Title
Entropy production fluctuations encode collective behavior in active matter

Permalink
https://escholarship.org/uc/item/648260bq

Journal
Physical Review E, 103(1)

ISSN
2470-0045

Authors
GrandPre, Trevor
Klymko, Katherine
Mandadapu, Kranthi K
et al.

Publication Date
2021

DOI
10.1103/physreve.103.012613

Peer reviewed

Entropy production fluctuations encode collective behavior in active matter

Trevor GrandPre,[1] Katherine Klymko,[2] Kranthi K. Mandadapu,[3, 4] and David T. Limmer[4, 5, 6, 7]

[1]*Department of Physics, University of California, Berkeley CA 94609*
[2]*Computational Research Division, Lawrence Berkeley National Laboratory CA 94609*
[3]*Department of Chemical and Biomolecular Engineering, University of California, Berkeley CA 94609*
[4]*Chemical Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94609*
[5]*Department of Chemistry, University of California, Berkeley CA 94609*
[6]*Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94609*
[7]*Kavli Energy NanoScience Institute, Berkeley, CA 94609**
(Dated: December 23, 2020)

We derive a general lower bound on distributions of entropy production in interacting active matter systems. The bound is tight in the limit that interparticle correlations are small and short-ranged, which we explore in four canonical active matter models. In all models studied, the bound is weak where collective fluctuations result in long-ranged correlations, which subsequently links the locations of phase transitions to enhanced entropy production fluctuations. We develop a theory for the onset of enhanced fluctuations and relate it to specific phase transitions in active Brownian particles. We also derive optimal control forces that realize the dynamics necessary to tune dissipation and manipulate the system between phases. In so doing, we uncover a general relationship between entropy production and pattern formation in active matter, as well as ways of controlling it.

Active matter systems are defined by forces that inject energy locally into individual particles, driving nonequilibrium steady-states that continuously dissipate energy. This persistent dissipation and its associated entropy production have been shown to have deep connections with structural and dynamic properties of active matter [1–10]. Subsequently, understanding the contributions to the entropy production in active matter is the first step in manipulating their emergent order [11–19], designing active metamaterials with novel responses [20–23], and utilizing active heat engines [24–30]. Stochastic thermodynamics provides a framework for studying entropy production and has supplied general theories that constrain its statistics [31–35] and its role in nonequilbrium response [36–50]. Here, we provide a general bound on the distributions of entropy production for interacting active matter using stochastic thermodynamics and large deviation theory [51]. While not universal like the thermodynamic uncertainty principle [52, 53], the specific consideration of active matter admits a tight bound generically, and one in which deviations can be physically understood. The bound we present is valid arbitrarily far from equilibrium for self-propelled particles and is saturated in the limit that the interparticle contribution to the entropy production is small. Near phase transitions, the bound is weak as fluctuations are enhanced due to emergent effective long-ranged interactions that we quantify. This work provides a link between entropy production fluctuations and collective phenomena in active matter.

We consider active matter systems that are self-propelled and whose equations of motion are of the form

$$\dot{\boldsymbol{r}}_i = v\,\boldsymbol{b}_i + \mu \boldsymbol{F}_i\left(\boldsymbol{r}^N\right) + \sqrt{2D_\mathrm{t}}\boldsymbol{\eta}_i \;, \qquad (1)$$

where $\boldsymbol{r}_i$ denotes the position of the i'th particle, v and $\boldsymbol{b}_i$ set the typical magnitude and direction of self-propulsion, μ is a single particle mobility, and $\boldsymbol{\eta}_i$ is a Gaussian white noise with $\langle \eta_i^\alpha(t) \rangle = 0$ and $\langle \eta_i^\alpha(t)\eta_j^\beta(t') \rangle = \delta_{ij}\delta_{\alpha,\beta}\delta(t - t')$ for the α and β components of the random force. The translational diffusion coefficient, D_t, satisfies a fluctuation-dissipation relation, $D_\mathrm{t} = \beta^{-1}\mu$ where β^{-1} is the temperature times Boltzmann's constant. Throughout, we take $\mu = D_\mathrm{t} = 1$. The interparticle forces are conservative, $\boldsymbol{F}(\boldsymbol{r}^N) = -\nabla U(\boldsymbol{r}^N)$, and in general depend on all N particles' positions, $\boldsymbol{r}^N$. This class of active matter has a non-conservative self-propulsion term, $v\boldsymbol{b}$, which is driven by a constant energy supply. Our formulation is independent of the statistics and dynamics of the self propulsion vector, $\boldsymbol{b}$, and may be correlated due to aligning interactions. The dynamics of the orientation vector $\boldsymbol{b}$ are model specific and discussed in Appendix A, however our results are largely independent of its form. For concreteness, below we will consider collections of interacting active Brownian particles (ABPs), active dumbbells (ADPs), run and tumble particles (RTPs), and active Ornstein-Uhlenbeck particles (AOUPs).

The entropy production follows from time reversal symmetry arguments of stochastic thermodynamics [54–58], $\Delta S = \ln P[\Gamma]/P[\tilde{\Gamma}]$, where $P[\Gamma]$ is the probability of a forward trajectory $\Gamma = \{\boldsymbol{r}^N(t), \boldsymbol{b}^N(t)\}$ and $P[\tilde{\Gamma}]$ is the probability of observing the time-reversed trajectory. We use the convention that the parameter v is even under time-reversal consistent with previous work [59–63]. This convention ensures that there is a nonvanishing entropy production in the limit of noninteracting particles with no external fields accounting for the energy injected into the single particles to drive persistent motion. The choice of convention for time reversal without an underlying microscopic model of self-propulsion is somewhat arbitrary [64]. However, all of the collective phenomena reported below are independent of the convention [65]. The
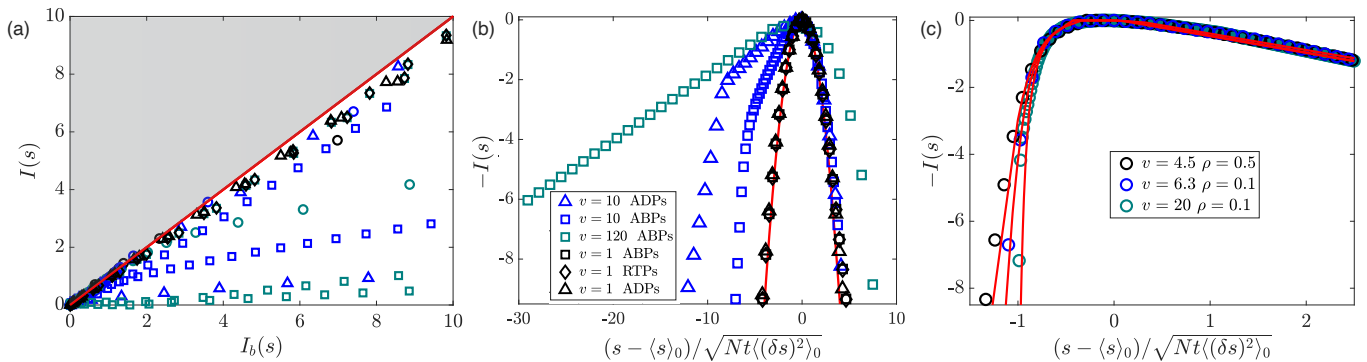
FIG. 1. Entropy production fluctuations for a variety of active matter systems. (a) Rate function obtained by importance sampling versus the bound in Eq. 8 with the symbols corresponding to the labels in (b) and (c). (b) Entropy production fluctuations for $\rho = 0.1$, ABPs with $D_\mathrm{r} = 3$, ADPs with spring constant $k = 100 \; \epsilon/\sigma^2$ and rest length $l = 1.5\sigma$, and RTPs with a tumble rate $\gamma = 1$ for different self-propulsion values. (c) Entropy production fluctuations for AOUPs for different parameter ranges and $D_\mathrm{r} = 1$. In all panels the red line denotes $I_b(s)$ and the errorbars are smaller than the symbols.

convention we follow is analogous to the active work [61] and can be derived independently from mechanical considerations. Under this time reversal convention in the long time limit, the entropy production is

$$\Delta S = \frac{v}{D_\mathrm{t}} \sum_{i=1}^{N} \int_0^t dt' \; \boldsymbol{b}_i \circ \dot{\boldsymbol{r}}_i \;, \qquad (2)$$

where $\circ$ denotes a Stratonovich product (see Appendix B). This definition codifies the amount of energy directly translated into motion in the form of persistent displacement [63, 64, 66].

One convenient way to characterize the statistics of ΔS is through its scaled cumulant generating function (CGF). For the time and system size intensive entropy production, $s = \Delta S/(Nt)$, the CGF is defined as,

$$\psi(\lambda) = \frac{1}{tN} \ln \left\langle e^{\lambda s(\Gamma)Nt} \right\rangle_0, \qquad (3)$$

where $\langle \ldots \rangle_0$ denotes average over paths and λ is the counting variable that probes rare fluctuations of the entropy production when nonzero. Cumulants of the entropy production are computable from $\psi(\lambda)$ through derivatives with respect to λ. We define a rate function

$$I(s) = -\frac{1}{Nt} \ln \left\langle \delta[s - s(\Gamma)] \right\rangle_0 \;, \qquad (4)$$

where $\delta(s)$ is Dirac's delta function. The rate function is the logarithm of the probability of s scaled by time and particle number. We are interested in the fluctuations of s in the macroscopic limit at long time and large system size, where $I(s)$ can be calculated by a Legendre-Fenchel transform, $I(s) = \max_\lambda [\lambda s - \psi(\lambda)]$. The transient fluctuations of entropy production would require an alternative method from the Legendre-Fenchel transform and are not considered in our work.

