Signatures of irreversibility in microscopic models of flocking

Federica Ferretti [1,2,*] Simon Grosse-Holz [3,4] Caroline Holmes,[5] Jordan L. Shivers [6,7]
Irene Giardina,[1,2,8] Thierry Mora [9] and Aleksandra M. Walczak [9]

[1]*Dipartimento di Fisica, Università Sapienza, 00185 Rome, Italy*
[2]*Istituto Sistemi Complessi, Consiglio Nazionale delle Ricerche, UOS Sapienza, 00185 Rome, Italy*
[3]*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[4]*Institut Curie, Paris 75005, France*
[5]*Department of Physics, Princeton University, Princeton, New Jersey 08544, USA*
[6]*Department of Chemical and Biomolecular Engineering, Rice University, Houston, Texas 77005, USA*
[7]*Center for Theoretical Biological Physics, Rice University, Houston, Texas 77030, USA*
[8]*INFN, Unità di Roma 1, 00185 Rome, Italy*
[9]*Laboratoire de Physique de l'École Normale Supérieure (PSL University), CNRS,*
Sorbonne Université, Université de Paris, 75005 Paris, France

Flocking in $d = 2$ is a genuine nonequilibrium phenomenon for which irreversibility is an essential ingredient. We study a class of minimal flocking models whose only source of irreversibility is self-propulsion and use the entropy production rate (EPR) to quantify the departure from equilibrium across their phase diagrams. The EPR is maximal in the vicinity of the order-disorder transition, where reshuffling of the interaction network is fast. We show that signatures of irreversibility come in the form of asymmetries in the steady-state distribution of the flock's microstates. These asymmetries occur as consequences of the time-reversal symmetry breaking in the considered self-propelled systems, independently of the interaction details. In the case of metric pairwise forces, they reduce to local asymmetries in the distribution of pairs of particles. This study suggests a possible use of pair asymmetries both to quantify the departure from equilibrium and to learn relevant information about aligning interaction potentials from data.

I. INTRODUCTION

Irreversibility is a distinguishing feature of active systems that enables remarkable collective phenomena not seen at equilibrium. Examples include motility-induced phase separation (MIPS), in which particles with strictly repulsive interactions segregate into dilute and dense phases spontaneously [1], and flocking, in which polar systems with short-ranged ferromagnetic interactions produce large-scale collective motion, even in $d = 2$ [2,3]. These systems differ from their passive counterparts only in their constituents' self-propulsion, which can be viewed as a source of effective interactions, e.g., the effective attraction seen in MIPS [4] or effective long-range alignment in flocking.

The key ingredient for the emergence of such collective phenomena is irreversibility. Self-propulsion can bring an active system out of equilibrium by injecting energy at the microscopic scale, even in the absence of alternative sources of irreversibility, like nonreciprocity of interactions or time delays [5–8]. Yet, if the dynamics obeys detailed balance, even when particles are motile a Hohenberg-Mermin-Wagner-type theorem holds, preventing any spontaneous breakdown of the rotational symmetry of an equilibrium system in two dimensions [9]. Irreversibility is then a necessary condition for the existence of 2D flocks exhibiting true long-range order.

The departure of active matter from equilibrium has become a topic of considerable interest in recent years [10–15] and can be quantitatively addressed using various measures for the breakdown of detailed balance, like effective temperatures, violations of the fluctuation-dissipation theorem [16–20], or entropy production [21–23]. Depending on the scale of interest, this is done employing either agent-based or field-theoretical descriptions. Most of the effort so far has focused on MIPS models, especially in the phase-separated state (see [24] for a review). Much less work, mainly using fluctuating hydrodynamic models, has been done to systematically quantify irreversibility in polar dry active matter [25–27].

Nonetheless, it has been noted that polar active matter has a stronger nonequilibrium character than apolar active matter. For instance, it is known that, in contrast to scalar field theories [28–30], activity is relevant (in the RG sense) for continuous flocking models, such that the critical properties of active polar systems are different from their passive counterparts [31–34]. Additionally, it has been shown that, in contrast to MIPS models where only repulsive forces are present, when nonconservative aligning torques are introduced pressure is no longer a state function [35,36].

*Current address: Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; fferrett@mit.edu

In this paper, we quantify the departure from equilibrium in a minimal agent-based model of flocking, akin to the standard Vicsek model [37], where the system is described as an active ferromagnet composed of self-propelled spins. We focus on microscopic descriptions of flocks and study how they depart from equilibrium by means of the entropy production rate [38,39]. We first use this quantity to measure the breakdown of the time-reversal symmetry across the phase diagram, which we find is maximal in the vicinity of the order-disorder transition, where reshuffling of neighbors is most efficient. We relate irreversibility to asymmetries in the steady-state distribution, the details of which depend on the interactions of the model. We propose that signatures of irreversibility in the steady-state distributions may then reveal relevant information about aligning interaction potentials in biological polar active systems.

