# Optimizing active work: Dynamical phase transitions, collective motion, and jamming

Takahiro Nemoto,<sup>1</sup> Étienne Fodor,<sup>2</sup> Michael E. Cates,<sup>2</sup> Robert L. Jack,<sup>2,3</sup> and Julien Tailleur<sup>4</sup>

<sup>1</sup>Philippe Meyer Institute for Theoretical Physics, Physics Department, École Normale Supérieure & PSL Research University,

24 rue Lhomond, 75231 Paris Cedex 05, France

<sup>2</sup>Department of Applied Mathematics and Theoretical Physics, University of Cambridge,

Wilberforce Road, Cambridge CB3 0WA, United Kingdom

<sup>3</sup>Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

<sup>4</sup>Laboratoire Matière et Systèmes Complexes, UMR 7057 CNRS/P7, Université Paris Diderot,

10 rue Alice Domon et Léonie Duquet, 75205 Paris cedex 13, France

(Received 8 May 2018; revised manuscript received 9 September 2018; published 8 February 2019)

Active work measures how far the local self-forcing of active particles translates into real motion. Using population Monte Carlo methods, we investigate large deviations in the active work for repulsive active Brownian disks. Minimizing the active work generically results in dynamical arrest; in contrast, despite the lack of aligning interactions, trajectories of high active work correspond to a collectively moving, aligned state. We use heuristic and analytic arguments to explain the origin of dynamical phase transitions separating the arrested, typical, and aligned regimes.

DOI: 10.1103/PhysRevE.99.022605

## I. INTRODUCTION

Active particles constitute an important class of nonequilibrium systems, with examples ranging from bacteria to synthetic colloidal swimmers [1–7]. These particles expend energy to propel themselves, driving active matter out of equilibrium at microscopic scales and causing rich dynamical behaviors. Some of these are *universal*, whereby systems that differ microscopically show similar emergent physics [8], such as motility-induced phase separation (MIPS) [9–15], collective motion [6,16–21], lane formation [22,23], or motile defects [24–29].

Many recent advances in our understanding of nonequilibrium systems are based on *large-deviation theory* (LDT) [30,31]. This extends the counting procedures of equilibrium statistical mechanics from configuration space to trajectory space, addressing collective phenomena such as dynamical phase transitions. It has been used to characterize dynamical symmetries [32–35], measure free-energy differences [36,37], and locate atypical trajectories, such as activated processes [38]. LDT has proven useful in fields ranging from dynamical systems [39] and glasses [40,41] to fluid mechanics [42] and geophysical flows [43]. In contrast, the full range of insights offered by LDT to active matter remains largely unexplored, despite a handful of pioneering studies [44–50].

Here we use LDT to study active Brownian particles (ABPs) interacting via repulsive central forces. We focus on the large deviations of the active work, defined as the particle-averaged inner product of propulsive force and velocity. This measures how far the local self-forcing of active particles translates into real motion. A recent study [45] used brute-force simulations to sample the fluctuations of active work in a dilute system of active dumbbells and found a low active work to correlate with the emergence of ordered clusters in this system. Here we use an advanced numerical method

[39,51–58] to explore the full large-deviation regime in all relevant regions of the phase diagram of our ABP model. We first show that finite systems always admit a large deviation principle, notwithstanding [45], but that they are flanked by two dynamical phase transitions. Indeed, minimizing the active work always leads to complete dynamical arrest, whether or not the unbiased system exhibits MIPS. Biasing instead toward high active work, we find a striking result: Trajectories now correspond to flocked states of aligned collective motion, despite the microscopic absence of aligning interactions. We explain the origin of the dynamical phase transitions separating these regimes using a combination of arguments including macroscopic fluctuation theory [59,60].

#### **II. MODEL**

We consider *N* active Brownian particles interacting via purely repulsive pairwise forces in two spatial dimensions [11–15]. The positions and orientations of the particles are  $r_i$  and  $\theta_i$ ; they evolve as

$$\dot{\boldsymbol{r}}_i = \mu \boldsymbol{F}_{i,\text{ex}} + v_p \boldsymbol{u}(\theta_i) + \sqrt{2D} \boldsymbol{\eta}_i; \quad \dot{\theta}_i = \sqrt{2D_r \xi_i}, \quad (1)$$

where  $\eta_i, \xi_i$  are zero-mean unit-variance Gaussian white noises,  $\mu$  is a particle's mobility,  $v_p$  its bare self-propulsion speed and  $u(\theta_i) = (\cos \theta_i, \sin \theta_i)$  its orientation vector, and  $D, D_r$  are translational and rotational diffusivities. Particles interact via a repulsive WCA potential, detailed in Appendix A, of range  $\sigma$ . For consistency with Ref. [12], we set  $D_r = 3D/\sigma^2$  and the WCA strength parameter to be  $D/\mu$ . Then, we choose space and time units such that  $\sigma = 1$  and  $\sigma/v_p = 1$ (see Appendix A). When the persistence length  $\ell_p \equiv v_p/D_r$ is much larger than the particle size,  $\ell_p/\sigma \gtrsim 15$ , the system undergoes MIPS: at high volume fractions, a vapor of motile particles coexists with dense macroscopic clusters [11–15]. For smaller  $\ell_p$ , the system remains uniform and the main effect of activity is to enhance the effective translational diffusivity.

For interacting particles, a natural measure of how efficiently active forces create motion is given by the propulsive speed  $v_i \equiv \dot{r_i} \cdot u(\theta_i)$ , which projects a particle's velocity along its orientation. This relates directly to the active work [45], the total work done by the active forces on the particles, which obeys (in the Stratonovich convention)

$$W_a(t) \equiv \frac{v_p}{\mu} \int_0^t \sum_{i=1}^N \dot{\boldsymbol{r}}_i(\tau) \cdot \boldsymbol{u}[\theta_i(\tau)] d\tau = \sum_{i=1}^N \int_0^t \frac{v_p v_i(\tau)}{\mu} d\tau.$$
(2)

For conservative interactions,  $W_a$  relates to the dissipation in the thermostat  $\int dt \dot{\mathbf{r}}_i \cdot (\dot{\mathbf{r}}_i - \sqrt{2D} \eta_i)$  [61–65], and thus to the entropy production in the full  $\{r_i, \theta_i\}$  configuration space [45]. (This is generally distinct from that measured in position space  $\{r_i\}$  [66]. See also Refs. [67,68] for a comparative study of different entropy productions.) It is convenient to consider a normalized rate of active work per particle,  $w \equiv W_a \mu / (v_p^2 N t)$ . The dilute limit of vanishing packing fraction  $\phi \to 0$  then leads to  $\langle w \rangle = 1$  which serves as a useful reference point.

For fixed N and large t, the distribution of w has a largedeviation form

$$p(w) \sim \exp[-tI(w)], \tag{3}$$

where I(w) is a rate function [31]. The corresponding cumulant generating function (CGF)

$$G(s) = \lim_{t \to \infty} \frac{1}{t} \log \langle e^{-stNw} \rangle \tag{4}$$

is related to I(w) by Legendre transformation. As shown in Appendix B, the functions I(w) and G(s) are convex, and G(s) obeys a fluctuation relation G(s) = G(a - s), with  $a = 3\ell_p/\sigma$ . The CGF is analogous to a free energy in equilibrium statistical mechanics [31,69–71]. Within this analogy, *trajectories* of our two-dimensional system (evolving in time) correspond to *configurations* of an anisotropic three-dimensional system. Suppose that one spatial dimension (the "length") of this anisotropic system becomes infinite, while the others remain fixed—this is analogous to considering trajectories with  $t \to \infty$  and fixed N. Phase transitions are not possible in such one-dimensional geometries, which is another way to see that G(s) must be convex (and analytic).

Now consider the limit  $N \to \infty$  (taken at fixed  $\phi$ , after  $t \to \infty$ ). In this case dynamical phase transitions are possible—the analogous thermodynamic system is becoming infinite along more than one spatial dimension [55,70,73]. The dynamical analogues of the (bulk) thermodynamic free-energy and entropy are

$$\mathcal{I}(w) = I(w)/N$$
 and  $\mathcal{G}(s) = G(s)/N.$  (5)

As in statistical mechanics, singularities in the large-*N* limits of these functions are interpreted as phase transitions [55,70,73–76].

To observe and measure large deviations of the active work, we use a cloning algorithm [39,53,54], also known as population Monte Carlo [77], whose optimized implementation using modified dynamics [54] is detailed in Appendix C. (See Refs. [51,58] for a lattice version of this algorithm and Ref. [46] for a recent application to active systems.) In essence, the method relies on evolving a large population of copies of the system to generate "biased ensembles" of trajectories that sample the average in Eq. (4) with a cost that scales linearly in t, allowing direct access to the large-tlimit. For positive and negative s, the biased ensembles are dominated by trajectories with atypically small and large w, respectively.