Calculating $\psi(\lambda)$ or $I(s)$ exactly for interacting systems is difficult because of many-body correlations. However, we find that $\psi(\lambda)$ can generally be rewritten by

factoring out the single particle part,

$$\psi(\lambda) = \psi_f(\lambda) + \frac{1}{Nt} \ln \left\langle e^{\lambda \Delta W} \right\rangle_{\boldsymbol{u}_\lambda} \;, \qquad (5)$$

where $\psi_f(\lambda)$ is the CGF for an isolated active particle. The remaining contribution to $\psi(\lambda)$ represents interparticle correlations and is given by the CGF of

$$\Delta W = \beta v \sum_{i=1}^{N} \int_0^t dt' \; \boldsymbol{b}_i \cdot \boldsymbol{F}_i \;, \qquad (6)$$

averaged over an ensemble with an additional force $\boldsymbol{u}_\lambda$. The force $\boldsymbol{u}_\lambda$ is the optimal control force to realize rare entropy production fluctuations for an isolated particle and its model specific form is considered below. The observable ΔW is the dimensionless work done on the surrounding particles due to self-propulsion. By applying Jensen's inequality to Eq. 5,

$$\psi(\lambda) \geq \psi_f(\lambda) + \beta \lambda v \left\langle \boldsymbol{b} \cdot \boldsymbol{F} \right\rangle_{\boldsymbol{u}_\lambda} \;, \qquad (7)$$

$\psi(\lambda)$ is bounded (see Appendix D). The correction over the single particle CGF can be interpreted as $\beta \lambda v$ times the effective drag a tagged particle feels in the direction of the self-propulsion due to the surrounding particles [67]. This gives rise to an effective velocity that is smaller than v and dependent on the density and λ [68].

Inserting the bound on the CGF in Eq. 7 into the Legendre-Fenchel transform, we derive a bound on the distribution of the entropy production, $I_b(s)$,

$$I(s) \leq I_b(s) = \max_\lambda \left[\lambda s - \psi_f(\lambda) - \beta \lambda v \left\langle \boldsymbol{b} \cdot \boldsymbol{F} \right\rangle_{\boldsymbol{u}_\lambda} \right]. \qquad (8)$$

By construction the bound recovers the correct mean dissipation and is tight far into the tails of the distribution in the limit that fluctuations in ΔW are small and the saddle point approximation to its CGF is accurate. Data in Fig. 1 confirms the upper bound for all of the active matter models studied. Throughout, $I(s)$ is computed using the cloning algorithm [69, 70] and $I_b(s)$ by

computing $\langle \Delta W \rangle_{\boldsymbol{u}_\lambda}$ from direct simulations. All simulations are done with a WCA interparticle potential [71], $U(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right] + \epsilon$ for $r \leq 2^{1/6}\sigma$ and zero otherwise. The parameter ϵ is the energy scale of the interactions and σ is the particle diameter. ADPs have an added harmonic potential between composite particles. Our results are presented with a non-dimensional v in units of D_t/σ, γ and D_r in units of D_t/σ^2, and bulk density ρ in units of $1/\sigma^2$ in two dimensions. Also, D_t, and β are set to 1. Data in Fig. 1a shows that there are large parameter regimes where the bound is tight. In practice, the bound is accurate when the system is away from dynamical phase transitions, this is valid when $\rho(v/D_\mathrm{r})^2 < 1$. Nevertheless, even when $\rho(v/D_\mathrm{r})^2 \approx 1$ we find the bound is still reasonably tight.

The detailed forms for $I(s)$ and $I_b(s)$ are distinct for different models of active matter. For ABPs, ADPs, and RTPs, the entropy production fluctuations are Gaussian for isolated particles, with $\psi_f(\lambda) = v^2\lambda(1+\lambda)/D_\mathrm{t}$ (see Appendix C). The corresponding control force, $\boldsymbol{u}_\lambda = 2\lambda v \boldsymbol{b}$, is appended to the existing forces in Eq. 1 such that rare entropy production fluctuations are realized by a renormalized velocity, $v_\lambda = v(1 + 2\lambda)$. This $\psi_f(\lambda)$ gives rise to a bound that is nearly Gaussian, as shown in Fig. 1b. For low densities and low velocities, $I(s) \approx I_b(s)$. Increasing v, the bound weakens for smaller than average entropy production fluctuations, $s < \langle s \rangle_0$. Fluctuations that result in larger than average entropy production, $s \gtrsim \langle s \rangle_0$, for large v are more probable than predicted by the bound due to neglecting contributions from interparticle correlations. However, the relative error between the entropy production distribution and the bound decreases into the tails due to the increasingly independent particle behavior elaborated upon below.

For isolated AOUPs, the entropy production fluctuations are generically non-Gaussian and $\psi_f(\lambda) = D_\mathrm{r}\left(1 - \sqrt{1 - 2v^2\lambda(1+\lambda)/D_\mathrm{t}D_\mathrm{r}}\right)$, where D_r is the rotational diffusion constant (see Appendix C). The fluctuations in s are Gaussian only near the mean and are asymmetric [61]. This is in contrast to the Gaussian distribution that would be predicted by the thermodynamic uncertainty relations, and reflects the finite memory in the self-propulsion vector. The control force includes the same renormalized velocity as for ABPs, but in addition includes a force on the particle's orientation, $\boldsymbol{u}_\lambda = \psi_f(\lambda)\boldsymbol{b}$. In Fig. 1c, we see that the bound gives an accurate prediction of the fluctuations across the densities and D_a's considered. The fluctuations are still enhanced relative to the bound for $s < \langle s \rangle_0$, though less so than in Fig. 1b.

In order to understand the origins of the deviations from the bound and the connections to collective behavior in active matter, we consider in detail a system of ABPs at conditions near and far from its motility induced phase separation (MIPS) transition. Additionally, the asymmetry of entropy production fluctuations about its average, motivates us to consider separately
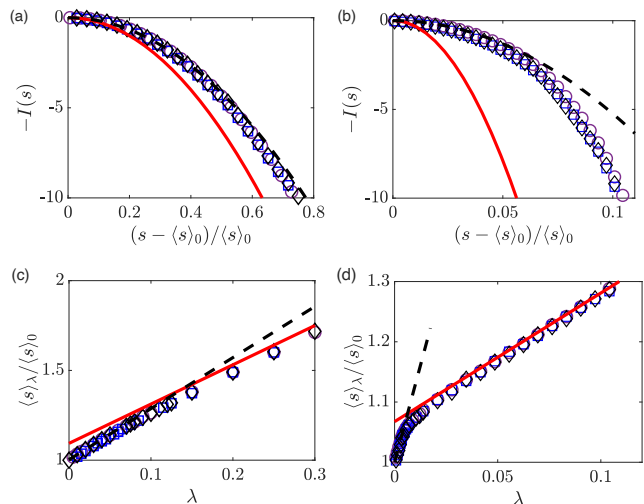

FIG. 2. Larger than average entropy production fluctuations for ABPs with $N = 10$ (purple circles), 20 (blue squares), and 40 (black diamonds). Distribution of entropy production for a) $v = 10$ and b) $v = 120$ with $\rho = 0.1$. In a) and b), the red lines are $I_b(s)$ and the dashed black lines are fits at $\lambda = 0$ to extract the second cumulant. The average entropy production at finite λ for c) $v = 10$ and d) $v = 120$ with $\rho = 0.1$. The dashed lines are from the cumulant fits in a) and b), and the red line is the non-interacting rate function.

fluctuations of $s > \langle s \rangle_0$ and $s < \langle s \rangle_0$. In Fig. 2, the distributions for $s > \langle s \rangle_0$ are shown for $v = 10$ and $v = 120$, for a variety of system sizes at fixed density, $\rho = 0.1$. While the probability is larger than predicted by the bound, it can be perturbatively corrected. Specifically, we can expand Eq. 5 up to the second cumulant, $\psi(\lambda) \approx \psi_f(\lambda) + \left(\lambda \langle \Delta W \rangle_0 + \lambda^2 \langle \delta \Delta W^2 \rangle_0 /2\right)/Nt$. The result of this approximation to the rate function is shown in Figs. 2 a) and b). For $v = 10$ the fluctuations are well described by the cumulant approximation, while for $v = 120$ asymptotic entropy production fluctuations are narrower than predicted.

The asymptotic behavior for $s \gg \langle s \rangle_0$ is well described by free particle motion for all v's. This can be seen by considering $d\psi/d\lambda = \langle s \rangle_\lambda$ from $\langle s \rangle_\lambda = \int ds\, s \exp\{Nt[-I(s) + \lambda s - \psi(\lambda)]\}$, which is a direct probe of the tails of $I(s)$. As shown in Figs. 2 c) and d), for both large and small v, $\langle s \rangle_\lambda$ exhibits a crossover from Gaussian statistics. Near $\lambda = 0$, $\langle s \rangle_\lambda$ varies linearly with λ with a slope given by the variance $\langle \delta s^2 \rangle_0$. For $\lambda \gg 0$, $\langle s \rangle_\lambda$ varies linearly with λ with a slope given by the free particle variance. An analogous crossover has been noted in the current statistics of an interacting tagged ABP [68]. The asymptotic free behavior implies that the most likely way for the system to produce large amounts of entropy is to suppress density correlations and decrease ΔW. This behavior results from the system adopting a net orientation for the particles' self-propulsion vector [60, 72]. If the net orientation persists in the thermodynamic limit, it would represent a spontaneous symmetry

breaking.

