j}\langle v_a^j \circ dv_a^j\rangle = \langle dU(\mathbf{r})\rangle = 0$ in the steady state, thus proceeding with the surviving terms to obtain

$$\frac{d\Delta\Xi^{\rm med}}{dt} = \sum_{a=1}^{N}\left[\gamma_{\rm R}\big(\beta I\langle\omega_a^2\rangle - 1\big) + \sum_{j=1}^{2} m\gamma v_0\beta\langle\hat{e}^j(\theta_a)v_a^j\rangle\right], \tag{A3}$$

illustrating a useful equality, valid in the steady state,

$$\sum_{a,j}\langle\hat{e}^j(\theta_a)v_a^j\rangle = \sum_{a,j}\frac{m\beta\langle(v_a^j)^2\rangle - 1}{m\beta v_0}. \tag{A4}$$

2. Odd propulsion

For the odd interpretation of θ (i.e., $\varepsilon\theta_a = \theta_a + \pi$, $\varepsilon\omega_a = \omega_a$) we have $A^{\rm ir}_{v_a^j} = -\gamma v_a^j + \gamma v_0\hat{e}^j(\theta_a)$ and $A^{\rm rev}_{v_a^j} = -\partial_{r_a^j}U_a(\mathbf{r})/m$. Applying Eq. (11) we obtain

$$d\Delta\Xi^{\rm med} = -\sum_{a=1}^{N}\beta\omega_a \circ d\omega_a + \beta\omega_a\mathcal{T}_a(\mathbf{r},\theta)dt + \sum_{j=1}^{2}\beta\big[-mv_a^j + mv_0\hat{e}^j(\theta_a)\big]\circ dv_a^j + \beta\big[-v_a^j + v_0\hat{e}^j(\theta_a)\big]\partial_{r_a^j}U_a(\mathbf{r})dt. \tag{A5}$$

Converting to Itō form, inserting the stochastic differentials dv_a^j and $d\omega_a$ and taking expectations yields

$$d\Delta\Xi^{\rm med} = \sum_{a=1}^{N}\left(\gamma_{\rm R}\big(\beta I\langle\omega_a^2\rangle - 1\big) + \sum_{j=1}^{2}\gamma\left\{\beta m\langle\big[v_a^j - v_0\hat{e}^j(\theta_a)\big]^2\rangle - 1\right\}\right), \tag{A6}$$

which again forms a basis for a per particle contribution. Again, however, in the steady state we may assume $\sum_{a,j}\langle v_a^j \circ dv_a^j\rangle = \langle dU(\mathbf{r})\rangle = 0$ such that we have

$$d\langle\Delta\Xi^{\rm med}\rangle = \sum_{a=1}^{N}\beta\langle\omega\mathcal{T}_a(\mathbf{r},\theta)\rangle dt + m\beta\gamma v_0^2 dt - \sum_{j=1}^{2} m\beta\gamma v_0\langle\hat{e}^j(\theta_a)v_a^j\rangle dt. \tag{A7}$$

Utilizing Eq. (A4) then gives

$$\frac{d\langle \Delta \Xi^{\text{med}} \rangle}{dt} = \sum_{a=1}^{N} \Bigg[\gamma_{\text{R}} \big(\beta I \langle \omega_a^2 \rangle - 1 \big) + \sum_{j=1}^{2} \gamma \left(\frac{\beta m v_0^2}{2} - \beta m \langle \big(v_a^j \big)^2 \rangle + 1 \right) \Bigg], \quad \text{(A8)}$$

giving a starker contrast of the total contributions under odd and even interpretations of TRS for θ.

APPENDIX B: HIDDEN ENTROPY PRODUCTION BEYOND UNDER-DAMPED TRANSLATIONAL MOTION

For a single particle, i.e., setting $\partial_{r_a^j} U_a(\mathbf{r}) = \mathcal{T}_a(\mathbf{r}, \theta) = 0$, for odd θ we have

$$d\langle \Delta \Xi^{\text{med}} \rangle = m\beta\gamma v_0^2 dt - \sum_{j=1}^{2} m\beta\gamma v_0 \langle \hat{e}^j(\theta) v_a^j \rangle dt, \quad \text{(B1)}$$

and for even θ,

$$d\langle \Delta \Xi^{\text{med}} \rangle = \sum_{j=1}^{2} m\beta\gamma v_0 \langle \hat{e}^j(\theta) v_a^j \rangle dt. \quad \text{(B2)}$$