II. STOCHASTIC THERMODYNAMICS OF THE VICSEK MODEL

We consider a continuous-time variant of the original 2D Vicsek model, where the system is described as a set of interacting active Brownian particles (ABPs) performing rotational but not translational diffusion, which self-propel at a fixed speed v_0 and align to each other through short-ranged ferromagnetic interactions. The stochastic equation of motion reads

$$d\mathbf{x}_i = v_0 \mathbf{e}(\theta_i) \, dt, \tag{1}$$

$$d\theta_i = F_i(\mathbf{X}, \boldsymbol{\Theta}) \, dt + \sqrt{2D} \, dW_i, \tag{2}$$

where $W_i(t)$ is a set of independent Wiener processes for $1 \leqslant i \leqslant N$, $\mathbf{e}(\theta) = (\cos\theta, \sin\theta)$ is the orientation vector in $d = 2$ and $F_i(\mathbf{X}, \boldsymbol{\Theta})$ is the torque that reorients the ith particle. Since we are interested in quantifying the effect of self-propulsion alone, we assume the torques are symmetric, ensuring that the action-reaction principle holds. In this way, the only source of irreversibility is particle motility. We note that models typically employed to simulate flocks, like the standard Vicsek model [37,40] or topological variants with fixed number of interacting neighbors [41–43], involve nonreciprocal interactions, generating additional irreversible phase space currents. Therefore we choose

$$F_i(\mathbf{X}, \boldsymbol{\Theta}) = -\frac{\partial \mathcal{H}_{XY}(\boldsymbol{\Theta}; \mathbf{n}(\mathbf{X}))}{\partial \theta_i}, \tag{3}$$

where

$$\mathcal{H}_{XY}(\boldsymbol{\Theta}; \mathbf{n}) = -\frac{J}{2} \sum_{ij} n_{ij} \cos(\theta_i - \theta_j) \tag{4}$$

is the Hamiltonian of an *XY* model defined on a graph with a given adjacency matrix **n**. Contrarily to the classical XY model, **n** is not constant but depends on time [through the **X** variables, Eq. (1)].

The dynamics described by Eqs. (1)–(2) is Markovian in the phase space of the N-body system, whose general coordinate is $z = (\mathbf{X}, \boldsymbol{\Theta})$, where $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ is the set of particle positions and $\boldsymbol{\Theta} = (\theta_1, \ldots, \theta_N)$ is that of velocity directions. We recall the definition of the average entropy production as the Kullback-Leibler divergence between the path

probability of a stochastic trajectory $z(t)$, for $0 < t < \tau$, and its time-reversed $z^\dagger(t)$. For a general Markov process, where $p[z(t)] = p[z(t)|z(0)]p_0[z(0)]$, the average entropy production is decomposed as follows, in the absence of external driving:

$$\mathcal{S}(\tau) = D_{KL}\{p[z(t)]||p[z^\dagger(t)]\} = \mathcal{S}^{hk}(\tau) + \Delta\mathcal{S}_0. \tag{5}$$

We denote by $\mathcal{S}^{hk} = \langle \log p[z(t)|z(0)] \rangle - \langle \log p[z^\dagger(t)|z^\dagger(0)] \rangle$ the housekeeping entropy production [44], with $\langle \cdot \rangle$ the average over the ensemble of (forward) trajectories, and by $\Delta\mathcal{S}_0 = \langle \log p[z(0)] \rangle - \langle \log p[z^\dagger(0)] \rangle$ the variation in the Shannon entropy of the initial conditions of forward and backward paths. Let $\dot{\mathcal{S}}$ be the entropy production *rate* (EPR), defined from (5) as

$$\dot{\mathcal{S}} = \lim_{\tau \to \infty} \mathcal{S}(\tau)/\tau. \tag{6}$$

In order to compute this quantity for the process in Eqs. (1)–(2), we need to identify the parity under time reversal (T) of all the state variable coordinates. Positions are time-reversal symmetric, i.e., $\mathbf{x}(t) \mapsto \mathbf{x}^\dagger(t) = \mathbf{x}(\tau - t)$, while we assume $\mathbf{e}(\theta)$ is time-reversal antisymmetric, $\mathbf{e}[\theta(t)] \mapsto \mathbf{e}[\theta(t)]^\dagger = -\mathbf{e}[\theta(\tau - t)]$, so that the angular degrees of freedom acquire a global π offset: $\theta^\dagger(t) = \theta(\tau - t) + \pi$. Let us notice that, in the absence of a translational diffusion term in Eq. (1), this choice for the parity of $v_0 \mathbf{e}(\theta)$—typically identified as the *self-propulsion*—is mandatory, since it must be interpreted as a physical velocity. Adopting, on the contrary, a T-even prescription would lead to a diverging EPR, due to the deterministic nature of Eq. (1) (see Appendix A 1).

Provided that orientations are T-odd, the first equation is perfectly reversible and does not contribute to $\dot{\mathcal{S}}$. We can therefore eliminate the positional degrees of freedom and transform them into external parameters controlling the temporal evolution of $\mathbf{n}(t) = \mathbf{n}[\mathbf{X}(t)]$. In this way, the reference framework becomes that of a quasistatically driven Langevin process, with state variables $\boldsymbol{\Theta}(t)$ and a set of time-varying parameters $\mathbf{n}(t)$ [45,46]. We remark that no separation of scales between the dynamics of $\mathbf{n}(t)$ and $\boldsymbol{\Theta}(t)$ is explicitly assumed, but the protocol $\mathbf{n}(t)$ can be considered as "effectively quasistatic" in the sense that, as we verified numerically, it ensures that the many-body system is in the same stationary state at the beginning and at the end of its evolution. This property lets the system behave as if it was in local equilibrium [47] and will be crucially exploited later.