Fluctuations for $s < \langle s \rangle_0$ are much larger than predicted by the bound and are collective in origin. Fig. 3 shows the distributions of entropy production and $\langle s \rangle_\lambda$ for $v = 10$ and $v = 120$ at $\rho = 0.1$ for 3 system sizes. The distributions in Figs. 3 a) and b) show significant finite size effects for $s < \langle s \rangle_0$. In Figs. 3 c), and d), this is evident by a transition between two types of behavior that sharpens with increasing N and occurs at larger λ with increasing v over the limited range of system sizes we can study numerically. These features are a hallmark of a dynamical phase transition, in this case between a dilute phase and a phase separated state reminiscent of MIPS [73–75]. As has been found previously [60], this shows that the most likely way for the system to produce little entropy is to condense, decreasing the particles' displacement by increasing the effective drag. We find we can describe $I(s)$ by explicitly assuming that each dynamical phase is well approximated by a Gaussian distribution. Specifically, assuming $\psi_i(\lambda)Nt = \lambda\langle\Delta S\rangle_i + \lambda^2\langle\delta\Delta S^2\rangle_i/2$ for $i = d, c$ being the dilute and condensed phases, the rate function can be computed from a contraction principle [51] for the CGF, $\psi(\lambda) = \max_\lambda[\psi_c(\lambda), \psi_d(\lambda)]$ (see Appendix A). The result is a Maxwell construction and is shown in Fig. 3 to be a good approximation in the infinite system size limit. Due to the exponential difficulty of sampling large deviations in interacting systems we are unable to study larger systems [76]. Effects from the relatively large persistent length for $v = 120$ may complicate the extrapolation of these finite size effects to larger systems.

For $s < \langle s \rangle_0$, it is not sufficient to perturbatively correct the bound even for $v = 10$, which is far from the MIPS transition. To understand this behavior we have developed a coarse-grained theory. We define a fluctuating density field as $\rho(\boldsymbol{r}, t) = \sum_{i=1}^{N} \delta[\boldsymbol{r} - \boldsymbol{r}_i(t)]$. With this field, ΔW can be computed by assuming that the collisions are concentrated directly in front of a tagged particle. Under that assumption, ΔW can be written in terms of $\rho(\boldsymbol{r}, t)$, $\Delta W \approx -\beta v \int dt \int d\boldsymbol{r}d\boldsymbol{r}' \rho(\boldsymbol{r}, t)F(|\boldsymbol{r} - \boldsymbol{r}'|)\rho(\boldsymbol{r}', t)/2$, which is a convolution of two points of the density field with the interparticle force. For simplicity we have assumed that $F(0) = 0$. Further assuming that the force can be Fourier transformed, we find

$$\Delta W = -\frac{\beta v}{2} \int dt \int dk \, |\hat{\rho}(k, t)|^2 \hat{F}(k) \,, \qquad (9)$$

where $\hat{\rho}(k, t)$ is the Fourier transformed isotropic density field and $\hat{F}(k)$ the Fourier transformed force.

In order to evaluate the statistics of ΔW, we require an evolution equation for $\hat{\rho}(k, t)$. From the equation of motion for the position and orientation of each particle in the presence of the single particle control force, standard techniques afford an exact equation of motion for $\rho(\boldsymbol{r}, t)$ [77]. Its solution is complicated by its non-locality and coupling to a polarization field arising from the dynamics of the particle's orientation (see Appendix E). Rather than deal with it directly, assuming the system is macro-
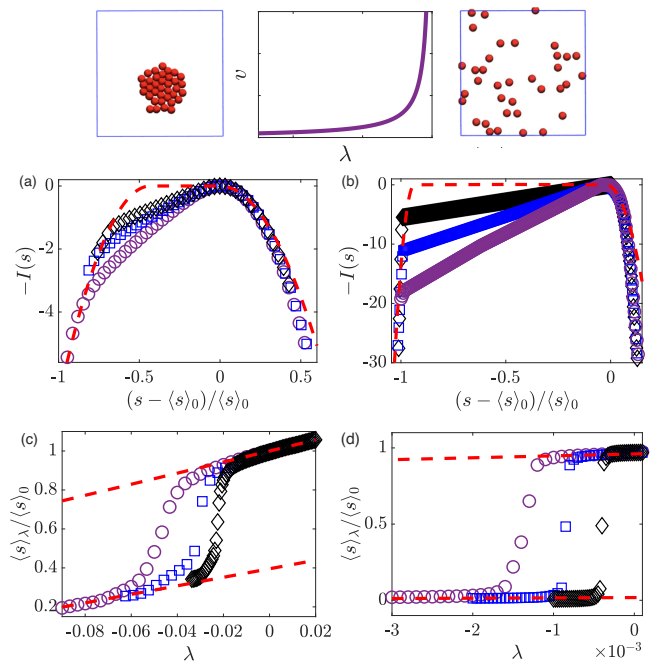

FIG. 3. Smaller than average entropy production fluctuations and dynamical phase transition for ABPs for different system sizes $N = 10$ (purple circles), 20 (blue squares), and 40 (black diamonds). The phase diagram and example structures are illustrated at the top with phase separation on the left of the phase diagram and a homogeneous state on the right. Distribution of entropy production for a) $v = 10$ and b) $v = 120$ with $\rho = 0.1$. In a) and b), the dashed red lines are a Maxwell construction for the dynamical phases. The average entropy production at finite λ for c) $v = 10$ and d) $v = 120$ with $\rho = 0.1$. The dashed red lines are from the Gaussian fits in a) and b) used in the Maxwell construction.

scopically homogeneous on the largest scales, we expect the density field to evolve diffusively. Thus, in the limit that $k \to 0$, the stochastic equation of motion for $\hat{\rho}(k, t)$ takes the form,

$$\frac{\partial \hat{\rho}(k, t)}{\partial t} \approx -k^2 \mathcal{D}_\lambda \hat{\rho}(k, t) + \sqrt{2\Delta_\lambda k^2} \hat{\eta}_\rho \,, \qquad (10)$$

where $\mathcal{D}_\lambda$ is the effective diffusion constant, Δ_λ is the effective mobility, and $\hat{\eta}_\rho$ is a complex noise [67, 74, 78–80]. Assuming that the polarization field relaxes quickly, and linearizing around a homogeneous density, these parameters can be derived explicitly for each active matter model.

Equation 10 has the form of an independent Ornstein-Uhlenbeck process for each Fourier mode of the density [80]. The CGF for ΔW can be solved exactly within this approximate linearized dynamics. Defining $\Delta\psi = \psi - \psi_f$,

$$\Delta\psi(\lambda) \approx \frac{1}{N} \sum_{k>0} k^2 \mathcal{D}_\lambda \left[1 - \sqrt{1 + \frac{\beta\lambda v_\lambda \hat{F}(k)\Delta_\lambda}{\mathcal{D}_\lambda^2 k^2}} \right] \,, \qquad (11)$$

we get an approximate correction to the bound in Eq. (7) due to interparticle correlations. This correction is valid for all positive λ, but becomes unstable at a critical value $\lambda_c \leq 0$ reflecting the breakdown in the linearized evolution equation for $\hat{\rho}(k,t)$. For a finite system with largest wavevector $k = 2\pi\sqrt{\rho/N}$, the location of the instability is found by setting the discriminant to zero, $\lambda_c \approx -4\pi^2\mathcal{D}_0^2\rho/\beta v \hat{F}(0)\Delta_0 N$ where for the short ranged forces considered, we can approximate the force as $\hat{F}(0)$ and we can neglect the λ dependence in $v_\lambda, \mathcal{D}_\lambda$, and Δ_λ. This instability signals the dynamical phase transition that occurs at $\lambda_c = 0^-$ in the thermodynamic limit and whose influence on the dynamics of active matter increases with v, and with increasing proximity to MIPS, consistent with the results in Fig. 3. In a phase separated state, ΔW is a large negative number which counteracts the free particle contribution and reduces the entropy production.

The origin of phase separation can be understood by noting that the optimal control potential which gives rise to rare entropy fluctuations is, for large interparticle separations $r/\sigma \gg 0$ and in the limit λ approaches zero (see Appendix E),

$$\hat{V}_\lambda \approx -\frac{\beta v\lambda}{N\mathcal{D}_0}\sum_{k>0}\frac{|\hat{\rho}(k,t)|^2}{2k^2}\hat{F}(k) \ . \tag{12}$$

The inverse Fourier transform will involve a convolution between the WCA force and $1/k^2$ which gives rise to a Bessel Function. Since the WCA potential quickly decays, the long range contribution in real space is a logrithmic potential,

$$V_\lambda(r) \approx -\frac{\beta\lambda v}{\mathcal{D}_0}\ln r/2 \ , \tag{13}$$

which is attractive for $\lambda < 0$ with a magnitude that depends on v and the control force is $\boldsymbol{u}_\lambda \approx (\beta\lambda v/\mathcal{D}_0)\nabla\ln r/2$. For negative enough λ or large enough v, this force will give rise to phase separation. This optimal control force is similar to other passive models near diffusive instabilities [80–82].

The long-ranged effective force demonstrates how effective attractions are introduced by self-propulsion in order to minimize the entropy production. This force is unique and encodes the way in which self-propelled particles interact provided the condition of obtaining a lower than average value of the entropy production. As such, it provides a sharp relationship between entropy production and emergent collective behavior in active matter. Correlations between entropy production and motility induced phase separation have been observed previously at the level of the mean behavior [1, 83, 84], however this work codifies that relationship on the level of fluctuations.

For both MIPS and the dynamical transition we discuss, phase separation is the result of a diffusive instability where density accumulates due to unbalanced fluxes made possible by the system being kept from thermal equilibrium. We have shown such collective behavior results from the reduction of entropy production and enhancement of density correlations. Large entropy production by contrast, arises through the suppression of density correlations. Thus, our results show how the structure of entropy production fluctuations are intimately connected to long-ranged correlations in active matter. We expect that deviations from the bound derived here can serve as a guide to identify criticality and novel phases of active matter generally.