#### **III. RESULTS**

We first consider a system whose parameters lie (as  $N \to \infty$ ) within the MIPS region,  $\ell_p = 40\sigma$  and  $\phi = 0.65$  (See Ref. [12] for the full phase diagram of the system). We compute  $\mathcal{G}(s)$ , and  $w(s) \equiv -\mathcal{G}'(s)$ , which is the mean value of the active work in the presence of the bias, and its inverse s(w). We also determine the rescaled rate function as  $\mathcal{I}(w) = -s(w)w - \mathcal{G}[s(w)]$ . Our numerical results (Fig. 1) show three regimes separated by dynamical phase transitions that we discuss below: a MIPS-like coexistence between vapor and dense phases near s = 0; a phase-separated arrest (PSA) at large positive s; and a collectively moving (CM) state at large negative s.

#### A. Large active work: Collective motion

For large negative *s*, the biased ensemble probes atypically large values of the active work. Despite the absence of aligning interactions, the biased ensemble is dominated by trajectories where particles' orientations are aligned with each other, and they move collectively as a flock. A global order parameter for this transition is the orientation

$$\hat{\nu} = \left| \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{u}(\theta_i) \right|, \quad \nu(s) = \langle \hat{\nu} \rangle_s, \tag{6}$$

where angle brackets are averages within the biased ensemble. Figure 2 shows the emergence of global orientation for s < 0. In contrast, in the unbiased case (s = 0) there is no emergent alignment:  $v(s = 0) \rightarrow 0$  at large N.

To understand the emergence of order, note that particle alignment naturally promotes active work: The CM state has far fewer collisions than if motion is incoherent, so that active forces translate more efficiently into particle motion. To confirm this interpretation, we compute the rate function  $J(\bar{\nu})$ of the time-integrated orientation  $\bar{\nu} = t^{-1} \int_0^t \hat{\nu}(\tau) d\tau$  whose probability distribution scales as  $p(\bar{\nu}) \sim e^{-tJ(\bar{\nu})}$ , analogous to Eq. (3). Since the orientational dynamics of ABPs is independent of their positions, the rate function J can be computed semi-analytically as shown in Appendices E and F. It measures the probability of rare events where rotational symmetry is spontaneously violated. Now, define  $I_2$  as the joint rate function for  $\bar{\nu}$  and w, and let  $\nu_*(w) = \nu(s(w))$  be the average global orientation for a biased system with active work w. Then  $I(w) = \inf_{v} I_2(w, v) = I_2(w, v_*(w))$  where the first equality is the contraction principle for large deviations [31] and the second follows because the infimum is achieved by  $v_*$ . Similarly  $J(v_*(w)) = \inf_{w'} I_2(w', v_*(w)) \leq I_2(w, v_*(w)),$ 



FIG. 1. (a) The rescaled rate function  $\mathcal{I}(w)$ , for system parameters where MIPS occurs ( $\phi = 0.65$ ,  $\ell_p = 40\sigma$ ). The *N*-dependence shows strong finite size effects for  $w < \langle w \rangle$ . The labels PSA and CM indicate respectively regimes of phase-separated arrest and collective motion. (b) The function w(s) exhibits two sharp crossovers, which we attribute to two dynamical phase transitions. (These appear in (a) as near-linear segments of the rate function.) (c, d) Snapshot configurations for N = 64 in the biased ensemble, corresponding to the arrested phase [s = 0.8 (d)] and the collective motion phase [s = -3.2 (c)]. Particles are colored according to their orientations. For corresponding movies see Supplemental Material Ref. [72].

and hence

$$\mathcal{I}(w) \geqslant \frac{J(\nu_*(w))}{N}.$$
(7)

Figure 2(b) shows that the inequality Eq. (7) is almost saturated when  $w > \langle w \rangle_0$ . Physically, this indicates that the probability cost for creating a large fluctuation of the active work is dominated, for s < 0, by the cost to create an improbable global orientation, with an associated spontaneous



FIG. 2. (a) The orientation vector v(s) in the biased ensemble. (b) The difference between the two quantities appearing in Eq. (7), showing saturation of this bound at  $w > \langle w \rangle_0$ .

symmetry breaking and an accompanying singularity in  $\mathcal{I}(w)$ . It appears that "the best strategy" for a set of active particles to move fast is for them to collectively align. Here "best" means least improbable within the microscopic stochastic dynamics specified by Eq. (1). Note that the emergence of macroscopic arrested clusters due to MIPS can be suppressed by local torques that limit the head-on collisions of particles [78–80]. It is thus quite remarkable that the most likely way to generate an efficient motion of each particle, and hence a large active work, is through the emergence of a collectively moving state, and not through such local rearrangements (which do not lead to a CM state).

#### B. Small active work: Dynamic arrest

For positive *s*, the biased ensemble selects trajectories with atypically low active work, so that propulsive effort leads to little motion. On increasing the bias, we find the system sharply transitions into a dynamically arrested state. (See Fig. 1(c) and movies in Supplemental Material Ref. [72].) A signature of this transition is the discontinuity in w(s)reported in Fig. 1(b), which signifies a first-order dynamical phase transition: The linear segment in  $\mathcal{I}(w)$ , for w < 0.5, is analogous to a Maxwell construction and the discontinuity in w(s) to a jump in the order parameter. These features should become strict singularities only as  $N \to \infty$  but are clearly visible for N = 32, 64. (Note, however, that G(s) and I(w)are well-defined for any finite N, notwithstanding [45].) As N increases, the critical value of s moves toward zero (see Appendix G for this finite-size scaling analysis), suggesting



FIG. 3. The active work w(s) and the orientation v(s) for two state points where the unbiased dynamic gives homogeneous steady states. In (a, b) the state point is  $(\phi, \ell_p) = (0.1, 40\sigma)$ ; in (c, d) it is  $(\phi, \ell_p) = (0.65, 6.7\sigma)$ . The crossovers shown in these figures separate the homogeneous fluid (s = 0) from a CM phase (s < 0) and a PSA state (s > 0). These phases are qualitatively similar to those shown in Figs. 1(c) and 1(d). See movies in Supplemental Material Ref. [72].

that bulk MIPS states live on the verge of a first order transition to complete arrest.

This situation is reminiscent of dynamical phase transitions arising for activity-biased kinetically constrained models (KCMs) of glassy systems [40,70]. Indeed, our findings for s > 0 can be qualitatively understood by generalizing arguments developed for KCMs. Specifically, we exploit a variational principle that allows the rate function I(w) to be computed by considering what auxiliary "control" forces need to be added to a system to realize the rare trajectories of interest [81–83]. Stabilizing a large, dense cluster in a system undergoing MIPS only requires applying forces on its boundaries, hence involving a subextensive number of particles. This argument, detailed in Appendix H, immediately leads to  $\lim_{N\to\infty} \mathcal{I}(w) = 0$  and a dynamical phase transition at s = 0. Since the argument is variational in nature, it can be exploited in numerical simulations: We have used it to obtain bounds on I(w) for  $w < \langle w \rangle$  in large systems as shown in Appendix H, which are consistent with the presence of a phase transition and complement the accurate results for I(w) in small systems that we show in Fig. 1.

#### C. Large deviations for homogeneous steady states

So far we considered systems with parameters for which the unbiased, large N dynamics shows steady-state MIPS. We now consider large deviations from a steady state that is homogeneous. We focus on two state points:

 $(\phi, \ell_p) = (0.1, 40\sigma)$ , corresponding to a reduced density but large  $\ell_p$ , and  $(\phi, \ell_p) = (0.65, 6.7\sigma)$ , corresponding to high density but smaller  $\ell_p$ . Figure 3 and movies in Supplemental Material Ref. [72] show the asymptotic phases observed for large positive and negative *s* to be similar in both cases: for s < 0, they again exhibit collective motion while for s > 0 the system undergoes phase-separated arrest. Compared to Fig. 1, the crossover to the PSA state in Fig. 3(c) is smoother; at this smaller value of  $\ell_p$ , the density fluctuations of the ABPs are smaller (there is no MIPS) and the instability to phase separation is weaker.