Introducing the Onsager-Machlup action of the process [48] into the formula of the housekeeping EPR and assuming stationarity, we obtain (see Appendix A 1)

$$\dot{\mathcal{S}} = \frac{1}{D} \sum_i \langle \dot{\theta}_i(t) \circ F_i(t) \rangle$$
$$= -\frac{J}{D} \sum_{ij} \langle \dot{\theta}_i(t) \circ n_{ij}(t) \sin[\theta_i(t) - \theta_j(t)] \rangle, \tag{7}$$

where $\circ$ indicates the Stratonovich prescription. According to the definition of stochastic heat of Langevin processes [49], Eq. (7) can also be interpreted as $\dot{\mathcal{S}} = D^{-1} \langle đq/dt \rangle$, with $đq$ the infinitesimal amount of heat dissipated into the medium (conventionally positive). The parameter D corresponds to both the rotational diffusion coefficient of the ABP and the

temperature of the heat bath that the active ferromagnet is in contact with. We recall that the stochastic heat we consider is not a physical heat, and the EPR as defined here is not the physical entropy: only when dealing with real thermodynamic systems and baths, an identification with the produced physical entropy is possible.

Another expression for the EPR can be obtained by noticing that to write Eq. (7) we exploited the ergodicity of the dynamics. Reintroducing an explicit time average allows us to integrate by parts (see Appendix A 2), yielding

$$
\dot{\mathcal{S}} = \lim_{\tau \to \infty} \frac{\langle -\mathcal{H}_{XY}(\boldsymbol{\Theta}(\tau); \mathbf{n}(\tau)) + \mathcal{H}_{XY}(\boldsymbol{\Theta}(0); \mathbf{n}(0)) \rangle}{D\tau}
$$
$$
- \frac{J}{2D} \sum_{ij} \langle \dot{n}_{ij}(t) \circ \cos[\theta_i(t) - \theta_j(t)] \rangle. \tag{8}
$$

The first term in Eq. (8) is proportional to the rate of change of the system's internal energy and vanishes under the assumption of stationarity. The second term corresponds to the work done on the system per unit time, divided by D. This interpretation is sound within the "external protocol" framework described above, where

$$
d\!\!{}^{\_}w_{\text{resh}} = -\frac{J}{2} \sum_{ij} dn_{ij}(t) \circ \cos[\theta_i(t) - \theta_j(t)]
$$
$$
= \sum_{ij} dn_{ij}(t) \circ \frac{\partial \mathcal{H}_{XY}(\boldsymbol{\Theta}; \mathbf{n})}{\partial n_{ij}}(t) \tag{9}
$$

represents the infinitesimal work of fictitious reshuffling forces which rewire the adjacency matrix. When the system is in a steady state, the average internal energy is constant, so the EPR includes only contributions from this irreversible work.

The fact that the dissipated heat coincides with the irreversible work of the external (fictitious) forces is typical of systems satisfying the local detailed balance condition [47]. Although there is no a priori general reason for this condition to be valid for a Vicsek-like model, the equality between Eq. (7) and the second term of Eq. (8) has been numerically verified in the whole phase diagram of all the variants of the model we studied [inset in Fig. 1(i) for Model I]. For the class of models described by Eqs. (1)–(2) the equality of Eq. (7) and Eq. (8) follows indeed from the combination of the Hamiltonian structure of the aligning force (see Appendix A 2), and the stationarity of the system [according to which the first term of Eq. (8) vanishes].

Generalizations of the EPR formulas to $d > 2$ are obtained in Appendix A 3.

III. NUMERICAL RESULTS

Let us now parametrize the connectivity matrix $\mathbf{n}(\mathbf{X})$ and use the formulas above to compute the EPR from numerical simulations of the model. We implement two variants of the Langevin-Vicsek model in Eq. (1)–(2) with short-range ferromagnetic interactions. In the first one (Model I) we model a

metric pairwise alignment with

$$
n_{ij}^{(\mathrm{I})}(\mathbf{X}) = \Theta(R - |\mathbf{x}_i - \mathbf{x}_j|), \tag{10}
$$

in which Θ is the Heaviside step function. The second variant (Model II) implements a topological multiparticle interaction, with

$$
n_{ij}^{(\mathrm{II})}(\mathbf{X}) = \begin{cases} 1 & \text{if } i, j \text{ Voronoi neighbors} \\ 0 & \text{otherwise.} \end{cases} \tag{11}
$$

At each time step, we build a tessellation of the periodic plane domain associated to the current particle configuration using the CGAL library [50]. Voronoi neighbors are pairs of particles belonging to adjacent cells in this tessellation.

Both models are known to exhibit a phase transition from a disordered isotropic phase to a polar ordered phase, but with different phenomenology. Model I undergoes a first order phase transition, where sharp phase coexistence is realized in a wide portion of the ordered phase [see Figs. 1(a) and 1(c) and [51]]. On the contrary, spatial heterogeneities are largely suppressed in Model II, and the transition seems to be of second order at the considered system sizes [Figs. 1(b) and 1(d)]. For a discussion on the nature of the transition of active models with metric-free interactions see [52–54].

Phase diagrams for the modulus of the polar order parameter $\Phi = \frac{1}{N} |\sum_{i=1}^{N} \mathbf{e}(\theta_i)|$ and for the EPR are plotted in Figs. 1(e)–1(l). We observe that, as control parameters are varied, the two models depart from equilibrium with roughly the same qualitative behavior: the entropy production rate peaks at intermediate D values, while it vanishes as $D \to \infty$ or $D \to 0$. A heuristic explanation for this behavior is readily provided if we recall that in the considered class of models nonequilibrium effects are entirely due to the rewiring of the interaction network.

The existence of the first equilibrium limit is not surprising [55]: at $D \to \infty$ the system behaves as an ideal gas of free ABPs. The existence of a second equilibrium limit at $D \to 0$ is less trivial, because of the presence of a factor D^{-1} in Eqs. (7)–(8). To have $\dot{\mathcal{S}} \to 0$, irreversible reshuffling [numerator of Eq. (8)] must occur on timescales that diverge faster than D^{-1}, corresponding to the ABP persistence time. The reference model in this second limit is a perfectly ordered passive ferromagnet, where reshuffling is suppressed by the strong alignment of particle velocities. This result justifies the use of approximate equilibrium descriptions for strongly interacting biological systems operating in this phase [56].