Appendix A: Simulation Details

1. Model definitions and parameters

a. ABPs and RTPs

For both ABPs and RTPs the orientation vector has a fixed magnitude, so in two dimensions it can be uniquely parameterized by an angle θ. For the ith particle, $\boldsymbol{b_i} = \{\cos(\theta_i)\hat{\boldsymbol{x}}_i, \sin(\theta_i)\hat{\boldsymbol{y}}_i\}$, where $\hat{\boldsymbol{x}}_i$ and $\hat{\boldsymbol{y}}_i$ are the unit vectors in the x and y directions, respectively. For ABPs, the dynamics of θ_i are Brownian,

$$\dot{\theta}_i(t) = \eta_i^\theta(t) \tag{A1}$$

where η_i^θ is a Gaussian white noise, satisfying $\langle\eta_i^\theta(t)\rangle = 0$ and $\langle\eta_i^\theta(t)\eta_j^\theta(t')\rangle = 2D_r\delta_{ij}\delta(t-t')$ with D_r the rotational diffusion constant. We take $D_r = 3D_t/\sigma^2$ throughout.

The dynamics of θ for RTPs are piecewise constant over waiting times, τ, satisfying a Poisson process[85, 86]. The waiting time distribution is given by an exponential distribution,

$$P(\tau) = \gamma e^{-\gamma\tau}, \tag{A2}$$

with constant reorientation rate γ. We take $\gamma = D_t/\sigma^2$. At each τ, the particles reorient by drawing a new θ chosen uniformly over the range $[0, 2\pi]$.

b. ADPs

Each ADP is composed of two particles that are tethered together by a harmonic bond. The harmonic bond

potential is given by $U_{\mathrm{H}}(r) = k(r - l)^2/2$, where k is the spring constant, l is the rest length, and r is the displacement between the two bonded particles. We take $k = 100 \; \epsilon/\sigma^2$ and $l = 1.5 \; \sigma$. The self-propulsion direction is along the bond vector. For the ith ADP, composed of monomers 1 and 2, $\boldsymbol{b}_i = \hat{\boldsymbol{r}}_{i,12}$ where $\hat{\boldsymbol{r}}_{i,12}$ is the unit displacement vector between monomers 1 and 2. The time evolution of the orientation vector is given by the time evolution of the displacement vector between the two composite particles as dictated by their individual equations of motion [87, 88].

c. AOUPs

For AOUPs, the self propulsion vector changes both its magnitude and direction. Its equation of motion takes the form of an Ornstein-Uhlenbeck process and given by

$$\dot{\boldsymbol{b}_i} = -D_{\mathrm{r}}\boldsymbol{b}_i + \boldsymbol{\xi_i} \tag{A3}$$

where $\boldsymbol{\xi_i}$ is a Gaussian random variable satisfying $\langle \xi_{i,\alpha}(t) \rangle = 0$ and $\langle \xi_i^\alpha(t)\xi_j^\beta(t') \rangle = D_{\mathrm{r}}\delta_{ij}\delta_{\alpha,\beta}\delta(t - t')$ for each α, β component [9, 48].

2. Bound and Cloning calculation details

For all simulations we used $N =$ 10, 20, or 40 particles. A 2-dimensional square box of length L with periodic bounds was used and the length chosen to give the desired density through the equation $L = \sqrt{N/\rho}$. The equations of motion are discretized using a first order Euler method. Calculations of the rate functions, $I(s)$, require enhanced sampling techniques in order to probe rare fluctuations. For this we use the cloning algorithm[69]. Cloning results were run with 2.4×10^4-1.5×10^7 walkers. The cloning parameters varied for the models considered. For ABPs, RTPs, and ADPs, we used a time step of $\delta t = 10^{-3} - 10^{-5}$, depending on the v, a branching time of $t_{\mathrm{int}} = 50\delta t$, and an observation time of $t = 10t_{\mathrm{int}}$. For AOUPs, we used a timestep of $\delta t = 10^{-4}$, a branching time of $t_{\mathrm{int}} = 10\delta t$, and an observation time $t = 30t_{\mathrm{int}}$.

Each estimate for the CGF at a specific lambda is the mean from 3 runs. They were checked for convergence in walker number and time [76, 89]. For the simulations of ABPs in Fig. 3, we used cloning with guiding forces to accelerate convergence of the estimate [60, 70]. This was done by adding the non-interacting control force $\boldsymbol{u}_\lambda$ to the equations of motion and using cloning with the weight ΔW (see Eq. 5). The full CGF was then obtained by adding back the non-interacting CGF, $\psi_f(\lambda)$.

In Fig. 4, we show the convergence in walkers (N_w) for negative λ. The critical λ is close to -0.0015. In the limit that the walkers go to infinity the hysteresis seen in the dip around the critical λ will disappear. In Figs. 5, 6, 7 we show the convergence as a function of walker number

for $\lambda = -6.94 \times 10^{-4}$, $\lambda = -0.0021$, and $\lambda = -0.0028$ which are before, close to, and after the critical point. The CGF estimate was easily converged for positive λ for 1.2×10^4 walkers which is consistent with [60].

To compute $I_b(s)$, we require a numerical estimate of $\langle \Delta W \rangle_{\boldsymbol{u}_\lambda}$, which was computed for all systems with $N = 40$ particles for an observation time of $t = 10^6\delta t$. The observation time and particle number were increased until convergence of the running average was obtained. The codes used to generate the data in this paper can be found at: **https://github.com/kklymko/active_work**

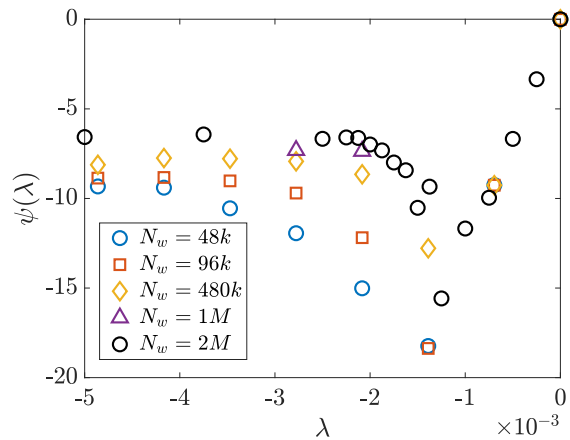
FIG. 4. The convergence of the estimate of the CGF for $N = 40$, $v = 120$ for $t = 500 \Delta t$ for different numbers of walkers N_w.
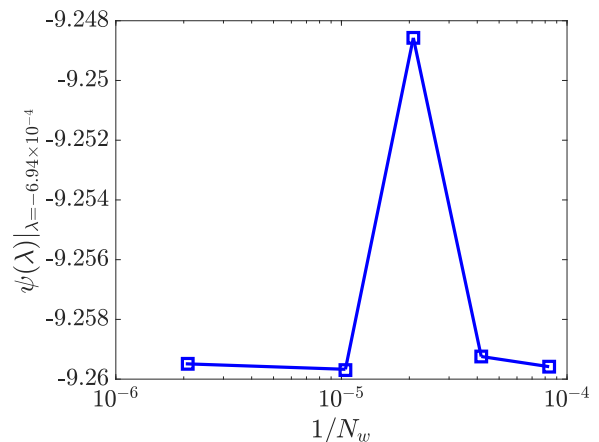

FIG. 5. Convergence of the CGF estimate for $v = 120$, $N = 40$, $t = 500\Delta t$, $\lambda = -6.94 \times 10^{-4}$, which is right before the phase transition. In theory, the CGF estimate is converged at zero on the y-axis.
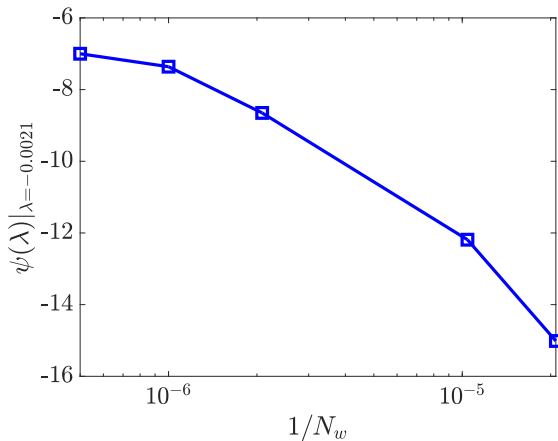
FIG. 6. Convergence of the CGF estimate for $v = 120$, $N = 40$, $t = 500\Delta t$, $\lambda = -0.0021$, which is close to λ_c. In theory, the CGF estimate is converged at zero on the y-axis.

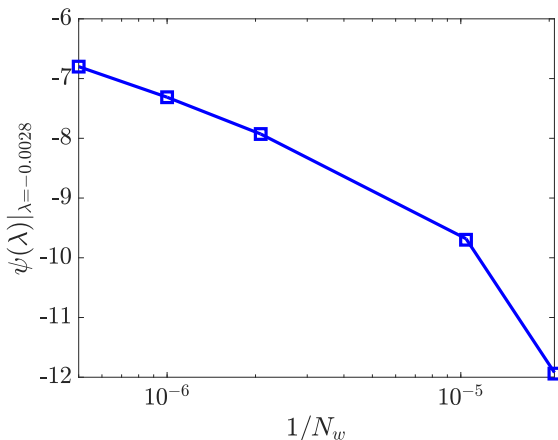

FIG. 7. Convergence of the CGF estimate for $v = 120$, $N = 40$, $t = 500\Delta t$, $\lambda = -0.0028$, which is after the phase transition. In theory, the CGF estimate is converged at zero on the y-axis.