This bias-induced phase separation can be explained by a hydrodynamic argument [60]. For a macroscopically homogeneous fluid, long-wavelength density fluctuations should obey an equation of the form [84]

$$\dot{\rho} = \nabla \cdot [D_{\text{eff}}(\rho)\nabla\rho + \sqrt{2\sigma(\rho)\xi}], \qquad (8)$$

where  $\rho$  is the local density,  $D_{\text{eff}}$  is a (density-dependent) diffusivity,  $\sigma(\rho)$  is a noise strength, and  $\xi$  is a Gaussian white noise. (Higher-order gradients, while relevant to MIPS [85–88], are negligible for the long-wavelength fluctuations of interest here.) It is then natural to approximate  $W[\rho] \simeq \frac{v_p}{\mu} \int dt dx \rho v(\rho)$ , where  $v(\rho)$  is the average of the effective active speed  $v_i$  in a homogeneous system of density  $\rho$ . This  $v(\rho)$  is known to decrease linearly with density in pairwiseforce active particles [11], so that the active work density  $\kappa(\rho) \equiv \frac{v_p}{\mu} \rho v(\rho)$  is a concave function. A density fluctuation  $\delta \rho$  then leads to a fluctuation of the active work  $\delta W = \int \frac{1}{2} \kappa''(\rho) (\delta \rho)^2 dx$  with  $\kappa''(\rho) < 0$ .

Large deviations of such observables in the setting of Eq. (8) are known to lead to phase separation in the large system limit  $L \to \infty$  whenever s > 0 and  $\kappa''(\rho) < 0$  [60]: A long-wavelength linear instability arises for  $s > \lambda_c/L^2$  with  $\lambda_c = (2\pi D_{\text{eff}})^2 / |\sigma \kappa''|$  [60,89,90]. This bias-induced instability arises in passive systems [60,90], and we argue that it applies to homogeneous, isotropic active fluids also, since the form of Eq. (8) is the same. Alongside it, any conventional phase separation, including MIPS, creates an instability even in the unbiased case, s = 0. This sets in as  $D_{\text{eff}}(\rho) \rightarrow 0$ . In that limit,  $\lambda_c \rightarrow 0$  so that the bias-induced and motilityinduced instabilities merge; physically, the bias reinforces the natural tendency to phase separate. (The convergence with N is slowest in the small persistence length region, Fig. 3(c), which is furthest from the MIPS regime.) In contrast, the collective motion regime observed for s < 0 has no passive counterpart and cannot be captured by Eq. (8), which assumes that the orientations are only weakly affected by the bias and can therefore be integrated out.

#### **IV. CONCLUSION**

We have shown, using a combination of numerical simulations and theoretical arguments, that active systems interacting via pairwise forces undergo several different dynamical phase transitions. Choosing a bias field to select trajectories of low active work, we found these trajectories to involve a coexistence of a dense jammed, arrested domain with a dilute vapor. This is the most likely way in which an active system that is normally a uniform bulk fluid can stop moving. Biasing in the other direction to find trajectories of high active work, we found collective motion with aligned propulsion directions despite the absence of aligning interactions microscopically.

We end by speculating about a link between large deviations and evolutionary biology, motivated by two observations. First, the cloning algorithm involves the evolution of a population of systems: the method balances their natural dynamics (which favor the unbiased steady state) and a selection pressure, which favors systems with atypical values of some *fitness* function (here, active work) [91,92]. Second, we have shown that alignment among ABPs tends to suppress collisions, leading to efficient motion. We have argued that alignment is an effective strategy for promoting particle motion, with a minimal cost (in probability). We suggest that this cost-minimization strategy might also be viewed as a possible evolutionary strategy for maximising active work in biological systems. We do not expect a general correspondence between evolutionary strategies and cost minimization, particularly since cost-minimization strategies may be complicated, perhaps requiring concerted motion across large length scales [71,73,81]. However, one may imagine that some robust characteristics (such as global alignment) might appear generically in both cost-minimization strategies and evolutionary strategies.

#### ACKNOWLEDGMENTS

This work was granted access to the HPC resources of CINES/TGCC under Allocation No. 2018-A0042A10457

made by GENCI and of MesoPSL financed by the Region Ile de France and the project Equip@Meso (Reference No. ANR-10-EQPX-29-01) of the program Investissements d'Avenir supervised by the Agence Nationale pour la Recherche. É.F. benefits from an Oppenheimer Research Fellowship from the University of Cambridge, and a Junior Research Fellowship from St Catherine's College. J.T. acknowledges support from the ANR grant Bactterns. Work funded in part by the European Research Council under the Horizon 2020 Programme, ERC Grant No. 740269. M.E.C. is funded by the Royal Society.

#### APPENDIX A: NONDIMENSIONALIZED TIME EVOLUTION EQUATIONS AND PECLET NUMBER

We use the particle radius  $\sigma$  and propulsion speed  $v_p$  to define dimensionless position and time as  $r_i/\sigma$  and  $v_p t/\sigma$ . We also define a dimensionless mobility  $\alpha = D/(D_r \sigma^2)$ . The dynamics Eqs. (1) then become

$$\dot{\mathbf{r}}_{i} = \alpha \frac{\sigma}{\ell_{p}} \tilde{\mathbf{F}}_{i,\text{ex}} + \mathbf{u}_{i}(\theta_{i}) + \sqrt{\alpha \frac{2\sigma}{\ell_{p}}} \boldsymbol{\eta}_{i}, \qquad (A1)$$

$$\dot{\theta}_i = \sqrt{\frac{2\sigma}{\ell_p}} \xi_i. \tag{A2}$$

The interaction force  $\tilde{F}_{i,\text{ex}}$  stems from the (dimensionless) WCA potential

$$V_{\rm ex} = \{4[(1/r)^{12} - (1/r)^6] + 1\}\Theta(2^{1/6} - r),$$
(A3)

where  $\Theta$  is a Heaviside step function. (Following Ref. [12], we chose the typical strength of the WCA potential to be  $D/\mu$  in the original units.) The Peclet number Pe used in Ref. [12] is given as

$$Pe = \frac{\ell_p}{\sigma \alpha},$$
 (A4)

with  $\alpha = 1/3$ . The normalized active work rate is  $w(t) = (1/(Nt)) \sum_{i=1}^{N} \int_{0}^{t} d\tilde{t} \, \dot{r}_{i}(\tilde{t}) \cdot \boldsymbol{u}_{i}$  in the dimensionless position and time units.

## APPENDIX B: FLUCTUATION RELATION AND CONVEXITY OF *I*(*w*)

The large deviation function of the active work satisfies the fluctuation theorem

$$\mathcal{G}(s) = \mathcal{G}(\operatorname{Pe} - s). \tag{B1}$$

This means that  $\mathcal{G}(s)$  takes its minimum value at  $s_{\min} = \text{Pe}/2$ , where  $\mathcal{G}'(s)$  vanishes. In Fig. 4, we show a numerical example of  $\mathcal{G}(s)$  for Pe = 1, which illustrates this symmetry property.

We now derive Eq. (B1). First, let  $P(\Omega)$  be the probability density of a trajectory  $\Omega = (\mathbf{r}_i(t'), \theta_i(t'))$  with  $t' \in [0, t]$ . We then define a time-reversed trajectory as  $\Omega^T = (\mathbf{r}_i(t - t'), \theta_i(t - t'))$ . Using standard methods [66], the ratio between  $P(\Omega)$  and  $P(\Omega^T)$  can be computed as

$$\frac{P(\Omega)}{P(\Omega^T)} = \exp\{\operatorname{Pe} tNw(t) - V(t) + V(0)\}, \qquad (B2)$$



FIG. 4. For Pe = 1 and  $\phi$  = 0.65, the cumulant generating function  $\mathcal{G}(s)$  is symmetric around  $s_{\min} = 1/2$ , as predicted by Eq. (B1).

where V(t) is the total potential energy of the system at time *t*. Multiplying both sides by  $e^{-stNw}P(\Omega^T)$  and summing upon all possible  $\Omega$ , we get

$$\langle e^{-stNw} \rangle = \langle e^{(s-\operatorname{Pe})tNw - V(0) + V(t)} \rangle.$$
 (B3)

The large time limit then immediately leads to the fluctuation theorem Eq. (B1).

An additional question within large-deviation theory is whether the limit in Eq. (4) is finite and whether the resulting G(s) is analytic. This discussion refers to finite systems, since it is clear from our results that the large-N limit of  $\mathcal{G}(s)$ can develop singularities. If G(s) is differentiable everywhere, then I(w) can be obtained from it by Legendre transformation and is also convex and continuous: All this follows from the Gärtner-Ellis theorem [31]. The finite-size scaling in our numerical work assumes that this theorem is applicable.

As noted in Sec 3.2 of Ref. [93], G(s) is the largest eigenvalue of a differential operator (which is called the tilted generator); and if this operator satisfies conditions for a Perron-Frobenius theorem then it has a finite spectral gap. This is sufficient to establish that G(s) is analytic, and hence the Gärtner-Ellis theorem. The Perron-Frobenius theorem can be applied to systems such as Eq. (1) as long as the number of particles is finite, the domain in which they move is compact, the noise terms in Eq. (1) are additive (and nonzero), and the forces are bounded.