A simple argument can be made to predict a power-law decay of the EPR in the two equilibrium limits, as detailed in Appendix A 5. By approximating the amplitude of correlations and reshuffling times in these two regimes, we deduce from Eq. (8) that the expected EPR scaling at high D is $\dot{\mathcal{S}} \sim D^{-2}$, whereas at low D the EPR must scale as $\dot{\mathcal{S}} \sim D^{1/2}$. Figures 1(l) and 1(h) show the agreement between these predictions and numerical results for Model II (both regimes) and Model I (high D regime only).

Model-specific features in the EPR curves of Model I and Model II are also visible. For instance, the two variants of the Vicsek model show an opposite trend of the EPR with the average density of the system, $\rho = N/L^2$, as shown in Figs. 1(g) and 1(k). In Model I an increase in ρ pushes the system
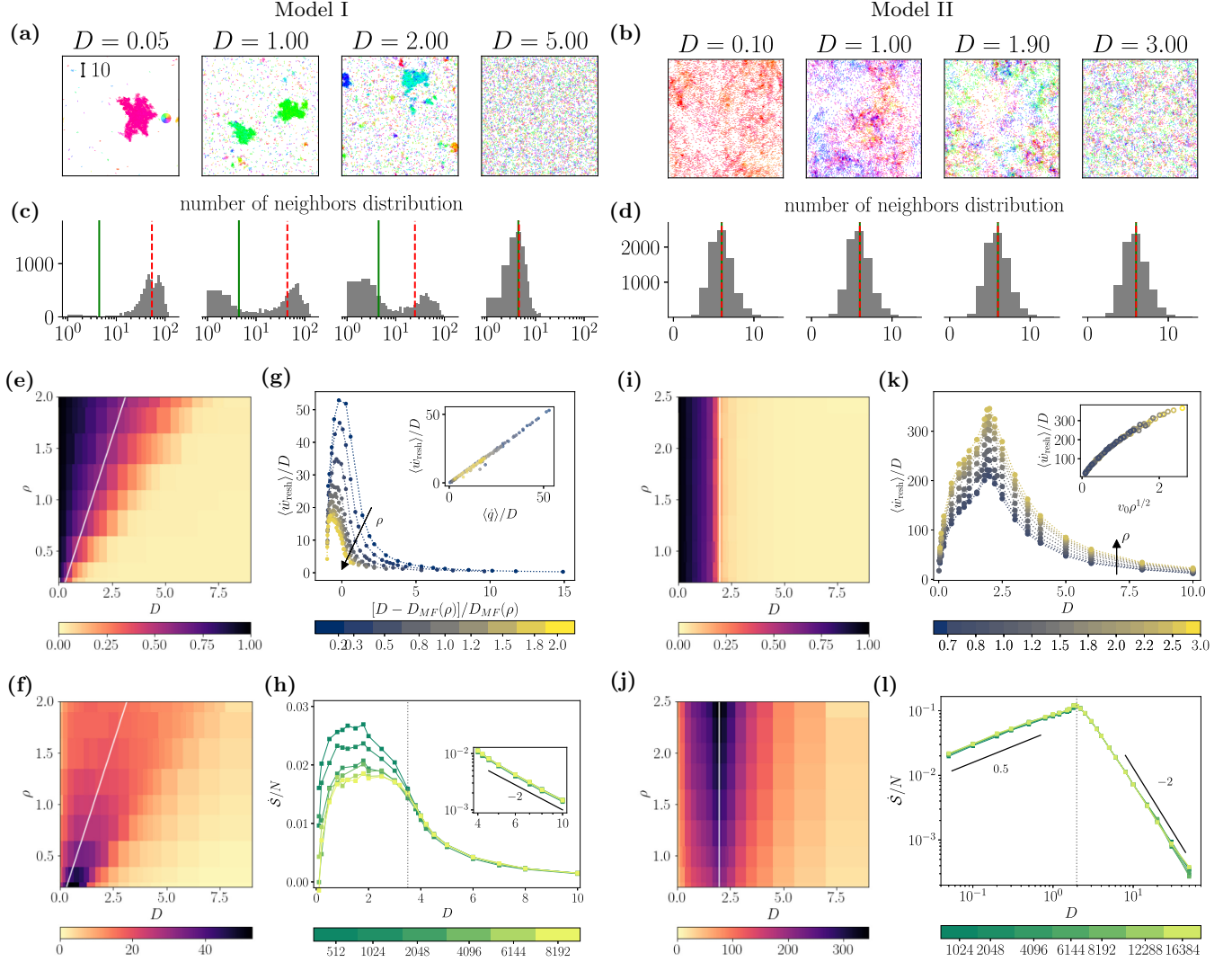
FIG. 1. [(a), (b)] Typical configurations for Model I and Model II at different D values. [(c), (d)] Distribution of the number of interacting neighbors associated to the configurations above. Bimodality at moderate to small D values indicates phase coexistence in Model I. For Model II the distribution has an almost invariant shape, hence the amount of irreversible reshuffling of the interaction network cannot be deduced from its static connectivity alone. Red dashed line: mean number of neighbors; green solid line: expected number of neighbors for a Poisson point pattern [$\pi \rho R^2 + 1$ in the metric case (c), 6 in (d)]. (e) Polar order parameter (OP) of Model I. The white line is the mean field transition point [40]. (f) EPR across the phase diagram of Model I. (g) EPR curves as a function of the rescaled distance from the mean field transition point, $D_{MF} = J\rho\pi R^2/2$. A possible explanation for the remaining density dependence is provided in the main text. Inset: Equivalence of formulas (7) and (8) is verified, guaranteeing stationarity of the observed processes. (h) EPR per particle (Model I). Color code is associated with the system size N. Strong finite-size effects are evident for small N in the symmetry-broken phase (where coexistence is realized), but $\dot{S}/N$ seems intensive as N is increased. The dashed vertical line represents the spinodal (extracted from the number of neighbors' distribution for $N = 8192$). (i) OP of Model II. (j) EPR across the phase diagram of Model II. The white line represents the alleged critical point. (k) Density dependence of the EPR curves (discussed in the main text). Inset: irreversibility is governed by a single control parameter, $v_0\rho^{1/2}$, at fixed D. (l) EPR per particle for Model II (different colors for different N values). All curves perfectly collapse on a single master curve, which converges to the two equilibrium limits as predicted. Simulation parameters: (Model I) $R = 1$, $J = 1$, $v_0 = 0.5$; when not explicitly indicated $N = 1024$, $\rho = 1$. (Model II) $J = 1$, $v_0 = 0.5$; when not explicitly indicated $N = 2048$.