3. Gaussian fits for Figs. 2 and 3

In Fig. 2, the Gaussian fits for small fluctuations for $v = 10$ and $v = 120$ is $\langle s \rangle_0 = v^2(1-0.84\rho)$, $\langle (\delta s)^2 \rangle = 3v^2$, and $\langle s \rangle_0 = v^2(1-0.63\rho)$, and $\langle (\delta s)^2 \rangle = 12v^2$ with $\rho = 0.1$ which are represented by black dotted lines. There is not a clear size dependence for the system sizes studied here and we have found that all three system sizes considered have the same best fit.

In Fig. 3, we fit the dense phase in a similar way. Although the transition has a system size dependence, once the system is within the phase separated state there is not a clear system size dependence in the variance. The Gaussian fit for the phase separated peak in Figure 3a) and 3b) is given by $\langle s \rangle_0 = v^2(1-0.84\rho)$, $\langle (\delta s)^2 \rangle = 2v^2$ for v=10 with $\rho = 0.58$, and $\langle s \rangle_0 = v^2(1 - 0.84\rho)$, $\langle (\delta s)^2 \rangle = 2v^2$ for v=120 with $\rho = 1.12$. Note that the variance

for both v's considered for the phase separated system is given by the non-interacting CGF. The averages used in the Maxwell construction and those in Fig. 3c), and d) are slightly different due to the shift in the mean in the thermodynamic limit given by $\langle s \rangle_{\lambda_c} = \langle s \rangle_0 + \lambda_c \langle (\delta s)^2 \rangle_0$ but the slopes are identical.

Appendix B: General form for the Entropy production

In order to derive the entropy production for each model, we assume that the self-propulsion is even under time reversal. The difference between choosing the self-propulsion to be even under time reversal is that there is a non-interacting term, as shown in Ref. [65]. All of the collective phenomena are thus going to be independent of the convention. The distinction between choice of sign is described in more detail in Ref. [61]. We note that the convention used in this manuscript is consistent with Ref. [59, 60, 62, 63]. We also note that the form of the active work is closely related to the swim pressure described in the literature [90, 91] and can be derived independently from mechanical considerations.

We take the standard definition of the entropy production based on the path probability and its time reversal,

$$\Delta S = \ln P[\Gamma]/P[\tilde{\Gamma}], \tag{B1}$$

$P[\Gamma]$ is the probability of observing a path denoted $\Gamma = (r^N(t), b^N(t))$, and $\tilde{\Gamma} = (\tilde{r}^N(t), \tilde{b}^N(t))$ is the time reserved path. In the time reversed path, we change the signs of functions with explicit time dependence, $\dot{\tilde{r}}_i(-t) = -\dot{r}_i(t)$ and $\dot{\tilde{b}}_i(-t) = -\dot{b}_i(t)$. In the subsequent sections, we write out $P[\Gamma]$ for ABPs and AOUPs and their corresponding entropy production. The ADPs and RTPs can be derived analogously. It is found that all models considered have the same form of ΔS in the long time limit.

a. ABPs

The probability of observing a path for a system of ABPs with conservative interactions in the Stratonovich convention is

$$P[\Gamma] \propto \exp\left[-\sum_{i=1}^{N} \int_0^t dt' \frac{\left(\dot{r}_i - v\,b_i - \mu F_i\left(r^N\right) \right)^2}{4D_t} \right.$$
$$\left. + \frac{\nabla_{r_i} \cdot \left(\mu F_i\left(r^N\right) \right)}{2} + \frac{\dot{b}_i^2}{4D_r} \right], \quad \text{(B2)}$$

where the gradient term in the second line follows from the Stratonovich convention. After performing the time

reversal operation and taking a ratio of path probabilities, the entropy production then becomes

$$\Delta S = \frac{1}{D_\mathrm{t}} \sum_{i=1}^{N} \int_0^t dt' \left[v\, \boldsymbol{b}_i \circ \dot{\mathbf{r}}_i(t') + \dot{\mathbf{r}}_i \circ \mu \boldsymbol{F}_i\left(\boldsymbol{r}^N\right) \right] \; , \tag{B3}$$

which is a sum of two terms. However, since we are using the Stratonovich convention the chain rule is preserved and the term

$$\sum_{i=1}^{N} \int_0^t dt' \dot{\mathbf{r}}_i \circ \boldsymbol{F}_i\left(\boldsymbol{r}^N\right) = U(\boldsymbol{r^N}(0)) - U(\boldsymbol{r^N}(t)) \;, \tag{B4}$$

does not grow with time, unlike the first term. In the long time limit it will become negligible, and can be neglected in the entropy production.

b. AOUPs

For AOUPS using the Stratonovich convention, the derivation of the form of the entropy production follows similarly as for the other models. Specifically, the path probability is

$$P[\Gamma] \propto \exp\left[-\sum_{i=1}^{N} \int_0^t dt' \frac{\left(\dot{\boldsymbol{r}}_i - v\,\boldsymbol{b}_i - \mu \boldsymbol{F}_i\left(\boldsymbol{r}^N\right) \right)^2}{4D_\mathrm{t}} \right.$$
$$+ \frac{\boldsymbol{\nabla}_{\boldsymbol{r}_i} \cdot \left(\mu \boldsymbol{F}_i\left(\boldsymbol{r}^N\right) \right)}{2}$$
$$\left. + \frac{\left(\dot{\boldsymbol{b}}_i + D_\mathrm{r}\boldsymbol{b}_i \right)^2}{2D_\mathrm{r}} - \frac{\boldsymbol{\nabla}_{\boldsymbol{b}_i} \cdot D_\mathrm{r}\boldsymbol{b}_i}{2} \right] , \tag{B5}$$

where the additional force on $\boldsymbol{b}_i$ results in the last two terms. After performing the time reversal operation, the entropy production is

$$\Delta S = \sum_{i=1}^{N} \int_0^t dt' \left(\frac{v\,\boldsymbol{b}_i \circ \dot{\boldsymbol{r}}_i}{D_\mathrm{t}} + \frac{\dot{\boldsymbol{r}}_i \circ \mu \boldsymbol{F}_i\left(\boldsymbol{r}^N\right)}{D_\mathrm{t}} \right.$$
$$\left. + 2\dot{\boldsymbol{b}}_i \circ \boldsymbol{b}_i \right) , \tag{B6}$$

where the first two terms are analogous to the ABPs. Both the second term and third term do not grow with time, and so in the long time limit the entropy production reduces to

$$\Delta S = \frac{v}{D_\mathrm{t}} \sum_{i=1}^{N} \int_0^t dt' \; \boldsymbol{b}_i \circ \dot{\boldsymbol{r}}_i \;, \tag{B7}$$

which is equivalent to the form found for the ABPs.

Appendix C: Free particle CGFs and optimal control forces

The free particle CGF is computable from the solution of a generalized eigenvalue equation of the form

$$L_\lambda \nu_\lambda = \psi_f(\lambda)\nu_\lambda \;, \tag{C1}$$

where L_λ is the Lebowitz-Spohn, or tilted, operator and ν_λ and $\psi_f(\lambda)$ are the maximum eigenvector eigenvalue pair. The tilted operator is derivable from the time evolution of the CGF and the relation to the spectrum of L_λ and the CGF follows from the long time limit. Generically, for a current-type variable [51, 54] the optimal control force that realizes rare entropy production fluctuation is given by

$$\boldsymbol{u}_\lambda = 2\lambda v \boldsymbol{b} + 2\boldsymbol{D} \cdot \nabla \ln \nu_\lambda \;, \tag{C2}$$

where $\boldsymbol{D}$ is a matrix of diffusion constants in define in space crossed with the self-propulsion vector dimension and $\nabla = \{\nabla_{\boldsymbol{r}}, \nabla_{\boldsymbol{b}}\}$. The optimal control force is encoded in the maximum eigenvector associated with L_λ [51].

In order to fully solve the eigenspectrum it is necessary to solve the eigenvalue problem for the adjoint tilted operator [92]

$$L_\lambda^\dagger q_\lambda = \psi_f(\lambda)q_\lambda \;. \tag{C3}$$

since in general L_λ is not Hermitian. The boundary conditions of the eigevectors must obey a normalization boundary condition $\nu_\lambda(\boldsymbol{b})q_\lambda(\boldsymbol{b}) \to 0$ as $b \to \infty$ [92]. The boundary condition can equivalently be written as

$$\int d\boldsymbol{b} \; q_\lambda(\boldsymbol{b})\nu_\lambda(\boldsymbol{b}) = 1, \tag{C4}$$

and for convenience we impose that

$$\int d\boldsymbol{b} \; q_\lambda(\boldsymbol{b}) = 1. \tag{C5}$$

a. ABPs

The tilted generator for the entropy production of an isolated ABP is

$$L_\lambda = v\, \boldsymbol{b} \cdot \left[\nabla_{\boldsymbol{r}} + \lambda \frac{v\,\boldsymbol{b}}{D_\mathrm{t}} \right] \tag{C6}$$
$$+ D_\mathrm{t} \left[\nabla_{\boldsymbol{r}} + \lambda \frac{v\,\boldsymbol{b}}{D_\mathrm{t}} \right] \cdot \left[\nabla_{\boldsymbol{r}} + \lambda \frac{v\,\boldsymbol{b}}{D_\mathrm{t}} \right] + \boldsymbol{\nabla}_{\boldsymbol{b}}^2 D_\mathrm{r} \;,$$

which can be solved on a periodic domain by a constant eigenvector, $\nu_\lambda = \mathrm{const}$. This is equivalent to assuming that the stationary state is uniform and isotropic for all λ. The CGF follows by noting $\boldsymbol{b} \cdot \boldsymbol{b} = 1$ and is

$$\psi_f(\lambda) = \lambda \frac{v^2}{D_\mathrm{t}} + \lambda^2 \frac{v^2}{D_\mathrm{t}} \;, \tag{C7}$$

and that the control force that realizes the rare dynamics reduces to $\boldsymbol{u}_\lambda = 2\lambda v\,\boldsymbol{b}$. The corresponding equation of motion is

$$\dot{\boldsymbol{r}}_i = v\,(1+2\lambda)\,\boldsymbol{b}_i + \sqrt{2D_\mathrm{t}}\boldsymbol{\eta}_i \ , \qquad (C8)$$

where we see explicitly that the control force acts to renormalize the self-propulsion velocity.