For the system considered here, there is one subtlety, which is that the interparticle forces  $F_{i,ex}$  in Eq. (1) are not bounded. This prevents direct application of the Perron-Frobenius theorem. From a mathematical point of view, this raises the possibility that large deviations might be realized by trajectories where two (or more) particles collapse onto the same point. We do not have a mathematical proof that such trajectories can be neglected, but we see no physical reason why they would be relevant, and there is no evidence for them in our numerical computations. For this reason, we argue that the unbounded interaction forces in Eq. (1) can be truncated when particles come very close to the same point, without

changing any of the behavior that we find. In such a modified system (with finite N), the Perron-Frobenius theorem applies, which means that G(s) and hence I(w) are both analytic and are related by Legendre transformation

## APPENDIX C: ENHANCED CONVERGENCE OF THE CLONING ALGORITHM USING MODIFIED DYNAMICS

Our cloning algorithm gives access to the cumulant generating function  $\mathcal{G}(s)$  in the limit of large number of clones. To enhance the convergence of the algorithm, a generic strategy is to rely on modified dynamics [54]. We now detail the implementation of this strategy to sample the large deviations of the active work in our model. We first introduce the following modification of dynamics Eq. (1):

$$\dot{\boldsymbol{r}}_{i} = \alpha \frac{\sigma}{\ell_{p}} \tilde{\boldsymbol{F}}_{i,\text{ex}} + (1+f)\boldsymbol{u}_{i}(\theta_{i}) + \sqrt{\alpha \frac{2\sigma}{\ell_{p}}} \boldsymbol{\eta}_{i}, \qquad (C1)$$

and

$$\dot{\theta}_i = \sqrt{\frac{2\sigma}{\ell_p}} \xi_i. \tag{C2}$$

We denote  $P_f(\Omega)$  the probability of  $\Omega$  in this new system. The following identity is then satisfied:

$$P(\Omega)e^{-stNw} = P_f(\Omega)e^{-stN\tilde{w}},\tag{C3}$$

where the new bias  $\tilde{w}$  is defined as

$$-stN\tilde{w} = -\left(s + \frac{f\text{Pe}}{2}\right)tNw + \sum_{i=1}^{N} \int_{0}^{t} d\tilde{t}$$
$$\times \left(\frac{f\text{Pe}}{2} + \frac{f^{2}\text{Pe}}{4} + \frac{f}{2}\tilde{F}_{i,\text{ex}} \cdot \boldsymbol{u}_{i}\right). \quad (C4)$$

Simulating dynamics Eqs. (C1) and (C2) with the bias Eq. (C4) is thus equivalent to simulating Eq. (1) with a bias -stNw. In practice we use f = -2s/Pe so that the new bias  $\tilde{w}$  reduces to

$$\tilde{w} = 1 - \frac{s}{\text{Pe}} + \frac{1}{tN\text{Pe}} \sum_{i=1}^{N} \int_{0}^{t} d\tilde{t} \tilde{F}_{i,\text{ex}} \cdot \boldsymbol{u}_{i}, \qquad (C5)$$

which indeed produced faster convergence with the number of clones used in the simulations.

To characterize the CM state, we add another modifying force described as follows:

$$\dot{\mathbf{r}}_i = \frac{1}{\text{Pe}} \mathbf{F}_{i,\text{ex}} + \left(1 - \frac{2s}{\text{Pe}}\right) \boldsymbol{u}(\theta_i) + \sqrt{\frac{2}{\text{Pe}}} \boldsymbol{\eta}_i, \quad (C6)$$

and

$$\dot{\theta}_i = -gN \frac{\partial}{\partial \theta_i} \hat{v}^2 + \sqrt{\frac{2}{\alpha \text{Pe}}} \xi_i, \qquad (C7)$$

where g is a parameter whose value is discussed later. Similarly, the probability of the trajectory  $\omega$  in this modified

system,  $P_{\text{mod}}$ , is given by

$$P_{\text{mod}}(\omega) \propto \exp\left\{-\frac{\text{Pe}}{4}\int dt \sum_{i} \left[\dot{\mathbf{r}}_{i} - \frac{1}{\text{Pe}}\mathbf{F}_{i,\text{ex}} - \left(1 - \frac{2s}{\text{Pe}}\right)\boldsymbol{u}(\theta_{i})\right]^{2} - \frac{\alpha\text{Pe}}{4}\int dt \sum_{i} \left(\dot{\theta}_{i} + gN\frac{\partial}{\partial\theta_{i}}\hat{v}^{2}\right)^{2} - \frac{1}{4\text{Pe}}\int dt \sum_{i} \frac{\partial}{\partial\mathbf{r}_{i}}\mathbf{F}_{i,\text{ex}} + \frac{1}{2}\int_{0}^{\tau} dt \sum_{i} gN\frac{\partial^{2}\hat{v}^{2}}{\partial\theta_{i}^{2}}\right\}.$$
(C8)

By taking the ratio between  $P(\omega)$  and  $P_{mod}(\omega)$ , we get

$$\frac{P_{\text{mod}}(\omega)}{P(\omega)} \simeq \exp\left\{-s\tau Nw + \int_0^\tau dt \left[sN - \frac{s^2}{\text{Pe}}N + \frac{s}{\text{Pe}}\sum_i \boldsymbol{u}_i \cdot \boldsymbol{F}_{i,\text{ex}} + g - gN\hat{v}^2 - \frac{g^2\alpha\text{Pe}}{\hat{v}^2}\sum_i \sin^2(\theta_i - \varphi)\right]\right\},\tag{C9}$$

where w is the active work introduced in the main text. By defining the modified active work  $w_{mod}$  as

$$w_{\text{mod}} = \frac{1}{\tau} \int_0^\tau dt \left[ 1 - \frac{s}{\text{Pe}} + \frac{1}{N\text{Pe}} \sum_i \boldsymbol{u}_i \cdot \boldsymbol{F}_{i,\text{ex}} + \frac{g}{sN} - \frac{g}{s} \hat{\boldsymbol{v}}^2 - \frac{g^2 \alpha \text{Pe}}{sN \hat{\boldsymbol{v}}^2} \sum_i \sin^2(\theta_i - \varphi) \right], \tag{C10}$$

we thus get

$$P(\omega)e^{-s\tau Nw} = P_{\rm mod}(\omega)e^{-s\tau Nw_{\rm mod}}.$$
 (C11)

Note that g is a free parameter here: The equality Eq. (C11) holds irrespective of the value of g.

As discussed in Ref. [54] and in Appendix H, there is an optimal modification to the dynamics—if this could be found, then the cloning algorithm would have zero error and  $w_{mod}$  in Eq. (C11) would become a simple number (independent of the trajectory), equal to -G(s)/s. However, finding the optimal modification is as difficult as solving the large-deviation problem analytically, and is out of reach for most problems, including this one. Hence, the modifying forces used here are not optimal in the sense of Ref. [54], but we may still choose g so as to enfore the following equality:

$$\langle -sw_{\text{mod}} \rangle_{\text{mod}} = \hat{G}(s),$$
 (C12)

where  $\langle \rangle_{\text{mod}}$  means the average in the modified dynamics (obtained from the cloning algorithm) and  $\hat{G}(s)$  is the estimator of the cumulant generating function within the cloning algorithm. We found that this is an efficient way to choose our modifying force.

In Fig. 5, we compare the standard cloning algorithm [lefthand side of Eq. (C12)] and the modified one [right-hand side of Eq. (C12)] by plotting the active work as a function of s. We see that for N = 16 and N = 32, both algorithm lead to the same function w(s), but that the modified dynamics converges much faster as the number of clones  $N_c$  increases. For N = 64, however, the standard algorithm shows a very slow convergence, unlike our modified algorithm.

## APPENDIX D: THE PARAMETERS USED FOR POPULATION MONTE CARLO METHOD

Here we summarize the parameters used to get the results in the main text. Convergence is obtained using the number of clones  $N_c = 25\,600$ . The time-step of the simulations is dt = 0.001; cloning steps are performed each  $\Delta t = 0.01$ . The simulation length varies from  $t = 30\,000$  to  $300\,000$ , depending on the values of N and s. We checked the convergence with respect to the cloning parameters  $N_c$ ,  $\delta t$ , t, for all values of s except for the immediate vicinity of the PSA transition point (s > 0). Around this transition point, we observe a slight unphysical concavity of  $\mathcal{G}(s)$  which is a signature that perfect convergence is out of reach of our simulations [see Fig. 7(a)].