towards an equilibrium-like mean field limit, by increasing the average number of neighbors and reducing the effect of reshuffling. Conversely, increasing ρ in Model II causes an effective increase in the activity of the system. Since the average number of Voronoi neighbors does not depend on the density, but density modifies typical interparticle distances, larger ρ is equivalent to larger self-propulsion speed v_0. As shown in the inset of Fig. 1(k), the EPR of Model II is in fact governed by a single control parameter, $v_0\rho^{1/2}$.

The EPR curves of the two models also show a different shape and different sensitivity to finite-size effects [cf. Figs. 1(g) vs 1(k) and Figs. 1(j) vs 1(l)]. In the topological Voronoi case (Model II), the EPR curves exhibit a kink at the transition point, similar to that observed in [15] for the "flocking dissipation rate" of the active Ising model, and a perfect collapse when rescaled by N. In the metric case, EPR curves are smoother and tend to flatten in the coexisting region as N is increased. We notice indeed that at low to intermediate

values of the D parameter, where the two types of curves are furthest from each other, Model I and Model II mostly differ in the features of their typical macroscopic configurations—most strikingly, the presence or absence of phase coexistence (at least at the observed sizes). This observation corroborates the recently proposed idea that structure and dissipation are deeply interrelated in active matter [57,58].

As a final note, the use of formulas (7) and (8) suggests that dissipation has a local origin in flocking models, explaining its extensive nature, but it does not seem to be spatially *localized*, even if the system exhibits phase coexistence. In contrast to MIPS models, we could not observe a spatial segregation of the particles that relates to their dissipation [21,22,59]. On the contrary, the heat or work contributions per single particle fluctuate in time, assuming both positive and negative values, with an amplitude comparable to the fluctuations of the mean EPR, rescaled by $\sqrt{N}$ (see Fig. 6 in Appendix A 6). However, since the EPR is a global quantity and its local decomposition is nonunique (we can always add a state function with a well-defined steady-state average value), we cannot exclude that different rewritings of $\dot{S}$ could unveil alternative interesting interpretations.

IV. SIGNATURES OF IRREVERSIBILITY

Explicit expressions for the EPR of Vicsek-like models allow us to identify model-dependent signatures of irreversibility in the steady-state distribution of a flock. For the sake of simplicity, we focus on Model I and compute

$$\dot{n}_{ij}^{(I)} = -\delta(R - |\mathbf{x}_i - \mathbf{x}_j|)\frac{(\mathbf{x}_i - \mathbf{x}_j) \cdot (\mathbf{v}_i - \mathbf{v}_j)}{|\mathbf{x}_i - \mathbf{x}_j|}. \quad (12)$$

The rewiring of the connectivity matrix in Model I depends only on how mutual distances between pairs of particles evolve. Let us insert Eq. (12) into Eq. (8) and assume stationarity to rewrite the EPR as

$$\dot{S} = \frac{Jv_0}{2D}\sum_{ij}\langle\cos\varphi_{ij}[\cos\hat{\alpha}_{ij} - \cos(\hat{\alpha}_{ij} - \varphi_{ij})]\rangle_{|\mathbf{x}_i - \mathbf{x}_j| = R}, \quad (13)$$

where $\langle\cdot\rangle_{|\mathbf{x}_i - \mathbf{x}_j| = R}$ is the conditional average over pairs of particles at distance R. The variables φ_{ij}, $\hat{\alpha}_{ij}$ are angles parametrizing the mutual alignment and the relative angular position of the two particles, respectively (see Fig. 2):

$$\varphi_{ij} = (\theta_j - \theta_i) \bmod 2\pi, \quad \hat{\alpha}_{ij} = (\alpha_{ij}^0 - \theta_i) \bmod 2\pi, \quad (14)$$

where α_{ij}^0 is the angle indicating the direction of the displacement vector $\mathbf{r}_{ij} = \mathbf{x}_j - \mathbf{x}_i$ in a fixed reference frame, in which the ith particle's orientation is θ_i. We can symmetrize Eq. (13) by introducing $\alpha_{ij} = \hat{\alpha}_{ij} - \varphi_{ij}/2$ and rewrite