b. AOUPs

The tilted generator for the entropy production of an isolated AOUP is

$$L_\lambda = v\,\boldsymbol{b}\cdot\left[\nabla_{\boldsymbol{r}} + \lambda\frac{v\,\boldsymbol{b}}{D_\mathrm{t}}\right] + D_\mathrm{t}\left[\nabla_{\boldsymbol{r}} + \lambda\frac{v\,\boldsymbol{b}}{D_\mathrm{t}}\right]\cdot\left[\nabla_{\boldsymbol{r}} + \lambda\frac{v\,\boldsymbol{b}}{D_\mathrm{t}}\right]$$
$$+ \frac{D_\mathrm{r}}{2}\nabla_{\boldsymbol{b}}^2 - D_\mathrm{r}\boldsymbol{b}\cdot\nabla_{\boldsymbol{b}} \ , \quad (C9)$$

which contains an additional convective term in $\boldsymbol{b}$ due to the constant restoring force. Assuming the system maintains a uniform and isotropic state at all λ, such that the eigenvector does not depend on $\boldsymbol{r}$, we can simplify the tilted operator,

$$L_\lambda = \lambda\frac{v^2|\boldsymbol{b}|^2}{D_\mathrm{t}} + \lambda^2\frac{v^2|\boldsymbol{b}|^2}{D_\mathrm{t}} + \frac{D_\mathrm{r}}{2}\nabla_{\boldsymbol{b}}^2 - D_\mathrm{r}\boldsymbol{b}\cdot\nabla_{\boldsymbol{b}} \ , \quad (C10)$$

where b is the magnitude of the vector $\boldsymbol{b}$. The domain of b is from 0 to ∞, the eigenvector from equation (C1) is

$$\nu_\lambda(\boldsymbol{b}) = \exp\left(\frac{|\boldsymbol{b}|^2\psi_f(\lambda)}{2D_\mathrm{r}}\right) \ , \qquad (C11)$$

and its corresponding eigenvalue is

$$\psi_f(\lambda) = D_\mathrm{r}\left(1 - \sqrt{1 - \frac{2v^2}{D_\mathrm{r}D_\mathrm{t}}\lambda(1+\lambda)}\right) \ , \qquad (C12)$$

which can be verified by inserting $\nu(\boldsymbol{b})$ back into Eq. C10 and noting that since it's in two dimensions it is split up into the x and y dimensions with $|\boldsymbol{b}|^2 = \boldsymbol{b}\cdot\boldsymbol{b} = b_x^2 + b_y^2$ and $\nabla_{\boldsymbol{b}}^2 = \nabla_{b_x}^2 + \nabla_{b_y}^2$. The left eigenvector can also be solved to obtain the normalization constant but it is not needed for the control force calculations.

The optimal control force in the $\boldsymbol{r}$ and $\boldsymbol{b}$ directions, $\boldsymbol{u}_\lambda = \{\boldsymbol{u}_\lambda^{\boldsymbol{r}}, \boldsymbol{u}_\lambda^{\boldsymbol{b}}\}$ are

$$\boldsymbol{u}_\lambda = \{2\lambda v\,\boldsymbol{b}, \boldsymbol{b}\psi_f(\lambda)\} \ , \qquad (C13)$$

which is the result for the control force for non-interacting AOUPs. The biased equations of motion become

$$\dot{\boldsymbol{r}}_i = v\,(1+2\lambda)\,\boldsymbol{b}_i + \sqrt{2D_\mathrm{t}}\boldsymbol{\eta}_i \ , \qquad (C14)$$

and

$$\dot{\boldsymbol{b}}_i = -D_\mathrm{r}\boldsymbol{b}_i\,(1 - \psi_f(\lambda)/D_\mathrm{r}) + \sqrt{D_\mathrm{r}}\boldsymbol{\xi}_i \ , \qquad (C15)$$

where the former is identical for ABPs and the latter is specific to AOUPs.

Appendix D: Entropy bounds from Girsanov transformation

The CGF for the entropy production can be rewritten as an average over the biased ensemble by preforming a change of measure, or Girsanov transformation, from the original path ensemble with probability $P[\Gamma]$,

$$\psi(\lambda) = \frac{1}{tN}\ln\int\mathcal{D}[\Gamma]P[\Gamma]e^{\lambda\Delta S}$$
$$= \frac{1}{tN}\ln\int\mathcal{D}[\Gamma]\frac{P[\Gamma]}{P_{\boldsymbol{u}_\lambda}[\Gamma]}P_{\boldsymbol{u}_\lambda}[\Gamma]e^{\lambda\Delta S}$$
$$= \frac{1}{tN}\ln\left\langle\frac{P[\Gamma]}{P_{\boldsymbol{u}_\lambda}[\Gamma]}e^{\lambda\Delta S}\right\rangle_{\boldsymbol{u}_\lambda} \ , \qquad (D1)$$

where $P_{\boldsymbol{u}_\lambda}[\Gamma]$ denotes a path ensemble with an additional force $\boldsymbol{u}_\lambda$ added to the original equations of motion, and $\langle\ldots\rangle_{\boldsymbol{u}_\lambda}$ denote ensemble average with respect to that measure. Using Jensen's inequality, we find a general bound within an arbitrary control ensemble [93],

$$\psi(\lambda) \geq \frac{1}{tN}\left(\lambda\langle\Delta S\rangle_{\boldsymbol{u}_\lambda} + \left\langle\ln\frac{P[\Gamma]}{P_{\boldsymbol{u}_\lambda}[\Gamma]}\right\rangle_{\boldsymbol{u}_\lambda}\right) \ , \qquad (D2)$$

which need not be tight. However, below we show how in the systems studied by choosing $\boldsymbol{u}_\lambda$ to be the optimal control force for the free particle, we can arrive at the tight bound on the entropy production.

a. ABPs

The relative actions with and without the single particle control force for a system of interacting ABPs is

$$\ln\frac{P[\Gamma]}{P_{\boldsymbol{u}_\lambda}[\Gamma]} = \sum_{i=1}^{N}\int_0^t dt'\,\frac{v^2\lambda(1+\lambda)}{D_\mathrm{t}} - \frac{v}{D_\mathrm{t}}\lambda\boldsymbol{b}_i\circ\dot{\boldsymbol{r}}_i$$
$$+ \beta v\lambda\boldsymbol{b}_i\cdot\boldsymbol{F}_i\left(\boldsymbol{r}^N\right) \ , \quad (D3)$$

which employs the identity $\boldsymbol{b}\cdot\boldsymbol{b} = 1$. We recognize the first term on the right hand side as $\psi_f(\lambda)$, the second term as the negative of the entropy production, and the final term as ΔW. Inserting this relative action into Eq. D1, we note that the entropy production terms cancel, and we can pull the constants out of the average. The bound can be shown to work analogously for the ADP and RTP models since the added control force does not change the statistics of the orientation, $\boldsymbol{b}$, and only changes the positional degrees of freedom.

b. AOUPs

The relative actions with and without the single particle control force for a system of interacting AOUPs is

$$\ln \frac{P[\Gamma]}{P_{\boldsymbol{u}_\lambda}[\Gamma]} =$$
$$\int_0^t dt' \left(\frac{v^2\lambda(1+\lambda)}{D_t} - \psi_f(1 - \psi_f/2D_r) \right) b^2$$
$$+ \psi_f(\lambda) - \frac{v}{D_t}\lambda \boldsymbol{b}_i \circ \dot{\boldsymbol{r}}_i + v\beta\lambda \boldsymbol{b_i} \cdot \boldsymbol{F}_i\left(\boldsymbol{r}^N\right) , \quad (D4)$$

which is more complicated than for the ABPs due to the fluctuating magnitude of the self-propulsion vector. We still can identify the same structure as before, with the free particle CGF, negative of the entropy production, and ΔW, however there is an additional first term in the parenthesis. Inserting the definition of ψ_f from Eq. C12 we find that the term proportional to v^2b^2 is identically 0. This leaves us with the bound for AOUPs.

Appendix E: Entropy production from coarse-grained density field

Here we elaborate on our coarse grained theory of the interacting term. Assuming that the important contributions to the interparticle entropy production come from forces that directly oppose self-propulsion, we approximate $\boldsymbol{b}_i \cdot \boldsymbol{F}(\boldsymbol{r}_{ij}) \approx -F(r_{ij})$ where $\boldsymbol{F}(\boldsymbol{r}_{ij})$ is the contribution of the ith particle's force due to particle j and $\boldsymbol{r}_{ij}$ is the displacement vector between particles i and j with magnitude r_{ij}. As presented, under this approximation the fluctuations of ΔW depend only on the time evolution of the density field. Below we first derive an approximate equation of motion for the density, in the limit of small k and small fluctuations from its mean. Then we describe the approximate calculation of the cumulant generating function and control force.