#### APPENDIX E: DERIVATION OF THE POLARIZATION DYNAMICS

This appendix is devoted to the derivation of the dynamics of the stochastic polarization  $\hat{\nu}$  defined in Eq. (6) of the main text. It can be written as

$$\hat{\nu} = \frac{1}{N} \sqrt{\sum_{\{i,j\}=1}^{N} \cos(\theta_i - \theta_j)}.$$
(E1)

We introduce a global phase  $\varphi$  such as

$$\hat{\nu}e^{\varphi} \equiv \frac{1}{N}\sum_{j=1}^{N}e^{i\theta_j}.$$
(E2)

Using Itô's lemma, taking time derivative of Eq. (E1) gives

$$\dot{\hat{\nu}} = -\frac{\sqrt{2D_{\rm r}}}{N^2 \hat{\nu}} \sum_{\{i,j\}=1}^{N} \xi_i \sin(\theta_i - \theta_j) + \frac{D_{\rm r}}{N} \sum_{k=1}^{N} \frac{d^2}{d\theta_k^2} \sqrt{\sum_{\{i,j\}=1}^{N} \cos(\theta_i - \theta_j)},$$
(E3)

where we have used  $\dot{\theta}_i = \sqrt{2D_r}\xi_i$ , and  $\xi_i$  is a zero-mean unit-variance Gaussian white noise. The second term can be written as

$$\sum_{k=1}^{N} \frac{d^2}{d\theta_k^2} \sqrt{\sum_{\{i,j\}=1}^{N} \cos(\theta_i - \theta_j)} = -\sum_{\{k,l\}=1}^{N} \frac{d}{d\theta_k} \frac{\sin(\theta_k - \theta_l)}{\sqrt{\sum_{i,j} \cos(\theta_i - \theta_j)}} = \frac{N - \sum_{k,l} \cos(\theta_k - \theta_l)}{\sqrt{\sum_{i,j} \cos(\theta_i - \theta_j)}} - \sum_{\{k,l,m\}=1}^{N} \frac{\sin(\theta_k - \theta_l) \sin(\theta_k - \theta_m)}{\left[\sum_{i,j} \cos(\theta_i - \theta_j)\right]^{3/2}} = \frac{1}{\hat{v}} - N\hat{v} - \frac{1}{(N\hat{v})^3} \sum_{\{k,l,m\}=1}^{N} \sin(\theta_k - \theta_l) \sin(\theta_k - \theta_m).$$
(E4)

The noise term appearing in Eq. (E3) is denoted by

$$\Lambda \equiv \sum_{\{i,j\}=1}^{N} \xi_i \sin(\theta_i - \theta_j).$$
(E5)

It is a zero-mean Gaussian white noise with correlations

$$\langle \Lambda(t)\Lambda(0) \rangle = \sum_{\{i,j,k,l\}=1}^{N} \langle \xi_i(t)\xi_k(0) \rangle \sin[\theta_i(t) - \theta_j(t)]$$
  
 
$$\times \sin[\theta_k(0) - \theta_l(0)]$$
  
 
$$= \delta(t) \sum_{\{j,k,l\}=1}^{N} \sin(\theta_k - \theta_j)\sin(\theta_k - \theta_l).$$
 (E6)

To proceed further, we note that the sum can be simplified, using Eq. (E2), as

$$\sum_{\{j,k,l\}=1}^{N} \sin(\theta_k - \theta_j) \sin(\theta_k - \theta_l)$$
$$= (N\hat{\nu})^2 \sum_{i=1}^{N} \sin(\theta_i - \varphi)^2$$
$$= (N\hat{\nu})^2 \int_0^{2\pi} \sin(\theta - \varphi)^2 \psi(\theta, t) d\theta, \qquad (E7)$$

where we have introduced the angular distribution  $\psi(\theta, t) \equiv \sum_i \delta[\theta - \theta_i(t)]$ . We now assume that  $\psi$  is close to uniform, which should hold for  $\hat{\nu} \ll 1$  and large *N*, to obtain

$$\sum_{\substack{\{j,k,l\}=1}}^{N} \sin(\theta_k - \theta_j) \sin(\theta_k - \theta_l)$$
$$\xrightarrow{\hat{\nu} \ll 1, N \gg 1} \frac{N^3 \hat{\nu}^2}{2\pi} \int_0^{2\pi} \sin(\theta - \varphi)^2 d\theta = \frac{N^3 \hat{\nu}^2}{2}.$$
 (E8)

Finally, substituting this result in Eqs. (E4) and (E6), then Eq. (E3) reduces to

$$\frac{d\hat{\nu}}{dt} = D_r \left[ \frac{1}{2N\hat{\nu}} - \hat{\nu} \right] + \sqrt{\frac{D_r}{N}} \xi,$$
(E9)

which is a closed (autonomous) equation for the evolution of  $\hat{\nu}$ .

#### APPENDIX F: CGF OF THE TIME-AVERAGED TOTAL ORIENTATION **<b>**

In this Appendix, we consider the cumulant generating function of  $\bar{\nu}$ , defined as

$$\mathcal{H}(\varsigma) = \frac{1}{N} \lim_{t \to \infty} \frac{1}{t} \log \langle e^{-tN\varsigma \bar{\nu}} \rangle.$$
 (F1)

We work under the assumption that  $\hat{\nu} \ll 1$ ,  $N \to \infty$  so that we can use the time-evolution equation for  $\hat{\nu}$  given by Eq. (E9). We introduce the rescaled variable  $q = \hat{\nu}\sqrt{N}$ , whose dynamics is given by

$$\frac{dq}{dt} = D_r \left[ \frac{1}{2q} - q \right] + \sqrt{D_r} \xi.$$
 (F2)

Note that Eq. (F2) is *independent* of N. We then consider the cumulant generating function of the time-averaged value of q:

$$f(k) = \lim_{t \to \infty} \frac{1}{t} \log \langle e^{-k \int_0^t d\tilde{t} q(\tilde{t})} \rangle.$$
 (F3)

The CGF f(k) is the largest eigenvalue of the following operator:

$$L_k[\cdot] = -\frac{\partial}{\partial q} \left[ D_r \left( \frac{1}{2q} - q \right) \cdot \right] + \frac{D_r}{2} \frac{\partial^2}{\partial q^2} [\cdot] - kq. \quad (F4)$$

Since  $L_k$  is independent of N, f(k) is a well-defined smooth function in the  $N \to \infty$  limit:

$$f(k) = \lim_{N,t \to \infty} \frac{1}{t} \log \left\langle e^{-k \int d\tilde{t} q(\tilde{t})} \right\rangle.$$
(F5)

Using  $k = \zeta \sqrt{N}$ ,  $\mathcal{H}(\zeta)$  can now be expressed as

$$\mathcal{H}(\varsigma) = \frac{f(\varsigma\sqrt{N})}{N},\tag{F6}$$

or, conversely,

$$f(k) = N\mathcal{H}\left(\frac{k}{\sqrt{N}}\right).$$
 (F7)

In Fig. 6, we numerically demonstrate Eq. (F7). To do so, we compare the results obtained by applying the cloning algorithm to the dynamics Eq. (A2) of N independent rotors, which yields the right-hand side of Eq. (F7), with the result of the numerical diagonalization of the operator Eq. (F4), which yields the left-hand side of Eq. (F7). The results of the cloning algorithm for several N clearly collapse onto a single function f(k). Note that this overlap is satisfied not only for positive s (where the assumption  $\hat{v} \ll 1$  is safely satisfied) but also for negative s close to the origin.



FIG. 5. The estimator of the active work obtained from the standard (std.) cloning algorithm [left-hand side of Eq. (C11)] and modified (mod.) cloning algorithm [right-hand side of Eq. (C11)] for N = 16 (a), N = 32 (b), and N = 64 (c). The modified algorithm shows a much faster convergence as the number of clones is increased. Its limiting value agrees with the standard algorithm when the latter has converged.

# APPENDIX G: FINITE-SIZE SCALING TO ESTIMATE PSA TRANSITION POINT IN $N \rightarrow \infty$

We denote by  $s_c(N)$  the PSA transition point for finite system size N. It is defined as the value of s that maximizes



FIG. 6. An example of f(k) obtained from numerical diagonalization of the operator Eq. (F4) in blue solid line. We set  $D_r =$ 1/40. We also plot  $N\mathcal{H}(k/\sqrt{N})$  obtained from the application of the cloning algorithm to N active rotors. The data points for N =8, 16, 32, 64 correspond to red-circle, blue-triangle green-square and yellow cross. We can see clear overlap of the data points to a single line, demonstrating the validity of Eq. (F7).

the second derivative of  $\mathcal{G}(s)$  for positive *s*. The obstacle to estimate  $s_c(N)$  is that there are strong finite-size effects with respect to the number of clones  $N_c$  around  $s_c(N)$  [that artificially violate the convexity of  $\mathcal{G}(s)$  as seen in Fig. 7(a)]. To overcome this difficulty, we extract  $s_c(N)$  from the crossing point of the straight lines obtained by fitting the data for  $s < s_c(N)$  and  $s_c(N) < s$ , as shown in Fig. 7(a). Due to the convexity of G(s), the crossing point determined in this way gives a good approximation of  $s_c(N)$  [55]. We then plot  $s_c(N)$ as a function of N and extrapolate  $\lim_{N\to\infty} s_c(N)$ . As seen from Fig. 7(b),  $s_c(N)$  is consistent with a convergence to zero, with a power law:  $s_c(N) \sim N^{-a}$ .