Thus explicit expressions depend on determining the steady-state correlation $\langle \hat{e}^j(\theta) v^j \rangle$. We can calculate this explicitly in the zero inertia limit for the dynamics of θ. Under such conditions we may write

$$d\theta = \sqrt{2D_{\text{R}}} dW_\theta, \quad \text{(B3)}$$

where $D_{\text{R}} = (\gamma_{\text{R}} I \beta)^{-1}$. By Itō's lemma we have

$$d\hat{e}^1 = -D_{\text{R}} \hat{e}^1 dt - \sqrt{2D_{\text{R}}(1 - (\hat{e}^1)^2)} dW_\theta, \quad \text{(B4)}$$

with analogous expression for $d\hat{e}^2$. This has integrating factor solution

$$\hat{e}^1(t) = \hat{e}^1(0) e^{-D_{\text{R}} t} - \int_0^t e^{-D_{\text{R}}(t-t')} \sqrt{1 - [\hat{e}^1(t')]^2} \sqrt{2D_{\text{R}}} dW_\theta(t'). \quad \text{(B5)}$$

Similarly, we have an integrating factor solution for $v^1(t)$

$$v^1(t) = v^1(0) e^{-\gamma t} + \int_0^t e^{-\gamma(t-t')} \gamma v_0 \hat{e}^1(t') dt' + \int_0^t e^{-\gamma(t-t')} \sqrt{\frac{2\gamma}{m\beta}} dW_{v^1}(t'), \quad \text{(B6)}$$

so

$$\langle \hat{e}^1(t_e) v^1(t_v) \rangle = \left\langle \left(\hat{e}^1(0) e^{-D_{\text{R}} t_e} - \int_0^{t_e} e^{-D_{\text{R}}(t_e - t_e')} \sqrt{2D_{\text{R}}\{1 - [\hat{e}^1(t_e')]^2\}} dW_\theta(t_e') \right) \times \left[v^1(0) e^{-\gamma t_v} + \int_0^{t_v} e^{-\gamma(t_v - t_v')} \gamma v_0 \hat{e}^1(t_v') dt_v' + \int_0^{t_v} e^{-\gamma(t_v - t_v')} \sqrt{\frac{2\gamma}{m\beta}} dW_{v^1}(t_v') \right] \right\rangle, \quad \text{(B7)}$$

or explicitly writing $\hat{e}^1(t_v')$

$$\langle \hat{e}^1(t) v^1(t) \rangle = \left\langle \left\{ \hat{e}^1(0) e^{-D_{\text{R}} t} - \int_0^t e^{-D_{\text{R}}(t-t')} \sqrt{1 - [\hat{e}^1(t')]^2} \sqrt{2D_{\text{R}}} dW_\theta(t') \right\} \times \left(v^1(0) e^{-\gamma t_v} + \int_0^{t_v} e^{-\gamma(t_v - t_v')} \gamma v_0 \left\{ \hat{e}^1(0) e^{-D_{\text{R}} t_v'} - \int_0^{t_v'} e^{-D_{\text{R}}(t_v' - t_v'')} \sqrt{2D_{\text{R}}} \sqrt{1 - [\hat{e}^1(t_v'')]^2} dW_\theta(t'') \right\} dt_v' + \int_0^{t_v} e^{-\gamma(t_v - t_v')} \sqrt{\frac{2\gamma}{m\beta}} dW_{v^1}(t_v') \right) \right\rangle. \quad \text{(B8)}$$

We compute this in the $t \to \infty$ limit corresponding to the steady state. When we do this all terms will disappear, either through the averaging, i.e., $\langle dW_i \rangle = 0$ or through vanishing exponentials, except the term that contains $dW_\theta dW_\theta$ which we write as

$$\langle \hat{e}^1(t) v^1(t) \rangle = \left\langle \gamma v_0 \int_0^{t_e} \int_0^{t_v} \int_0^{t_v'} e^{-\gamma(t_v - t_v')} e^{-D_{\text{R}}(t_e - t_e')} e^{-D_{\text{R}}(t_v' - t_v'')} 2D_{\text{R}} \sqrt{1 - [\hat{e}^1(t_e')]^2} \sqrt{1 - [\hat{e}^1(t_v'')]^2} dW_\theta(t_e') dW_\theta(t_v'') dt_v' \right\rangle. \quad \text{(B9)}$$