1. Equation of motion for the density

We are interested in the density fluctuations with the added control force which changes the self propulsion speed proportional to lambda as $v_\lambda = v(1 + 2\lambda)$. To arrive at an effective equation of motion for the density we first define the instantaneous density field as,

$$\rho(\boldsymbol{r}, t) = \sum_{i=1}^N \delta[\boldsymbol{r} - \boldsymbol{r}_i(t)] , \quad (E1)$$

and corresponding polarization field as

$$\boldsymbol{P}(\boldsymbol{r}, t) = \sum_{i=1}^N \delta[\boldsymbol{r} - \boldsymbol{r}_i(t)]\boldsymbol{b_i}(t) , \quad (E2)$$

where δ are Dirac's delta function. In principle, higher order multipoles in the orientation field are needed to completely describe the dynamics, however we neglect quadrupole and higher fields. For the homogeneous states considered, this has been shown to be a good approximation [94, 95]. Following the standard procedures [77, 85] a set of coupled stochastic equation of motion for both fields. For the density field,

$$\frac{\partial \rho(\boldsymbol{r}, t)}{\partial t} = -\nabla_{\boldsymbol{r}} \left[\mu \rho(\boldsymbol{r}, t) \int d\boldsymbol{r}' \boldsymbol{F}(\boldsymbol{r} - \boldsymbol{r}')\rho(\boldsymbol{r}', t) \right.$$
$$\left. + v_\lambda \boldsymbol{P}(\boldsymbol{r}, t) \right] + D_t \nabla_{\boldsymbol{r}}^2 \rho(\boldsymbol{r}, t) + \nabla_r \sqrt{2\Delta_\lambda}\boldsymbol{\eta}_\rho(\boldsymbol{r}, t) \quad (E3)$$

where $\Delta_\lambda = D_t \rho(\boldsymbol{r}, t)$ is the mobility and the noise obeys the statistics $\langle \boldsymbol{\eta}_\rho(\boldsymbol{r}, t) \rangle = 0$ and $\langle \eta_\rho^\alpha(\boldsymbol{r}, t)\eta_\rho^\beta(\boldsymbol{r}', t') \rangle = \delta_{\alpha,\beta}\delta(t - t')\delta(\boldsymbol{r} - \boldsymbol{r}')$. For the polarization field,

$$\frac{\partial \boldsymbol{P}(\boldsymbol{r}, t)}{\partial t} = -\nabla_{\boldsymbol{r}} \left[\mu \boldsymbol{P}(\boldsymbol{r}, t) \int d\boldsymbol{r}' F(\boldsymbol{r} - \boldsymbol{r}')\rho(\boldsymbol{r}', t) \right]$$
$$- \nabla_{\boldsymbol{r}} \frac{v_\lambda \rho(\boldsymbol{r}, t)}{2} + D_t \nabla_{\boldsymbol{r}}^2 \boldsymbol{P}(\boldsymbol{r}, t)$$
$$- D_r \boldsymbol{P}(\boldsymbol{r}, t) + \nabla_{\boldsymbol{r}} \sqrt{2\Lambda_P}\boldsymbol{\eta}_P(\boldsymbol{r}, t) \quad (E4)$$

where $\eta_P(\boldsymbol{r}, t)$ has the same noise statistics as $\boldsymbol{\eta}_\rho$ and $\Lambda_P = D_t \boldsymbol{P}(\boldsymbol{r}, t)$.

We assume there is a separation of time scales between the density field, which we assume to be slow, and the polarization field, which we assume to relax quickly. Further we assume that on the scale of density fluctuations, the polarization is constant and homogeneous [67]. These so-called adiabatic assumptions are standard in the treatment of instabilities in the ABP system. Under these assumptions, the polarization is stationary and can be averaged separately from the density and we can neglect its gradient terms. Rearranging the remaining terms, we have an explicit relation between the polarization and density fields,

$$P(\boldsymbol{r}, t) = -\frac{v_\lambda}{2D_r}\nabla_{\boldsymbol{r}}\rho(\boldsymbol{r}, t) , \quad (E5)$$

which effectively separates the evolution of the two fields. Inserting this into Eq. (E3) we arrive at a closed equation of motion for the density,

$$\frac{\partial \rho(\boldsymbol{r}, t)}{\partial t} = -\nabla_{\boldsymbol{r}} \left[\mu \rho(\boldsymbol{r}, t) \int d\boldsymbol{r}' \boldsymbol{F}(\boldsymbol{r} - \boldsymbol{r}')\rho(\boldsymbol{r}', t) \right]$$
$$+ \mathcal{D}_\lambda \nabla_{\boldsymbol{r}}^2 \rho(\boldsymbol{r}, t) + \nabla_{\boldsymbol{r}} \sqrt{2\Delta_\lambda}\boldsymbol{\eta}_\rho(\boldsymbol{r}, t) \quad (E6)$$

with $\mathcal{D}_\lambda = D_t + v_\lambda^2/2D_r$ as the effective diffusion constant.

While the equation is closed, it is still nonlinear due to the fluctuating convective term from the interparticle interactions. While more sophisticated expansions exist, for the low densities we consider, we can linearize the evolution equation by simply dropping the second order term in the density,

$$\frac{\partial \rho(\boldsymbol{r}, t)}{\partial t} = \mathcal{D}_\lambda \nabla_{\boldsymbol{r}}^2 \rho(\boldsymbol{r}, t) + \nabla_{\boldsymbol{r}} \sqrt{2\Delta_\lambda}\eta_\rho(\boldsymbol{r}, t) , \quad (E7)$$

which results in a standard fluctuating diffusion equation. Corrections due to interactions can be included phenomenologically by making $\mathcal{D}_\lambda$ and Δ_λ depend on the mean density.

Introducing the Fourier transforms, for the density

$$\hat{\rho}(\boldsymbol{k}, t) = \int d\boldsymbol{r} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \rho(\boldsymbol{r}, t) , \qquad (E8)$$

and the noise,

$$\hat{\boldsymbol{\eta}}(\boldsymbol{k}, t) = \int d\boldsymbol{r} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \boldsymbol{\eta}(\boldsymbol{r}, t) , \qquad (E9)$$

we can arrive at the equation of motion in Eq. 10.

2. CGF and optimal control force

The equation of motion for the Fourier transformed density takes the form of a set of uncoupled, complex Ornstein-Uhlenbeck processes for each wavevector. The large deviations of such a system for observables like ΔW have been considered in detail in Ref. 80. The tilted operator for which the CGF of ΔW is the largest eigenvalue and has the form

$$\mathcal{L}_\lambda = \sum_{k>0} -k^2 \mathcal{D}_\lambda \hat{\rho}_k \nabla_{\hat{\rho}_k} + k^2 \Delta_\lambda \nabla_{\hat{\rho}_k}^2 - \lambda \frac{\beta v_\lambda}{2} \hat{F}(k) |\hat{\rho}_k|^2 , \qquad (E10)$$

which has to be solved for both the real and imaginary parts of $\hat{\rho}_k$. This can be done following the method of Ref. 80. The resulting CGF is

$$\Delta\psi(\lambda) = \frac{1}{N} \sum_{k>0} \Delta\psi_k(\lambda) , \qquad (E11)$$

where for each k,

$$\Delta\psi_k(\lambda) = k^2 \mathcal{D}_\lambda \left[1 - \sqrt{1 + \frac{\beta\lambda v_\lambda \hat{F}(k)\Delta_\lambda}{\mathcal{D}_\lambda^2 k^2}} \right] , \qquad (E12)$$

and the corresponding eigenvector

$$\nu_\lambda = \prod_{k>0} \exp\left[\frac{|\rho(k,t)|^2}{2\Delta_\lambda k^2} \Delta\psi_k(\lambda) \right] , \qquad (E13)$$

factorizes into a product of independent modes, each quadratic in the density. For a density type variable, the optimal control force is a gradient force, and so can be written as a potential. It is computable following Refs. 80 and 96, which in the limit that λ approaches zero we recover Eq. 13.

[1] C. Nardini, É. Fodor, E. Tjhung, F. Van Wijland, J. Tailleur, and M. E. Cates, Physical Review X 7, 021007 (2017).

[2] É. Fodor, T. Nemoto, and S. Vaikuntanathan, New Journal of Physics 22, 013052 (2020).

[3] C. del Junco, L. Tociu, and S. Vaikuntanathan, Proceedings of the National Academy of Sciences 115, 3569 (2018).

[4] P. Gaspard and R. Kapral, The Journal of chemical physics 148, 134104 (2018).

[5] K. Dasbiswas, K. K. Mandadapu, and S. Vaikuntanathan, Proceedings of the National Academy of Sciences 115, E9031 (2018).

[6] V. Soni, E. S. Bililign, S. Magkiriadou, S. Sacanna, D. Bartolo, M. J. Shelley, and W. T. Irvine, Nature Physics 15, 1188 (2019).

[7] Z. Liao, W. T. Irvine, and S. Vaikuntanathan, Physical Review X 10, 021036 (2020).

[8] F. Caballero and M. E. Cates, Phys. Rev. Lett. 124, 240604 (2020).

[9] É. Fodor, C. Nardini, M. E. Cates, J. Tailleur, P. Visco, and F. van Wijland, Physical review letters 117, 038103 (2016).

[10] D. Mandal, K. Klymko, and M. R. DeWeese, Physical review letters 119, 258001 (2017).

[11] L. Metselaar, J. M. Yeomans, and A. Doostmohammadi, Physical Review Letters 123, 208001 (2019).

[12] Y. Li and P. R. ten Wolde, Physical review letters 123, 148003 (2019).

[13] C. Wang, Y.-k. Guo, W.-d. Tian, and K. Chen, The Journal of chemical physics 150, 044907 (2019).

[14] R. G. Morris and M. Rao, Physical Review E 100, 022413 (2019).

[15] N. Bain and D. Bartolo, Science 363, 46 (2019).