# APPENDIX H: OPTIMAL CONTROL ARGUMENT FOR DYNAMICAL ARREST OF A MIPS CLUSTER (PSA TRANSITION)

We consider a system that is obtained by adding additional "control" forces to Eq. (1), leading to

$$\dot{\boldsymbol{r}}_{i} = \boldsymbol{B}_{i}^{r}(\boldsymbol{r},\theta) + \mu \boldsymbol{F}_{i,\mathrm{ex}} + \boldsymbol{v}_{p}\boldsymbol{u}(\theta_{i}) + \sqrt{2D\boldsymbol{\eta}_{i}},$$
  
$$\dot{\theta}_{i} = \boldsymbol{B}_{\theta}^{\theta}(\boldsymbol{r},\theta) + \sqrt{2D_{r}}\boldsymbol{\xi}_{i}.$$
(H1)

The control forces  $B^r$ ,  $B^\theta$  depend on the co-ordinates of all particles. A general result in large-deviation theory (see, e.g., Eq. (54) of Ref. [82] as well as Ref. [81] and the discussion in Sec. 4 of Ref. [83]) is that

$$I(w) = \inf_{B^r, B^\theta: \langle w \rangle_{\text{control}} = w} \Phi(B^r, B^\theta),$$
(H2)

where  $\Phi$  is a "cost function," and the infimum runs over those forces for which the steady state of Eq. (H1) has a mean active work w. The cost function  $\Phi$  is the relative entropy between two ensembles of trajectories, which are the (unbiased) steady state of the original system, and the steady state of the controlled system. This relative entropy is related to the large-deviation rate function at level-2.5: in the present



FIG. 7. (a) How to estimate  $s_c(N)$  from  $\mathcal{G}(s)$ . For *s* slightly larger than  $s_c$ , the required number of clones  $N_c$  needed to observe the convergence of the cloning algorithm rapidly increases beyond what can be reached numerically. As a consequence, we see an artificial violation of the concavity of  $\mathcal{G}(s)$ . To interpolate the correct shape of  $\mathcal{G}(s)$ , and locate  $s_c(N)$ , we use the data where the concavity is not violated. In both panels, dashed and dotted straight lines are obtained by fitting the data in inactive  $[s \gg s_c(N)]$  and active  $[s \ll s_c(N)]$ regions, respectively. Assuming a sharp kink [55], we obtain the estimated value of  $s_c(N)$  as the crossing point of these two lines (indicated as the black circles for each N). (b)  $s_c(N)$  estimated from the finite-size scaling on the data up to N = 64 (for  $\ell_p = 40\sigma$ ,  $\phi =$ 0.65) shows a power law decay with respect to N:  $s_c(N) \sim N^{-a}$  with a positive constant *a*. In particular, this implies  $\lim_{N\to\infty} s_c(N) = 0$ .

context it is simply [82, Eq. (76)]

$$\Phi(B^r, B^{\theta}) = \left\langle \sum_{i} \left[ \frac{\left(B_i^r\right)^2}{4D} + \frac{\left(B_i^{\theta}\right)^2}{4D_r} \right] \right\rangle_{\text{control}}, \quad (\text{H3})$$

where the average is taken in the steady state of Eq. (H1).

From Eq. (H2) one has that for any control forces *B* that realise the desired active work, then  $I(w) \leq \Phi(B^r, B^\theta)$ . To establish the existence of a phase transition at s = 0, it is sufficient to find (for each *N*) some *B* that realize the desired active work, with  $\lim_{N\to\infty} N^{-1}\Phi(B^r, B^\theta) = 0$  (that is,  $\Phi$ is subextensive). In this case  $\mathcal{I}_{\infty}(w) \equiv \lim_{N\to\infty} N^{-1}I(w) \leq$ 



FIG. 8. Schematic figure to show  $\sum_{j} \hat{\boldsymbol{e}}_{i,j} \simeq 0$  when a particle (indicated as gray color) is surrounded by six particles.

 $\lim_{N\to\infty} N^{-1}\Phi(B^r, B^\theta) = 0$ . The rate function is nonnegative so this is sufficient to show that  $\mathcal{I}_{\infty}(w) = 0$ .

To illustrate how this argument works, recall the case of kinetically constrained models. In that case the variational principle Eq. (H2) simplifies [83], because the controlled systems are at equilibrium and are fully characterized by their Boltzmann distributions. One may then find *B* such that the system is localized in a single state and the corresponding  $\Phi$  is the escape rate from that site. In KCMs there are configurations for which this escape rate is subextensive, leading



FIG. 9. (a) Snapshot of the original dynamics, for N = 4096 particles, leading to  $\langle w \rangle \simeq 0.25$ . (b) Snapshot of the dynamics with the control torque Eq. (H4) and g = 0.22. The control torques clearly reduce the number of particles in the gas phase, leading to a lower active work  $\langle w \rangle \simeq 0.05$ . The color code corresponds to the contribution of each particle to the cost function Eq. (H3), normalized by  $g^2$ . The cost is clearly dominated by the subextensive contribution of the particles localized at the boundary of the main cluster. (c) The upper bound  $\Phi$  of the large deviation function as a function of  $\langle w \rangle_g = \langle w \rangle_{\text{control}}$ , normalized by the number of particles. The bound decreases as the system size increases.

to  $\mathcal{I}_{\infty}(a) = 0$  for  $0 < a < \langle a \rangle$  (where *a* is the dynamical activity).

In the present context, we suppose that the natural state of the system is phase-separated (due to MIPS) and we consider a controlled steady state that is also phase-separated. We take  $B^r = 0$ , and we apply torques  $B^{\theta}$  that act on the particles near the boundary of the dense cluster, which favor orientations pointing toward the cluster. These torques will help to reinforce the MIPS state, and they also act to compress the cluster, so that its density will increase, which tends to reduce particle motion. For any control force of this type, the only terms which contribute in Eq. (H3) are from particles on the cluster boundary, so the number of such terms is subextensive. Hence  $\Phi$  is subextensive (assuming that the  $B_i^{\theta}$ are bounded). This means that  $\mathcal{I}_{\infty}(w) = 0$  for any value of the active work that can be realized by a perturbation of this type. Our data (Fig. 1) indicate that values of w close to zero can be achieved with a subextensive cost, just as happens in KCMs.

Building such control forces and torques explicitly, that would apply only to particles located at the boundary of the cluster, is a numerical challenge. We can nevertheless test our hypothesis by considering the following protocol

$$B_i^{\theta} = -g \frac{\partial}{\partial \theta_i} \sum_j \hat{\boldsymbol{e}}_{i,j} \cdot \boldsymbol{u}_i(\theta_i)$$
(H4)

(with  $B_i^r = 0$ ). Here, g > 0 is a constant parameter and  $\hat{e}_{i,j}$  is a unit vector from the particle *j* to the particle *i* when they interact and zero otherwise:  $\hat{e}_{i,j} = \Theta(2^{1/6} - r_{i,j})r_{i,j}/r_{i,j}$  with  $r_{i,j} \equiv r_i - r_j$ . The torques Eq. (H4) will favor head-on collisions between interacting particles. At the boundary of the cluster, such torques lower the tendency of particles to rotate and leave the cluster. It will play little role in the gas phase, where there are few collisions. For the particles inside the dense arrested clusters,  $B_i$  is also small by symmetry (see Fig. 8). Therefore, we expect that the dynamics with the control torque Eq. (H4) will lead to a reinforcement of MIPS and hence a lower active work, with a cost function  $(B_i^{\theta})^2/4D$  nearly vanishing outside the boundaries of the cluster.

Simulations using the control force Eq. (H4) indeed show reduced numbers of gas-phase particles in Figs. 9(a) and 9(b), leading to smaller values of the active work when compared to the original dynamics. Figure 9(c) shows that, furthermore, as the system sizes are increased, the upper bound of the LDF  $\Phi(w)$  strongly decreases. These numerical results support our theory because they illustrate how a phase-separated arrested state can indeed be stabilised using a cost that is dominated by boundary contributions. Note that for much larger sizes, however, our cost function might saturate because the torque  $B_i^{\theta}$  does not vanish exactly in the bulk of the cluster and gas phases. Only a protocol that would be exactly restricted to the boundary region could be used to achieve the  $N \to \infty$  limit, which is anyway far beyond what we can do numerically.