Sifting out with the delta correlated Wiener processes, i.e.,

$$\int_0^t \int_0^{t'} f(t'', t''') dW_\theta(t'') dW_\theta(t''') = \int_0^t \int_0^{t'} f(t'', t''') \delta(t'' - t''') dt'' dt''' = \int_0^{t'} f(t'', t'') dt'' \qquad (t > t') \quad \text{(B10)}$$

and considering $t_e = t_v$ such that $t_e > t_v'$ we write

$$\langle \hat{e}^1(t) v^1(t) \rangle = \left\langle \gamma v_0 \int_0^{t_v} \int_0^{t_v'} e^{-\gamma(t_v - t_v')} e^{-D_{\text{R}}(t_e - t_v'')} e^{-D_{\text{R}}(t_v' - t_v'')} 2D_{\text{R}} (1 - (\hat{e}^1(t_v''))^2) dt_v'' dt_v' \right\rangle, \quad \text{(B11)}$$

which becomes

$$\langle \hat{e}^1(t) v^1(t) \rangle = \gamma v_0 \int_0^{t_v} \int_0^{t_v'} e^{-\gamma(t_v - t_v')} e^{-D_{\text{R}}(t_e - t_v'')} e^{-D_{\text{R}}(t_v' - t_v'')} 2D_{\text{R}} (1 - \langle [\hat{e}^1(t_v'')]^2 \rangle) dt_v'' dt_v'. \quad \text{(B12)}$$

In the $t_e = t_v \to \infty$ limit we may safely write $\langle[\hat{e}^1(t_v'')]^2\rangle = 1/2$ and thus

$$\langle \hat{e}^1(t)v^1(t)\rangle = \gamma v_0 \int_0^{t_v}\int_0^{t_v'} e^{-\gamma(t_v - t_v')}e^{-D_{\rm R}(t_e - t_v'')}e^{-D_{\rm R}(t_v' - t_v'')}D_{\rm R}dt_v''dt_v'. \tag{B13}$$

Computing the integral and setting $t_e = t_v = t$ we find

$$\langle \hat{e}^1(t)v^1(t)\rangle = \gamma v_0 e^{-D_{\rm R}t}\frac{D_{\rm R}\cosh(D_{\rm R}t) - \gamma\sinh(D_{\rm R}t) - D_{\rm R}e^{-\gamma t}}{(D_{\rm R}+\gamma)(D_{\rm R}-\gamma)}. \tag{B14}$$

Considering the $t \to \infty$ limit then gives

$$\langle \hat{e}^1(t)v^1(t)\rangle = \frac{\gamma v_0}{2(D_{\rm R}+\gamma)}, \tag{B15}$$

and so by symmetry

$$\sum_{j=1}^{2}\langle \hat{e}^j(t)v^j(t)\rangle = \frac{\gamma v_0}{(D_{\rm R}+\gamma)}, \tag{B16}$$

and thus for odd θ gives

$$\begin{aligned} d\langle\Delta\Xi_{\rm med}\rangle &= m\beta\gamma v_0^2 dt - \sum_{j=1}^{2}\langle v^j(t)\hat{e}^j(t)\rangle m\beta\gamma v_0 dt \\ &= m\beta\gamma v_0^2 dt - \frac{m\beta\gamma^2 v_0^2}{D_{\rm R}+\gamma}dt = \frac{m\beta\gamma D_{\rm R}v_0^2}{D_{\rm R}+\gamma}dt \\ &= \frac{m\gamma\beta v_0^2}{1+\gamma\gamma_{\rm R}\beta I}dt, \end{aligned} \tag{B17}$$

and for even θ gives

$$\begin{aligned} d\langle\Delta\Xi_{\rm med}\rangle &= \sum_{j=1}^{2}\langle v^j(t)\hat{e}^j(t)\rangle m\beta\gamma v_0 dt \\ &= \frac{m\beta\gamma^2 v_0^2}{D_{\rm R}+\gamma}dt \\ &= \frac{m\gamma^2\beta^2 I\gamma_{\rm R}v_0^2}{1+\gamma\gamma_{\rm R}\beta I}dt, \end{aligned} \tag{B18}$$

in agreement with Ref. [33].

The above, however, relies upon an over-damped description for the dynamics in θ, not consistent with the full under-damped equation of motion. Direct computation of $\langle e^1(t)v^1(t)\rangle$ as above suffers from the nonlinearity of the transform of the unit vector. However, we can utilize Eq. (A4) to consider instead the long term variance $\langle[v^1(t)^2]\rangle$. We can construct this quantity using the same integrating factor solution in Eq. (B6). Squaring the expression for $v^1(t)$, expanding into constituent terms and exchanging the order of integration and expectation through Fubini's theorem, we take the limit $t \to \infty$. Expectations of individual Wiener processes, e.g., $\langle dW_{v^1}\rangle$, then vanish yielding

$$\begin{aligned} &\langle(v^1(t)^2)\rangle \\ &= \lim_{t\to\infty}\frac{2\gamma}{m\beta}\int_0^t\int_0^t e^{-\gamma(2t-t_1-t_2)}\langle dW_{v^1}(t_1)dW_{v^1}(t_2)\rangle \\ &\quad + \lim_{t\to\infty}\gamma^2 v_0^2\int_0^t dt_1\int_0^t dt_2 e^{-\gamma(2t-t_1-t_2)}\langle e^1(t_1)e^2(t_2)\rangle \\ &= \frac{1}{m\beta} + \lim_{t\to\infty}\gamma^2 v_0^2\int_0^t dt_1\int_0^t dt_2 e^{-\gamma(2t-t_1-t_2)}\langle e^1(t_1)e^2(t_2)\rangle \end{aligned} \tag{B19}$$