$$\dot{S} = \frac{Jv_0}{2D}N^2 g(R)\int_{[0,2\pi]^2} d\alpha\, d\varphi\, q(\alpha, \varphi)\varepsilon(\alpha, \varphi), \quad (15)$$

where $g(r) = \frac{1}{N^2}\langle\sum_{ij}\delta(|\mathbf{x}_j - \mathbf{x}_i| - r)\rangle$ is the pair correlation function and

$$\varepsilon(\alpha, \varphi) = \cos\varphi\left[\cos\left(\alpha - \frac{\varphi}{2}\right) - \cos\left(\alpha + \frac{\varphi}{2}\right)\right] \quad (16)$$
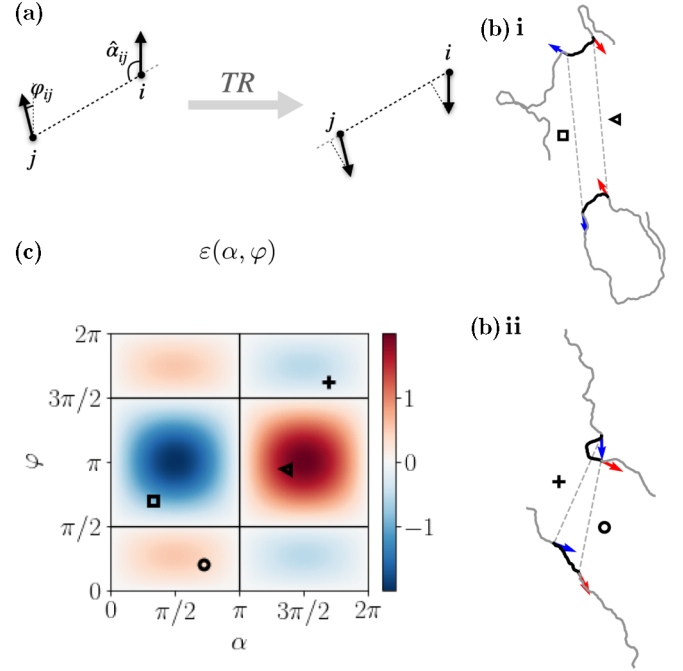

FIG. 2. (a) The time-reversal (TR) operator acts on the system's variables by flipping velocities and keeping positions unchanged. Out of equilibrium, the two configurations cannot be equally probable. (b) Sample trajectories from a system of $N = 1024$ particles ($D = 3$, $J = 1$, $v_0 = 0.5$, $\rho = 1$, $\Phi \simeq 0.21$) elucidating the dissipation mechanism in Model I. Black stretches indicate the portion of trajectory in which the two particles are at a distance smaller than the interaction radius; gray stretches those where particles are at a distance $r > R$ and do not interact. In (b.i) strongly antialigned particles start interacting with an initial phase difference of 0.95π (triangle, converging red arrows, corresponding to a positive EPR contribution $\varepsilon_{\text{in}} \simeq 1.76$). Due to rotational diffusion, they later leave each other's neighborhood at a new relative phase and displacement angle (square, blue arrows), corresponding to $\varepsilon_{\text{out}} \simeq -0.92$. In (b.ii) converging particles (cross) enter the interaction disk contributing to the EPR negatively ($\varepsilon_{\text{in}} = -0.34$). After the interaction, they diverge being more aligned than before (circle), so that the EPR contribution associated to this configuration surmounts the previous one: $\varepsilon_{\text{out}} = 0.38$. (c) Contour plot of Eq. (16).

is proportional to an EPR density per pair of particles. The quantity $q(\alpha, \varphi)$ is the (normalized) distribution of particle pairs at distance R:

$$q(\alpha, \varphi) = \frac{\langle\sum_{ij}\delta(|\mathbf{r}_{ij}| - R)\delta(\hat{\alpha}_{ij} - \varphi_{ij}/2 - \alpha)\delta(\varphi_{ij} - \varphi)\rangle}{N^2 g(R)}. \quad (17)$$

The time-reversal operator acts on the newly introduced angular variables in the following way:

$$\varphi^\dagger(t) = \varphi(\tau - t), \quad \alpha^\dagger(t) = \alpha(\tau - t) + \pi. \quad (18)$$

Hence, breakdown of the time-reversal symmetry, $\dot{S} \neq 0$, implies from Eq. (15) a symmetry breaking in the pair distribution: $q(\alpha, \varphi) \neq q(\alpha + \pi, \varphi)$. This means that two mirror configurations like those in Fig. 2(a) cannot be equally probable. Specifically, since the EPR is nonnegative, for aligned pairs ($\cos\varphi > 0$) configurations with diverging particles are
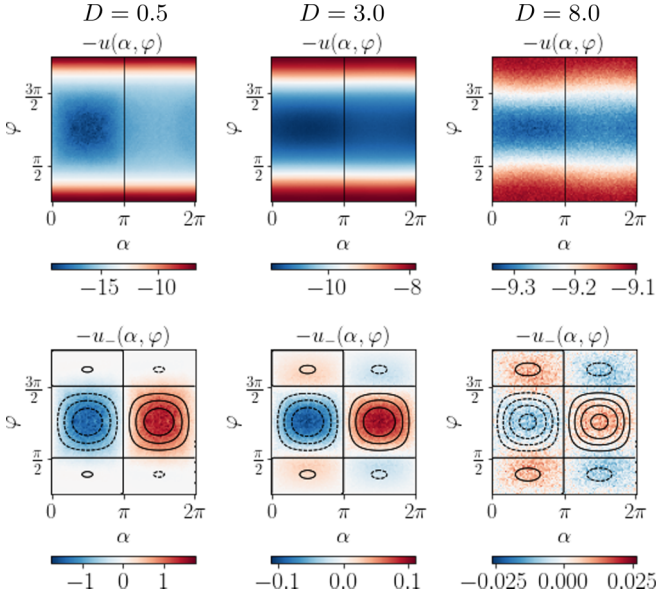
FIG. 4. Scalar measure for the degree of asymmetry of the log distribution of particle pairs at distance r: $\|u_-\|_1(r) = \int d\alpha \, d\varphi |u_-(\alpha, \varphi; r)|$. Curves peak at $r = R$ (dashed line) for all the considered D values. Lowering the rotational diffusion coefficient, the decay to zero at large r gets slower because of the presence of heterogeneous structures affecting the pair distribution. The trend of these curves is qualitatively reminiscent of $-\partial_r n(r) = \delta(r - R)$. All data are collected form simulations of systems of $N = 2048$ particles.