- H. C. Berg, *E. coli in Motion* (Springer Science & Business Media, Berlin, 2008).
- [2] M. E. Cates, Diffusive transport without detailed balance in motile bacteria: Does microbiology need statistical physics? Rep. Prog. Phys. 75, 042601 (2012).
- [3] J. R. Howse, Richard A. L. Jones, A. J. Ryan, T. Gough, R. Vafabakhsh, and R. Golestanian, Self-Motile Colloidal Particles: From Directed Propulsion to Random Walk, Phys. Rev. Lett. 99, 048102 (2007).
- [4] I. Theurkauff, C. Cottin-Bizonne, J. Palacci, C. Ybert, and L. Bocquet, Dynamic Clustering in Active Colloidal Suspensions with Chemical Signaling, Phys. Rev. Lett. 108, 268303 (2012).
- [5] J. Palacci, S. Sacanna, A. P. Steinberg, D. J. Pine, and P. M. Chaikin, Living crystals of light-activated colloidal surfers, Science 339, 936 (2013).
- [6] A. Bricard, J.-B. Caussin, N. Desreumaux, O. Dauchot, and D. Bartolo, Emergence of macroscopic directed motion in populations of motile colloids, Nature 503, 95 (2013).
- [7] C. Bechinger, R. D. Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe, Active particles in complex and crowded environments, Rev. Mod. Phys. 88, 045006 (2016).
- [8] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, and R. Aditi Simha, Hydrodynamics of soft active matter, Rev. Mod. Phys. 85, 1143 (2013).
- [9] M. E. Cates and J. Tailleur, Motility-induced phase separation, Ann. Rev. Condensed Matter Phys. 6, 219 (2015).
- [10] J. Tailleur and M. E. Cates, Statistical Mechanics of Interacting Run-and-Tumble Bacteria, Phys. Rev. Lett. 100, 218103 (2008).

- [11] Y. Fily and M. C. Marchetti, Athermal Phase Separation of Self-Propelled Particles with No Alignment, Phys. Rev. Lett. 108, 235702 (2012).
- [12] G. S. Redner, M. F. Hagan, and A. Baskaran, Structure and Dynamics of a Phase-Separating Active Colloidal Fluid, Phys. Rev. Lett. **110**, 055701 (2013).
- [13] J. Bialké, H. Löwen, and T. Speck, Microscopic theory for the phase separation of self-propelled repulsive disks, Europhys. Lett. 103, 30008 (2013).
- [14] A. Wysocki, R. G. Winkler, and G. Gompper, Cooperative motion of active Brownian spheres in three-dimensional dense suspensions, Europhys. Lett. **105**, 48004 (2014).
- [15] J. Stenhammar, D. Marenduzzo, R. J. Allen, and M. E. Cates, Phase behavior of active Brownian particles: The role of dimensionality, Soft Matter 10, 1489 (2014).
- [16] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, Novel Type of Phase Transition in a System of Self-Driven Particles, Phys. Rev. Lett. 75, 1226 (1995).
- [17] G. Grégoire and H. Chaté, Onset of Collective and Cohesive Motion, Phys. Rev. Lett. 92, 025702 (2004).
- [18] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini *et al.*, Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study, Proc. Natl. Acad. Sci. USA **105**, 1232 (2008).
- [19] T. Vicsek and A. Zafeiris, Collective motion, Phys. Rep. 517, 71 (2012).

- [20] A. P. Solon, H. Chaté, and J. Tailleur, From Phase to Microphase Separation in Flocking Models: The Essential Role of Nonequilibrium Fluctuations, Phys. Rev. Lett. **114**, 068101 (2015).
- [21] J. Deseigne, O. Dauchot, and H. Chaté, Collective Motion of Vibrated Polar Disks, Phys. Rev. Lett. 105, 098001 (2010).
- [22] K. Klymko, P. L. Geissler, and S. Whitelam, Microscopic origin and macroscopic implications of lane formation in mixtures of oppositely driven particles, Phys. Rev. E 94, 022608 (2016).
- [23] C. del Junco, L. Tociu, and S. Vaikuntanathan, Energy dissipation and fluctuations in a driven liquid, Proc. Natl. Acad. Sci. USA 115, 3569 (2018).
- [24] V. Narayan, S. Ramaswamy, and N. Menon, Long-lived giant number fluctuations in a swarming granular nematic, Science 317, 105 (2007).
- [25] T. Sanchez, D. T. N. Chen, S. J. De Camp, M. Heymann, and Z. Dogic, Spontaneous motion in hierarchically assembled active matter, Nature 491, 431 (2012).
- [26] L. Giomi, M. J. Bowick, X. Ma, and M. Cristina Marchetti, Defect Annihilation and Proliferation in Active Nematics, Phys. Rev. Lett. **110**, 228101 (2013).
- [27] S. P. Thampi, R. Golestanian, and J. M. Yeomans, Instabilities and topological defects in active nematics, Europhys. Lett. 105, 18001 (2014).
- [28] S. J. De Camp, G. S. Redner, A. Baskaran, M. F. Hagan, and Z. Dogic, Orientational order of motile defects in active nematics, Nat. Mater. 14, 1110 (2015).
- [29] B. Mahault, X.-c. Jiang, E. Bertin, Y.-q. Ma, A. Patelli, X.-q. Shi, and H. Chaté, Self-Propelled Particles with Velocity Reversals and Ferromagnetic Alignment: Active Matter Class with Second-Order Transition to Quasi-Long-Range Polar Order, Phys. Rev. Lett. **120**, 258002 (2018).
- [30] B. Derrida, Non-equilibrium steady states: Fluctuations and large deviations of the density and of the current, J. Stat. Mech.: Theory Exper. (2007) P07023.
- [31] H. Touchette, The large deviation approach to statistical mechanics, Phys. Rep. 478, 1 (2009).
- [32] G. Gallavotti and E. G. D. Cohen, Dynamical Ensembles in Nonequilibrium Statistical Mechanics, Phys. Rev. Lett. 74, 2694 (1995).
- [33] J. Kurchan, Fluctuation theorem for stochastic dynamics, J. Phys. A: Math. Gen. 31, 3719 (1998).
- [34] J. L. Lebowitz and H. Spohn, A Gallavotti-Cohen-type symmetry in the large deviation functional for stochastic dynamics, J. Stat. Phys. 95, 333 (1999).
- [35] G. E. Crooks, Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences, Phys. Rev. E 60, 2721 (1999).
- [36] C. Jarzynski, Nonequilibrium Equality for Free Energy Differences, Phys. Rev. Lett. 78, 2690 (1997).
- [37] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies, Nature 437, 231 (2005).
- [38] R. V. Kohn, M. G. Reznikoff, and E. Vanden-Eijnden, Magnetic elements at finite temperature and large deviation theory, J. Nonlin. Sci. 15, 223 (2005).
- [39] J. Tailleur and J. Kurchan, Probing rare physical trajectories with Lyapunov weighted dynamics, Nat. Phys. 3, 203 (2007).