$$= \frac{1}{m\beta} + \frac{\gamma^2 v_0^2\mathcal{G}}{2}, \tag{B20}$$

defining $\mathcal{G}$, with the step from the first to the second line due to the sifting property of the expectation of the square of the Wiener process. When paired with Eq. (A4) this gives, for odd θ,

$$\langle\Delta\dot{\Xi}^{\rm tot}\rangle = m\beta\gamma v_0^2(1-\gamma^2\mathcal{G}), \tag{B21}$$

and for even θ,

$$\langle\Delta\dot{\Xi}^{\rm tot}\rangle = m\beta\gamma^3 v_0^2\mathcal{G}. \tag{B22}$$

When θ is described by over-damped equations of motion then $\langle e^1(t_1)e^2(t_2)\rangle_{\rm over} = (1/2)e^{-(I\beta\gamma_{\rm R})^{-1}|t_1-t_2|}$ such that $\mathcal{G} = \mathcal{G}_{\rm over} = (\gamma(\gamma + (\beta I\gamma_{\rm R})^{-1}))^{-1}$ also in agreement with Eqs (B17) and (B18).

However, $\langle e^1(t_1)e^2(t_2)\rangle$ differs under an under-damped description. To find such a form we first consider $\langle(\theta(t_2) - \theta(t_1))^2\rangle$. First we integrate to find $\theta(t)$

$$\begin{aligned} \theta(t) &= \theta(0) + \omega(0)\int_0^t dt'\, e^{-\gamma_{\rm R}t'} \\ &\quad + \sqrt{\frac{2\gamma_{\rm R}}{I\beta}}\int_0^t dt'\int_0^{t'} dW(t'')\, e^{-\gamma_{\rm R}(t'-t'')} \\ &= \theta(0) + \frac{\omega(0)}{\gamma_{\rm R}}(1-e^{-\gamma_{\rm R}t}) \\ &\quad + \sqrt{\frac{2\gamma_{\rm R}}{I\beta}}\int_0^t dW(t'')\int_{t''}^t dt'\, e^{-\gamma_{\rm R}(t'-t'')} \\ &= \theta(0) + \frac{\omega(0)}{\gamma_{\rm R}}(1-e^{-\gamma_{\rm R}t}) \\ &\quad + \sqrt{\frac{2}{I\beta\gamma_{\rm R}}}\int_0^t (1-e^{-\gamma_{\rm R}(t-t'')})dW(t''), \end{aligned} \tag{B23}$$

from which we obtain

$$\begin{aligned} &\langle(\theta(t)-\theta(0))^2\rangle \\ &= \frac{\langle\omega^2(0)\rangle}{\gamma_{\rm R}^2}(1-e^{-\gamma_{\rm R}t})^2 + \frac{2}{I\beta\gamma_{\rm R}}\int_0^t (1-e^{-\gamma_{\rm R}(t-t')})^2 dt' \\ &= \frac{\langle\omega^2(0)\rangle}{\gamma_{\rm R}^2}(1-e^{-\gamma_{\rm R}t})^2 + \frac{2}{I\beta\gamma_{\rm R}}\frac{2\gamma_{\rm R}t - 3 - e^{-2\gamma_{\rm R}t} + 4e^{-\gamma_{\rm R}t}}{2\gamma_{\rm R}}. \end{aligned} \tag{B24}$$

With $\langle\omega^2(0)\rangle = (\beta I)^{-1}$ corresponding to the steady state we have

$$\langle(\theta(t)-\theta(0))^2\rangle = \frac{2}{\gamma_{\mathrm{R}}^2\beta I}(\gamma_{\mathrm{R}}t - 1 + e^{-\gamma_{\mathrm{R}}t}),\ t>0. \quad \text{(B25)}$$

Expecting time translation invariance and symmetry we then have

$$\langle(\theta(t_2)-\theta(t_1))^2\rangle = \frac{2}{\gamma_{\mathrm{R}}^2\beta I}(\gamma_{\mathrm{R}}|t_2-t_1| - 1 + e^{-\gamma_{\mathrm{R}}|t_2-t_1|}). \quad \text{(B26)}$$