FIG. 3. First row: Reconstructed log distributions of pairs of particles at distance R from numerical simulations of Model I ($N = 1024$, $v_0 = 0.5$, $\rho = 1$, $J = 1$). Second row: Antisymmetric part of the log distributions: $-u_-(\alpha, \varphi) = u(\alpha + \pi, \varphi) - u(\alpha, \varphi)$. A positive correlation with $\varepsilon(\alpha, \varphi)$, shown in Fig. 2(c), is evident. Black lines represent the level curves of the fitted $\lambda\varepsilon(\alpha, \varphi)$ function, with λ free parameter.

expected to be more probable than ones with converging particles. On the contrary, for antialigned pairs ($\cos\varphi < 0$), configurations with colliding particles are expected to be more probable than ones with divaricating particles [cf. Fig. 2(b)]. This scenario reflects the fact that birds leaving each other's neighborhood have been interacting in the past and are typically more aligned than those which have not interacted directly before.

It is convenient to consider the logarithm of the particle pair distribution function, $u(\alpha, \varphi) = -\log q(\alpha, \varphi)$ and decompose it into T-symmetric and T-antisymmetric parts: $u_\pm(\alpha, \varphi) = \frac{1}{2}[u(\alpha, \varphi) \pm u(\alpha + \pi, \varphi)]$. The irreversibility condition now reads $u_-(\alpha, \varphi) \neq 0$, while Eq. (15) is rewritten as

$$\dot{S} \propto \int_{[0,2\pi]^2} d\alpha \, d\varphi \, e^{-u_+(\alpha,\varphi)} \sinh\left[-u_-(\alpha, \varphi)\right]\varepsilon(\alpha, \varphi). \quad (19)$$

The positivity of the EPR is translated into a positive correlation between the log distribution of particle pairs $-u_-(\alpha, \varphi)$ and the EPR density $\varepsilon(\alpha, \varphi)$. We show in Fig. 3 the numerically reconstructed function $-u_-(\alpha, \varphi)$, for different parameter values. In the disordered phase, the log distribution looks almost T-symmetric, as the system is close to equilibrium (large D in Fig. 3, second row). In the ordered phase nonnegligible asymmetries are visible in the reconstructed distributions: these are especially concentrated in the low-probability region where particles are antialigned ($\varphi \sim \pi$).

It is worth remarking that the discussed features are *local*, as they are observed at the scale of the interaction radius R, which is much smaller than the system size or the typical size of polar clusters in the ordered phase. Asymmetries gradually disappear when we look at larger scales, as shown in
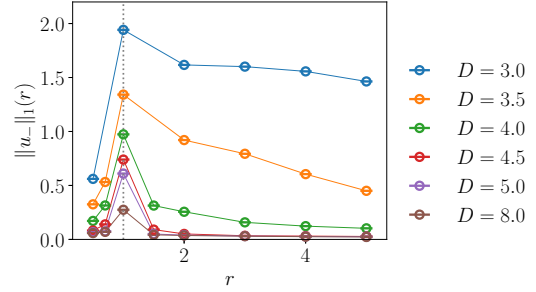
Fig. 4. This property is not specific to Model I, but holds independently of the precise parametrization of $\mathbf{n}(\mathbf{X})$, provided that it describes a pairwise short-ranged interaction. Interestingly, a similar asymmetric scenario has also been observed in [60] in a variant of Model I with nonadditive, nonpairwise interactions.

Model II is another example of a system with nonpairwise interactions, since $n_{ij}(\mathbf{X})$ is not a simple function of the mutual distance between particles i and j. As a consequence, we do not have a simple expression for $\dot{n}_{ij}$ to fill in Eq. (8). Since the number of Voronoi cells must be conserved, reshuffling can occur only via the formation of m-fold vertices, with $m \geqslant 4$. Such transitional configurations are realized when the particles form an m-sided polygon inscribed in a circle, so signatures of irreversibility must be sought for in m-particle densities (especially $m = 4$, since other transitional configuration than the lowest order fourfold vertex are unlikely to occur).

V. IRREVERSIBILITY-INDUCED EXPLICIT SYMMETRY BREAKING

The fact that irreversibility constrains asymmetries in the two-body distribution (for Model I) is an example of a more general result concerning systems with T-odd state variable coordinates. Following the line of reasoning of [61], it can be shown that for a whole class of stationary Langevin processes, the irreversibility condition implies a symmetry breaking in the nonequilibrium steady-state distribution of the system. This symmetry breaking only depends on the parity of the state variable coordinates under time reversal and it is easy to predict, without solving the steady-state Fokker-Planck equation. Nonetheless it may be typically hard to observe because it requires reconstructing the full steady-state distribution, while it is not guaranteed to survive projection, coarse graining, or any other dimensional reduction of the variable space.