[16] A. Morin and D. Bartolo, Physical Review X 8, 021037 (2018).

[17] A. Bricard, J.-B. Caussin, N. Desreumaux, O. Dauchot, and D. Bartolo, Nature 503, 95 (2013).

[18] A. Mietke, V. Jemseena, K. V. Kumar, I. F. Sbalzarini, and F. Jülicher, Physical review letters 123, 188101 (2019).

[19] A. Morozov, Science 355, 1262 (2017).

[20] A. Souslov, B. C. Van Zuiden, D. Bartolo, and V. Vitelli, Nature Physics 13, 1091 (2017).

[21] C. Ropp, N. Bachelard, D. Barth, Y. Wang, and X. Zhang, Nature Photonics 12, 739 (2018).

[22] S. C. Takatori and A. Sahu, Physical Review Letters 124, 158102 (2020).

[23] D. Chaudhuri, Physical Review E 90, 022131 (2014).

[24] P. Pietzonka, E. Fodor, C. Lohrmann, M. E. Cates, and U. Seifert, Phys. Rev. X 9, 041032 (2019).

[25] S. Krishnamurthy, S. Ghosh, D. Chatterji, R. Ganapathy, and A. Sood, Nature Physics 12, 1134 (2016).

[26] R. Zakine, A. Solon, T. Gingrich, and F. van Wijland, Entropy **19**, 193 (2017).

[27] D. Martin, C. Nardini, M. E. Cates, and É. Fodor, EPL (Europhysics Letters) **121**, 60005 (2018).

[28] A. Saha, R. Marathe, P. Pal, and A. Jayannavar, Journal of Statistical Mechanics: Theory and Experiment **2018**, 113203 (2018).

[29] T. Ekeh, M. E. Cates, and E. Fodor, Phys. Rev. E **102**, 010101 (2020).

[30] S. Chaki and R. Chakrabarti, Physica A: Statistical Mechanics and its Applications **511**, 302 (2018).

[31] C. Jarzynski, Physical Review Letters **78**, 2690 (1997).

[32] G. E. Crooks, Physical Review E **60**, 2721 (1999).

[33] G. Gallavotti and E. G. D. Cohen, Physical review letters **74**, 2694 (1995).

[34] P. Gaspard and R. Kapral, The Journal of chemical physics **147**, 211101 (2017).

[35] J. M. Horowitz and T. R. Gingrich, Nature Physics , 1 (2019).

[36] I. Di Terlizzi and M. Baiesi, Journal of Physics A: Mathematical and Theoretical **52**, 02LT03 (2018).

[37] J. A. Owen, T. R. Gingrich, and J. M. Horowitz, Physical Review X **10**, 011066 (2020).

[38] C. Nardini and H. Touchette, The European Physical Journal B **91**, 16 (2018).

[39] C. Y. Gao and D. T. Limmer, The Journal of chemical physics **151**, 014101 (2019).

[40] A. Dechant and S.-i. Sasa, Proceedings of the National Academy of Sciences **117**, 6430 (2020).

[41] M. Polettini and M. Esposito, Journal of Statistical Physics **176**, 94 (2019).

[42] M. Barbier and P. Gaspard, Journal of Physics A: Mathematical and Theoretical **51**, 355001 (2018).

[43] J. M. Epstein and K. K. Mandadapu, Physical Review E **101**, 052614 (2020).

[44] C. Hargus, K. Klymko, J. M. Epstein, and K. K. Mandadapu, The Journal of Chemical Physics **152**, 201102 (2020).

[45] C. G. Wagner, M. F. Hagan, and A. Baskaran, Physical Review E **100**, 042610 (2019).

[46] K. Asheichyk, A. P. Solon, C. M. Rohwer, and M. Krüger, The Journal of chemical physics **150**, 144111 (2019).

[47] S. Dal Cengio, D. Levis, and I. Pagonabarraga, Physical Review Letters **123**, 238003 (2019).

[48] L. Caprini, U. M. B. Marconi, and A. Vulpiani, Journal of Statistical Mechanics: Theory and Experiment **2018**, 033203 (2018).

[49] H. Merlitz, H. D. Vuijk, J. Brader, A. Sharma, and J.-U. Sommer, The Journal of Chemical Physics **148**, 194116 (2018).

[50] Z. Liao, M. Han, M. Fruchart, V. Vitelli, and S. Vaikuntanathan, The Journal of chemical physics **151**, 194108 (2019).

[51] R. Chetrite and H. Touchette, in *Annales Henri Poincaré*, Vol. 16 (Springer, 2015) pp. 2005–2057.

[52] A. C. Barato and U. Seifert, Physical review letters **114**, 158101 (2015).

[53] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England, Phys. Rev. Lett. **116**, 120601 (2016).

[54] J. Lebowitz and H. Spohn, Journal of Statistical Physics **95**, 333 (1999).

[55] U. Seifert, Physical review letters **95**, 040602 (2005).

[56] C. Maes and K. Netočný, Journal of statistical physics **110**, 269 (2003).

[57] L. Caprini, U. M. B. Marconi, A. Puglisi, and A. Vulpiani, Journal of Statistical Mechanics: Theory and Experiment **2019**, 053203 (2019).

[58] S. Chaki and R. Chakrabarti, Physica A: Statistical Mechanics and its Applications **530**, 121574 (2019).

[59] F. Cagnetta, F. Corberi, G. Gonnella, and A. Suma, Physical review letters **119**, 158002 (2017).

[60] T. Nemoto, É. Fodor, M. E. Cates, R. L. Jack, and J. Tailleur, Physical Review E **99**, 022605 (2019).

[61] S. Shankar and M. C. Marchetti, Physical Review E **98**, 020604 (2018).

[62] G. Szamel, Physical Review E **100**, 050603 (2019).

[63] P. Pietzonka and U. Seifert, Journal of Physics A: Mathematical and Theoretical **51**, 01LT01 (2017).

[64] L. Dabelow, S. Bo, and R. Eichhorn, Physical Review X **9**, 021009 (2019).

[65] T. Speck, arXiv preprint arXiv **1707** (2017).

[66] T. Speck, EPL (Europhysics Letters) **123**, 20007 (2018).

[67] T. Speck, A. M. Menzel, J. Bialké, and H. Löwen, The Journal of chemical physics **142**, 224109 (2015).

[68] T. GrandPre and D. T. Limmer, Physical Review E **98**, 060601 (2018).

[69] C. Giardina, J. Kurchan, and L. Peliti, Physical review letters **96**, 120603 (2006).

[70] U. Ray, G. K.-L. Chan, and D. T. Limmer, Physical review letters **120**, 210602 (2018).

[71] J. D. Weeks, D. Chandler, and H. C. Andersen, The Journal of chemical physics **54**, 5237 (1971).

[72] F. Cagnetta and E. Mallmin, Physical Review E **101**, 022130 (2020).

[73] G. S. Redner, M. F. Hagan, and A. Baskaran, Physical review letters **110**, 055701 (2013).

[74] M. E. Cates and J. Tailleur, Annu. Rev. Condens. Matter Phys. **6**, 219 (2015).

[75] P. Chiarantoni, F. Cagnetta, F. Corberi, G. Gonnella, and A. Suma, Journal of Physics A: Mathematical and Theoretical (2020).

[76] U. Ray, G. K.-L. Chan, and D. T. Limmer, The Journal of chemical physics **148**, 124120 (2018).

[77] D. S. Dean, Journal of Physics A: Mathematical and General **29**, L613 (1996).

[78] J. Tailleur and M. Cates, Physical review letters **100**, 218103 (2008).

[79] S. Chakraborti, S. Mishra, and P. Pradhan, Physical Review E **93**, 052606 (2016).

[80] J. Dolezal and R. L. Jack, Journal of Statistical Mechanics: Theory and Experiment **2019**, 123208 (2019).

[81] A. Das and D. T. Limmer, The Journal of Chemical Physics **151**, 244123 (2019).

[82] L. Tociu, É. Fodor, T. Nemoto, and S. Vaikuntanathan, Physical Review X **9**, 041026 (2019).

[83] E. Crosato, M. Prokopenko, and R. E. Spinney, Physical Review E **100**, 042613 (2019).

[84] R. Nigmatullin and M. Prokopenko, arXiv:1912.08948 (2019).

[85] A. P. Solon, J. Stenhammar, R. Wittkowski, M. Kardar, Y. Kafri, M. E. Cates, and J. Tailleur, Physical review letters **114**, 198301 (2015).

[86] Y. Fily, A. Baskaran, and M. F. Hagan, The European Physical Journal E **40**, 1 (2017).

[87] H. Löwen, EPL (Europhysics Letters) **121**, 58001 (2018).

[88] R. G. Winkler, Soft matter **12**, 3737 (2016).

[89] E. G. Hidalgo, T. Nemoto, and V. Lecomte, Physical Review E **95**, 062134 (2017).

[90] S. C. Takatori, W. Yan, and J. F. Brady, Physical review letters **113**, 028103 (2014).

[91] A. K. Omar, Z.-G. Wang, and J. F. Brady, Physical Review E **101**, 012604 (2020).

[92] H. Touchette, Physica A: Statistical Mechanics and its Applications **504**, 5 (2018).

[93] R. Chetrite and H. Touchette, Journal of Statistical Mechanics: Theory and Experiment **2015**, P12001 (2015).

[94] A. P. Solon, M. Cates, and J. Tailleur, The European Physical Journal Special Topics **224**, 1231 (2015).

[95] J. Bialké, H. Löwen, and T. Speck, EPL (Europhysics Letters) **103**, 30008 (2013).

[96] F. Angeletti and H. Touchette, Journal of Mathematical Physics **57**, 023303 (2016).