Crucially, a theorem of centered Gaussian variables states [60]

$$\langle\cos[\theta(t_2)]\cos[\theta(t_1)]\rangle = \tfrac{1}{2}e^{-\frac{1}{2}\langle[\theta(t_2)-\theta(t_1)]^2\rangle} \quad \text{(B27)}$$

such that we finally have (introducing subscripts for the nature of the *rotational* dynamics)

$$\begin{aligned}&\langle e^1(t_1)e^2(t_2)\rangle_{\text{under}}\\ &= \tfrac{1}{2}e^{-(\gamma_{\mathrm{R}}^2\beta I)^{-1}(\gamma_{\mathrm{R}}|t_2-t_1|-1+\exp[-\gamma_{\mathrm{R}}|t_2-t_1|])}\\ &= \langle e^1(t_1)e^2(t_2)\rangle_{\text{over}}\exp\left[\left(\gamma_{\mathrm{R}}^2\beta I\right)^{-1}(1-e^{-\gamma_{\mathrm{R}}|t_2-t_1|}\right]),\end{aligned} \quad \text{(B28)}$$

revealing that $\mathcal{G}_{\text{under}} \geqslant \mathcal{G}_{\text{over}}$. That is, an over-damped description systematically underestimates the rotational correlation times. Consequently, for odd and even θ, $\langle\Delta\dot{\Xi}^{\text{tot}}\rangle_{\text{under}} \leqslant \langle\Delta\dot{\Xi}^{\text{tot}}\rangle_{\text{over}}$ and $\langle\Delta\dot{\Xi}^{\text{tot}}\rangle_{\text{under}} \geqslant \langle\Delta\dot{\Xi}^{\text{tot}}\rangle_{\text{over}}$, respectively, for stationary free ABP dynamics with under-damped translational motion.

That in the steady state for a given TRS interpretation for θ, coarse-graining in the rotational dynamics causes an over or underestimation of the entropy production for any system parameters leads to the claim of a hidden entropy production associated with such coarse-graining.

The integral for $\mathcal{G}_{\text{under}}$ generally has no closed form solution, but we can calculate it in the special case where $\gamma = \gamma_{\mathrm{R}} = \beta I = 1$. In this case, it reads

$$\begin{aligned}\mathcal{G}_{\text{under}} &= \lim_{t\to\infty}\int_0^t dt_1\int_0^t dt_2\ e^{-(2t-t_1-t_2)}e^{-|t_1-t_2|+1-\exp(-|t_1-t_2|)}\\ &= \lim_{t\to\infty} e^{-2t}\{-2e^{2t} + e^{1+t-\cosh(t)+\sinh(t)}(1+e^t)\\ &\quad - e\mathrm{Ei}(-1) + e\mathrm{Ei}[-\cosh(t)+\sinh(t)]\}\\ &= e-2,\end{aligned} \quad \text{(B29)}$$

where $\mathrm{Ei}(\cdot)$ is the exponential integral. For the same parameters $\mathcal{G}_{\text{over}} = 1/2$ and so the ratio $\mathcal{G}_{\text{under}}/\mathcal{G}_{\text{over}} = 2e-4 \simeq 1.44$, leading to the the ratios $\langle\Delta\dot{\Xi}^{\text{tot}}\rangle_{\text{under}}/\langle\Delta\dot{\Xi}^{\text{tot}}\rangle_{\text{over}} = 2e-4 \simeq 1.44$ for even θ and $\langle\Delta\dot{\Xi}^{\text{tot}}\rangle_{\text{under}}/\langle\Delta\dot{\Xi}^{\text{tot}}\rangle_{\text{over}} = 6-2e \simeq 0.563$ for odd θ.

APPENDIX C: INTEGRABLE TORQUE FUNCTIONS

In the main text we use the specific torque function $\mathcal{T}_a(\mathbf{r},\theta) = -K\sum_{b\neq a} g(\mathbf{r}_{ab})\sin(\theta_a-\theta_b)$, where $g(\mathbf{r}_{ab}) = 1$ for $\mathbf{r}_{ab} < 2R$ and 0 otherwise. This can be seen to emerge from a potential function

$$U_\theta(\mathbf{r},\theta) = K\sum_{a=1}^{N}\sum_{b\neq a} g(\mathbf{r}_{ab})\cos(\theta_a-\theta_b). \quad \text{(C1)}$$