- [40] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland, Dynamical First-Order Phase Transition in Kinetically Constrained Models of Glasses, Phys. Rev. Lett. 98, 195702 (2007).
- [41] L. O. Hedges, R. L. Jack, J. P. Garrahan, and D. Chandler, Dynamic Order-Disorder in Atomistic Models of Structural Glass Formers, Science 323, 1309 (2009).
- [42] T. Grafke, R. Grauer, and T. Schäfer, The instanton method and its numerical implementation in fluid mechanics, J. Phys. A: Math. Theore. 48, 333001 (2015).
- [43] F. Bouchet, C. Nardini, and T. Tangarife, Stochastic averaging, large deviations, and random transitions for the dynamics of 2D and geostrophic turbulent vortices, Fluid Dynam. Res. 46, 061416 (2014).
- [44] A. G. Thompson, J. Tailleur, M. E. Cates, and R. A. Blythe, Lattice models of nonequilibrium bacterial dynamics, J. Stat. Mech.: Theory Exp. (2011) P02029.
- [45] F. Cagnetta, F. Corberi, G. Gonnella, and A. Suma, Large Fluctuations and Dynamic Phase Transition in a System of Self-Propelled Particles, Phys. Rev. Lett. 119, 158002 (2017).
- [46] S. Whitelam, K. Klymko, and D. Mandal, Phase separation and large deviations of lattice active matter, J. Chem. Phys. 148, 154902 (2018).
- [47] D. Chaudhuri, Active Brownian particles: Entropy production and fluctuation response, Phys. Rev. E 90, 022131 (2014).
- [48] C. Ganguly and D. Chaudhuri, Stochastic thermodynamics of active Brownian particles, Phys. Rev. E 88, 032102 (2013).
- [49] D. Mandal, K. Klymko, and M. R. De Weese, Entropy Production and Fluctuation Theorems for Active Matter, Phys. Rev. Lett. 119, 258001 (2017).
- [50] S. Shankar and M. C. Marchetti, Hidden entropy production and work fluctuations in an ideal active gas, Phys. Rev. E. 98, 020604(R) (2018).
- [51] C. Giardinà, J. Kurchan, and L. Peliti, Direct Evaluation of Large-Deviation Functions, Phys. Rev. Lett. 96, 120603 (2006).
- [52] P. I. Hurtado and P. L. Garrido, Test of the Additivity Principle for Current Fluctuations in a Model of Heat Conduction, Phys. Rev. Lett. **102**, 250601 (2009).
- [53] C. Giardina, J. Kurchan, V. Lecomte, and J. Tailleur, Simulating rare events in dynamical processes, J. Stat. Phys. 145, 787 (2011).
- [54] T. Nemoto, F. Bouchet, R. L. Jack, and V. Lecomte, Populationdynamics method with a multicanonical feedback control, Phys. Rev. E 93, 062123 (2016).
- [55] T. Nemoto, R. L. Jack, and V. Lecomte, Finite-size Scaling of a First-Order Dynamical Phase Transition: Adaptive Population Dynamics and an Effective Model, Phys. Rev. Lett. 118, 115702 (2017).
- [56] U. Ray, Garnet K.-L. Chan, and D. T. Limmer, Exact Fluctuations of Nonequilibrium Steady States from Approximate Auxiliary Dynamics, Phys. Rev. Lett. **120**, 210602 (2018).
- [57] K. Klymko, P. L. Geissler, J. P. Garrahan, and S. Whitelam, Rare behavior of growth processes via umbrella sampling of trajectories, Phys. Rev. E 97, 032123 (2018).
- [58] T. Brewer, S. R. Clark, R. Bradford, and R. L. Jack, Efficient characterization of large deviations using population dynamics, J. Stat. Mech. (2018) 053204.
- [59] L. Bertini, A. D. Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim, Macroscopic fluctuation theory for stationary nonequilibrium states, J. Stat. Phys. 107, 635 (2002).

- [60] R. L. Jack, I. R. Thompson, and P. Sollich, Hyperuniformity and Phase Separation in Biased Ensembles of Trajectories for Diffusive Systems, Phys. Rev. Lett. 114, 060601 (2015).
- [61] K. Sekimoto and S.-I. Sasa, Complementarity relation for irreversible process derived from stochastic energetics, J. Phys. Soc. Jpn. 66, 3326 (1997).
- [62] S. Toyabe, T. Okamoto, T. Watanabe-Nakayama, H. Taketani, S. Kudo, and E. Muneyuki, Nonequilibrium Energetics of a Single F<sub>1</sub>-ATPase Molecule, Phys. Rev. Lett. **104**, 198103 (2010).
- [63] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines., Rep. Prog. Phys. 75, 126001 (2012).
- [64] É. Fodor, K. Kanazawa, H. Hayakawa, P. Visco, and F. van Wijland, Energetics of active fluctuations in living cells, Phys. Rev. E 90, 042724 (2014).
- [65] É. Fodor, W. W. Ahmed, M. Almonacid, M. Bussonnier, N. S. Gov, M.-H. Verlhac, T. Betz, P. Visco, and F. van Wijland, Nonequilibrium dissipation in living oocytes, Europhys. Lett. 116, 30008 (2016).
- [66] É. Fodor, C. Nardini, M. E. Cates, J. Tailleur, P. Visco, and F. van Wijland, How Far from Equilibrium is Active Matter? Phys. Rev. Lett. 117, 038103 (2016).
- [67] A. Puglisi and U. M. Bettolo Marconi, Clausius relation for active particles: What can we learn from fluctuations, Entropy 19, 356 (2017).
- [68] U. M. B. Marconi, A. Puglisi, and C. Maggi, Heat, temperature and Clausius inequality in a model for active Brownian particles, Sci. Rep. 7, 46496 (2017).
- [69] V. Lecomte, C. Appert-Rolland, and F. van Wijland, Thermodynamic formalism for systems with Markov dynamics, J. Stat. Phys. 127, 51 (2007).
- [70] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland, First-order dynamical phase transition in models of glasses: An approach based on ensembles of histories, J. Phys. A: Math. Gen. 42, 075007 (2009).
- [71] R. Chetrite and H. Touchette, Nonequilibrium Microcanonical and Canonical Ensembles and Their Equivalence, Phys. Rev. Lett. 111, 120601 (2013).
- [72] See Supplemental Material at http://link.aps.org/supplemental/ 10.1103/PhysRevE.99.022605 for movies.
- [73] R. L. Jack and P. Sollich, Large deviations and ensembles of trajectories in stochastic models, Progr. Theoret. Phys. Suppl. 184, 304 (2010).
- [74] T. Bodineau and B. Derrida, Distribution of current in nonequilibrium diffusive systems and phase transitions, Phys. Rev. E 72, 066110 (2005).
- [75] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim, Current Fluctuations in Stochastic Lattice Gases, Phys. Rev. Lett. 94, 030601 (2005).
- [76] S. Vaikuntanathan, T. R. Gingrich, and P. L. Geissler, Dynamic phase transitions in simple driven kinetic networks, Phys. Rev. E 89, 062108 (2014).

- [77] P. Del Moral, Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications (Springer-Verlag, New York, 2004).
- [78] O. Pohl and H. Stark, Dynamic Clustering and Chemotactic Collapse of Self-Phoretic Active Particles, Phys. Rev. Lett. 112, 238303 (2014).
- [79] R. Matas-Navarro, R. Golestanian, T. B. Liverpool, and S. M. Fielding, Hydrodynamic suppression of phase separation in active suspensions, Phys. Rev. E 90, 032304 (2014).
- [80] A. Zöttl and H. Stark, Hydrodynamics Determines Collective Motion and Phase Behavior of Active Colloids in Quasi-Two-Dimensional Confinement, Phys. Rev. Lett. 112, 118101 (2014).
- [81] T. Nemoto and S.-I. Sasa, Thermodynamic formula for the cumulant generating function of time-averaged current, Phys. Rev. E 84, 061113 (2011).
- [82] R. Chetrite and H. Touchette, Variational and optimal control representations of conditioned and driven processes, J. Stat. Mech. (2015) P12001.
- [83] R. L. Jack and P. Sollich, Effective interactions and large deviations in stochastic processes, Eur. Phys. J: Special Topics 2351 (2015).
- [84] L. Bertini, A. D. Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim, Macroscopic fluctuation theory, Rev. Mod. Phys. 87, 593 (2015).
- [85] R. Wittkowski, A. Tiribocchi, J. Stenhammar, R. J. Allen, D. Marenduzzo, and M. E. Cates, Scalar  $\varphi$  4 field theory for active-particle phase separation, Nat. Commun. **5**, 4351 (2014).
- [86] C. Nardini, É. Fodor, E. Tjhung, F. van Wijland, J. Tailleur, and M. E. Cates, Entropy Production in Field Theories Without Time-Reversal Symmetry: Quantifying the Non-Equilibrium Character of Active Matter, Phys. Rev. X 7, 021007 (2017).
- [87] A. P. Solon, J. Stenhammar, M. E. Cates, Y. Kafri, and J. Tailleur, Generalized thermodynamics of phase equilibria in scalar active matter, Phys. Rev. E 97, 020602 (2018).
- [88] E. Tjhung, C. Nardini, and M. E. Cates, Cluster Phases and Bubbly Phase Separation in Active Fluids: Reversal of the Ostwald Process, Phys. Rev. X 8, 031080 (2018).
- [89] C. Appert-Rolland, B. Derrida, V. Lecomte, and F. van Wijland, Universal cumulants of the current in diffusive systems on a ring, Phys. Rev. E 78, 021122 (2008).
- [90] V. Lecomte, J. P. Garrahan, and F. van Wijland, Inactive dynamical phase of a symmetric exclusion process on a ring, J. Phys. A 45, 175001 (2012).
- [91] T. Brotto, G. Bunin, and J. Kurchan, Population aging through survival of the fit and stable, J. Stat. Mech.: Theory Exp. (2016) 033302.
- [92] T. Brotto, G. Bunin, and J. Kurchan, A model with Darwinian dynamics on a rugged landscape, J. Stat. Phys. 166, 1065 (2017).
- [93] R. Chetrite and H. Touchette, Nonequilibrium Markov processes conditioned on large deviations, Annales Henri Poincaré 16, 2005 (2015).