Such a potential function would also lead to a singular translational force at $\mathbf{r}_{ab} = 2R$ [deriving from $\partial_{\mathbf{r}_a} g(\mathbf{r}_{ab})$], which can be neglected in expectation. Consequently, in the stationary state, we have both $\langle\omega_a \circ d\omega_a\rangle = 0$ and $\langle dU_\theta(\mathbf{r},\theta)\rangle = \langle\mathcal{T}_a(\mathbf{r},\theta)\omega_a\rangle dt = 0$. This consequently allows for the specific simplification of the steady-state entropy productions to

$$d\Delta\Xi^{\text{med}} = \sum_{a=1}^{N}\sum_{j=1}^{2}\gamma\left\{\beta m\left\langle\left[v_a^j - v_0\hat{e}^j(\theta_a)\right]^2\right\rangle - 1\right\}, \quad \text{(C2)}$$

for odd θ and

$$\frac{d\Delta\Xi^{\text{med}}}{dt} = \sum_{a=1}^{N}\sum_{j=1}^{2}\gamma\left[\beta m\left\langle\left(v_a^j\right)^2\right\rangle - 1\right] \quad \text{(C3)}$$

for even θ. We emphasize that this simplification holds for integrable torque functions only.

APPENDIX D: NUMERICAL INTEGRATION

We utilize a stochastic velocity Verlet algorithm described in detail in Ref. [61]. For each particle, indexed by a, there are three momenta variables ω_a, v_a^1, and v_2^2. Each such variable requires two zero mean, unit variance, Gaussian distributed pseudorandom numbers, written $\phi_a^{1,1}, \phi_a^{1,2}, \phi_a^{2,1}, \phi_a^{2,2}, \phi_a^{\theta,1}, \phi_a^{\theta,2}$, which are all mutually independent (i.e., $\langle\phi_a^{x,y}\phi_b^{m,n}\rangle = \delta_{ab}\delta_{x,m}\delta_{y,n}$). Recalling $\mathbf{r} = \{\mathbf{r}_1,\ldots,\mathbf{r}_N\}$ (with similarly defined $\mathbf{v}$, θ, ω) and $\mathbf{r}_a = \{r_a^1, r_a^2\}$, $\mathbf{v}_a = \{v_a^1, v_a^2\}$, the algorithm then reads

$$\begin{aligned}C_a^1(t) &= \frac{(\Delta t)^2}{2}\left\{m^{-1}\mathcal{F}_a^1[\mathbf{r}(t),\theta(t)] - \gamma v_a^1(t)\right\} + \sqrt{\frac{\gamma}{m\beta}}\frac{(\Delta t)^{3/2}}{2}\left(\phi_a^{1,1} + \frac{\phi_a^{1,2}}{\sqrt{3}}\right),\\ C_a^2(t) &= \frac{(\Delta t)^2}{2}\left\{m^{-1}\mathcal{F}_a^2[\mathbf{r}(t),\theta(t)] - \gamma v_a^2(t)\right\} + \sqrt{\frac{\gamma}{m\beta}}\frac{(\Delta t)^{3/2}}{2}\left(\phi_a^{2,1} + \frac{\phi_a^{2,2}}{\sqrt{3}}\right),\\ C_a^\theta(t) &= \frac{(\Delta t)^2}{2}\left\{I^{-1}\mathcal{T}_a[\mathbf{r}(t),\theta(t)] - \gamma_{\mathrm{R}}\omega_a(t)\right\} + \sqrt{\frac{\gamma_{\mathrm{R}}}{I\beta}}\frac{(\Delta t)^{3/2}}{2}\left(\phi_a^{\theta,1} + \frac{\phi_a^{\theta,2}}{\sqrt{3}}\right),\\ r_a^1(t+\Delta t) &= r_a^1(t) + v_a^1(t)\Delta t + C_a^1(t),\\ r_a^2(t+\Delta t) &= r_a^2(t) + v_a^2(t)\Delta t + C_a^2(t),\\ \theta_a(t+\Delta t) &= \theta_a(t) + \omega_a(t)\Delta t + C_a^\theta(t),\end{aligned}$$

$$v_a^1(t+\Delta t) = v_a^1(t) - \gamma v_a^1(t)\Delta t + \frac{\Delta t}{2m}\left\{\mathcal{F}_a^1[\mathbf{r}(t),\theta(t)] + \mathcal{F}_a^1[\mathbf{r}(t+\Delta t),\theta(t+\Delta t)]\right\} - \gamma C_a^1(t) + \sqrt{\frac{\gamma\Delta t}{m\beta}}\phi_a^{1,1},$$

$$v_a^2(t+\Delta t) = v_a^2(t) - \gamma v_a^2(t)\Delta t + \frac{\Delta t}{2m}\left\{\mathcal{F}_a^2[\mathbf{r}(t),\theta(t)] + \mathcal{F}_a^2[\mathbf{r}(t+\Delta t),\theta(t+\Delta t)]\right\} - \gamma C_a^2(t) + \sqrt{\frac{\gamma\Delta t}{m\beta}}\phi_a^{2,1},$$

$$\omega_a(t+\Delta t) = \omega_a(t) - \gamma\omega_a(t)\Delta t + \frac{\Delta t}{2I}\left\{\mathcal{T}_a^\theta[\mathbf{r}(t),\theta(t)] + \mathcal{T}_a^\theta[\mathbf{r}(t+\Delta t),\theta(t+\Delta t)]\right\} - \gamma C_a^\theta(t) + \sqrt{\frac{\gamma_R\Delta t}{I\beta}}\phi_a^{\theta,1}.$$

These equations were simulated using $\Delta t = 0.008$.

[1] S. Ramaswamy, Annu. Rev. Condens. Matter Phys. **1**, 323 (2010).
[2] M. C. Marchetti, J.-F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, and R. A. Simha, Rev. Mod. Phys. **85**, 1143 (2013).
[3] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe, Rev. Mod. Phys. **88**, 045006 (2016).
[4] S. Ramaswamy, J. Stat. Mech.: Theor. Exp. 5 (2017) 054002.
[5] J. Bialké, T. Speck, and H. Löwen, J. Non-Cryst. Solids **407**, 367 (2015).
[6] A. Czirók, M. Matsushita, and T. Vicsek, Phys. Rev. E **63**, 031915 (2001).
[7] A. Sokolov, R. E. Goldstein, F. I. Feldchtein, and I. S. Aranson, Phys. Rev. E **80**, 031903 (2009).
[8] B. Szabo, G. J. Szöllösi, B. Gönci, Z. Jurányi, D. Selmeczi, and T. Vicsek, Phys. Rev. E **74**, 061908 (2006).
[9] J. K. Parrish, S. V. Viscido, and D. Grünbaum, Biol. Bull. **202**, 296 (2002).
[10] J. Buhl, D. J. T. Sumpter, I. D. Couzin, J. J. Hale, E. Despland, E. R. Miller, and S. J. Simpson, Science **312**, 1402 (2006).
[11] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, and M. Viale, Proc. Natl. Acad. Sci. USA **107**, 11865 (2010).
[12] C. Maggi, M. Paoluzzi, N. Pellicciotta, A. Lepore, L. Angelani, and R. Di Leonardo, Phys. Rev. Lett. **113**, 238303 (2014).
[13] A. Argun, A.-R. Moradi, E. Pinçe, G. B. Bagci, A. Imparato, and G. Volpe, Phys. Rev. E **94**, 062150 (2016).
[14] C. Maggi, M. Paoluzzi, L. Angelani, and R. Di Leonardo, Sci. Rep. **7**, 17588 (2017).
[15] T. Vicsek and A. Zafeiris, Phys. Rep. **517**, 71 (2012).
[16] L. Schimansky-Geier, M. Mieth, H. Rosé, and H. Malchow, Phys. Lett. A **207**, 140 (1995).
[17] É. Fodor, C. Nardini, M. E. Cates, J. Tailleur, P. Visco, and F. van Wijland, Phys. Rev. Lett. **117**, 038103 (2016).
[18] M. C. Marchetti, Y. Fily, S. Henkes, A. Patch, and D. Yllanes, Curr. Opin. Colloid Interface Sci. **21**, 34 (2016).
[19] Y. Fily and M. C. Marchetti, Phys. Rev. Lett. **108**, 235702 (2012).
[20] G. S. Redner, M. F. Hagan, and A. Baskaran, Phys. Rev. Lett. **110**, 055701 (2013).
[21] G. S. Redner, A. Baskaran, and M. F. Hagan, Phys. Rev. E **88**, 012305 (2013).
[22] M. E. Cates and J. Tailleur, Annu. Rev. Condens. Matter Phys. **6**, 219 (2015).
[23] A. P. Solon, J. Stenhammar, R. Wittkowski, M. Kardar, Y. Kafri, M. E. Cates, and J. Tailleur, Phys. Rev. Lett. **114**, 198301 (2015).
[24] L. F. Cugliandolo, P. Digregorio, G. Gonnella, and A. Suma, Phys. Rev. Lett. **119**, 268002 (2017).
[25] P. Digregorio, D. Levis, A. Suma, L. F. Cugliandolo, G. Gonnella, and I. Pagonabarraga, Phys. Rev. Lett. **121**, 098003 (2018).
[26] J. T. Siebert, F. Dittrich, F. Schmid, K. Binder, T. Speck, and P. Virnau, Phys. Rev. E **98**, 030601(R) (2018).
[27] F. D. C. Farrell, M. C. Marchetti, D. Marenduzzo, and J. Tailleur, Phys. Rev. Lett. **108**, 248101 (2012).
[28] A. Martín-Gómez, D. Levis, A. Díaz-Guilera, and I. Pagonabarraga, Soft Matter **14**, 2610 (2018).
[29] S. C. Kapfer and W. Krauth, Phys. Rev. Lett. **114**, 035702 (2015).
[30] C. Nardini, É. Fodor, E. Tjhung, F. van Wijland, J. Tailleur, and M. E. Cates, Phys. Rev. X **7**, 021007 (2017).
[31] P.-S. Shim, H.-M. Chun, and J. D. Noh, Phys. Rev. E **93**, 012113 (2016).
[32] D. Mandal, K. Klymko, and M. R. DeWeese, Phys. Rev. Lett. **119**, 258001 (2017).
[33] S. Shankar and M. C. Marchetti, Phys. Rev. E **98**, 020604(R) (2018).
[34] Here, and throughout "under-damped" is intended to refer to a full phase space dynamics with momentum variables, in contrast to an over-damped dynamics which integrates them out. We do not require that the system is in an under-damped *limit*.
[35] P. Pietzonka and U. Seifert, J. Phys. A: Math. Theor. **51**, 01LT01 (2017).
[36] T. Speck, Europhys. Lett. **114**, 30006 (2016).
[37] U. Seifert, Annu. Rev. Condens. Matter Phys. **10**, 171 (2019).
[38] L. Dabelow, S. Bo, and R. Eichhorn, Phys. Rev. X **9**, 021009 (2019).
[39] F. Cagnetta, F. Corberi, G. Gonnella, and A. Suma, Phys. Rev. Lett. **119**, 158002 (2017).
[40] T. Nemoto, E. Fodor, M. E. Cates, R. L. Jack, and J. Tailleur, Phys. Rev. E **99**, 022605 (2019).
[41] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).
[42] U. Seifert, Eur. Phys. J. B **64**, 423 (2008).
[43] R. E. Spinney and I. J. Ford, Phys. Rev. Lett. **108**, 170603 (2012).
[44] H. Risken, *The Fokker-Planck Equation* (Springer, Berlin, 1996), pp. 63–95.
[45] R. E. Spinney and I. J. Ford, Phys. Rev. E **85**, 051113 (2012).

[46] L. Berthier and J. Kurchan, Nat. Phys. **9**, 310 (2013).
[47] T. F. F. Farage, P. Krinninger, and J. M. Brader, Phys. Rev. E **91**, 042310 (2015).
[48] E. Ben-Isaac, Y. K. Park, G. Popescu, F. L. H. Brown, N. S. Gov, and Y. Shokef, Phys. Rev. Lett. **106**, 238103 (2011).
[49] 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 **158**, 822 (2014).
[50] É. Fodor, M. Guo, N. Gov, P. Visco, D. Weitz, and F. van Wijland, Europhys. Lett. **110**, 48005 (2015).
[51] S. Chaki and R. Chakrabarti, Physica A **511**, 302 (2018).
[52] R. E. Spinney, J. T. Lizier, and M. Prokopenko, Phys. Rev. E **98**, 032141 (2018).
[53] We note that the sum of Eqs. (20) and (21) is a constant equal to $m\beta\gamma v_0^2$, consistent with results for a lattice model reported in Ref. [35].
[54] M. Esposito, Phys. Rev. E **85**, 041125 (2012).
[55] A. Celani, S. Bo, R. Eichhorn, and E. Aurell, Phys. Rev. Lett. **109**, 260603 (2012).
[56] K. Kawaguchi and Y. Nakayama, Phys. Rev. E **88**, 022147 (2013).
[57] J. A. Acebrón, L. L. Bonilla, C. J. Pérez Vicente, F. Ritort, and R. Spigler, Rev. Mod. Phys. **77**, 137 (2005).
[58] Given a radius d, we define the local density as $\mathcal{X}_d(\mathbf{x}) = \sum_{a=1}^{N} \lambda(C(d,\mathbf{x}) \cap C(R,\mathbf{r}_a))/\pi d^2$, where $C(z,\mathbf{u})$ is the set of points in a disk with radius z centered on $\mathbf{u}$, $\cap$ represents set intersection and λ is the area (i.e., the Lebesgue measure). We find the distribution by averaging $p(\hat{\mathcal{X}}_d) = \langle \delta[\hat{\mathcal{X}}_d - \mathcal{X}_d(\mathbf{x})] \rangle_{\mathbf{x}}$.
[59] E. Crosato, R. E. Spinney, R. Nigmatullin, J. T. Lizier, and M. Prokopenko, Phys. Rev. E **97**, 012120 (2018).
[60] W. T. Coffey and Y. P. Kalmykov, *The Langevin Equation With Applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering*, 3rd ed. (World Scientific, Singapore, 2012).
[61] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Stochastic Modelling and Applied Probability (Springer-Verlag, Berlin/Heidelberg, 1992).