Let us consider a stationary additive process $\mathbf{z}(t)$ described by a Langevin equation

$$\dot{z}_\alpha = A_\alpha(\mathbf{z}) + B_{\alpha\beta}\xi_\beta, \quad (20)$$

whose coordinates have a definite parity under time reversal, i.e., $z_\alpha^\dagger(t) = \epsilon_\alpha z_\alpha(\tau - t)$ with $\epsilon_\alpha = \pm 1$. The condition $\dot{\mathcal{S}} > 0$ implies that at least one of the two following statements is violated (see Appendix B):

$$A_\alpha^{\mathrm{irr}}(\mathbf{z}) + D_{\alpha\beta}\partial_\beta\phi_+(\mathbf{z}) = 0 \ \ \forall\alpha, \tag{21}$$

$$\partial_\alpha\phi_-(\mathbf{z}) = 0 \ \ \forall\alpha. \tag{22}$$

Here $\mathbf{D} = \frac{1}{2}\mathbf{B}^\top\mathbf{B}$ is the diffusion matrix of process (20); $\mathbf{A}^{\mathrm{rev}}$ and $\mathbf{A}^{\mathrm{irr}}$ indicate the common decomposition of the drift term $\mathbf{A}$ into a reversible and an irreversible part [44]; $\phi_\pm$ are the T-symmetric and T-antisymmetric parts of the quasipotential $\phi = -\log\psi(\mathbf{z})$, where $\psi(\mathbf{z})$ is the nonequilibrium steady-state (NESS) distribution of the system.

When the state variable contains only even coordinates with respect to time reversal, Eq. (21) must be violated, since $\phi_-(\mathbf{z}) = 0$. In contrast, when $\mathbf{A}^{\mathrm{rev}}(\mathbf{z}) = 0$, the stationarity condition, combined with irreversibility, implies $\phi_-(\mathbf{z}) \neq 0$ (see Appendix B). This means that any Langevin additive process with null reversible drift and having coordinates in the state variable that change sign under time reversal exhibits signatures of irreversibility in its NESS distribution.

The considered class of 2D flocking models corresponds to this second scenario, independently of the interaction matrix parametrization, as better described in Appendix B. Hence the NESS distribution must be asymmetric under time reversal:

$$\log\psi(\mathbf{X}, \mathbf{\Theta}) \neq \log\psi(\mathbf{X}, \mathbf{\Theta} + \pi). \tag{23}$$

The discrete T symmetry which is here explicitly broken can be viewed as a π rotation in the velocity subspace $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_N)$, where $\mathbf{v}_k = e^{i\theta_k}$. Therefore, a continuous rotational symmetry in the velocity subspace cannot hold; we can think of this fact as an irreversibility-induced explicit breakdown.

However, a different symmetry is preserved: the Fokker-Planck operator and the NESS distribution are invariant if arbitrary identical rotations are performed on both the external space of positions and the internal space of the polar order parameter:

$$\psi(e^{i\theta_0}\mathbf{X}, \mathbf{\Theta} + \theta_0) = \psi(\mathbf{X}, \mathbf{\Theta}) \quad \forall\theta_0 \in \mathbb{R}. \tag{24}$$

As a consequence of Eq. (24), the marginalized pdf $\psi(\mathbf{\Theta}) = \int_{[0,L]^{2N}} d\mathbf{X}\,\psi(\mathbf{X}, \mathbf{\Theta})$ is rotationally invariant, even when the dynamics of the polar active system is irreversible. Similarly, the marginalized distribution of positions, $\psi(\mathbf{X}) = \int_{[0,2\pi]^N} d\mathbf{\Theta}\,\psi(\mathbf{X}, \mathbf{\Theta})$, is invariant under rotations.

In conclusion, the above discussion does not compel explicit symmetry breaking in $\psi(\mathbf{\Theta})$, but underlines the lack of continuous symmetry under velocity rotations in the coupled $(\mathbf{X}, \mathbf{V})$ space, which appears as a hypothesis in known Hohenberg-Mermin-Wagner theorems (including [9]). In our Langevin-Vicsek model, this symmetry is recovered only at equilibrium, where the absence of reshuffling disentangles positions and velocities.

VI. CONCLUSION

Real-life flocks and flocking models are often given as examples of strongly out of equilibrium systems. However,

analysis of real flocks has showed that they can function close to equilibrium if the self-propulsion leads the interaction network to rearrange on slow timescales compared to the local orientational dynamics [56]. Motivated by the observation that self-propulsion and irreversibility are not always synonymous and that the former alone is not a sufficient condition to explain the spontaneous emergence of collective behavior in polar active matter [9], we measured how flocks depart from equilibrium across their phase diagram. We employed the entropy production rate as the natural quantifier for the breakdown of detailed balance and exploited its positiveness to identify signatures of irreversibility in minimal agent-based flocking models.

While general statements can be deduced from the parity of the state variables alone, the way such signatures are manifested conveys model-specific information about alignment interactions in the flock. For the considered class of Vicsek-like models, violations of the time-reversal symmetry are indeed due to the interplay between self-propulsion and the (otherwise equilibrium-like) alignment interactions. Self-propulsion makes particles motile, but interaction is required to probe the effect of motility, through the rewiring of the interaction network.

It is worth noting that rewiring could in principle be reversible, if it occurs in a symmetric way. However, the way information is transferred from velocity to positional degrees of freedom prevents this from happening. The evolution of bird positions can be seen as an effective external protocol for the dynamics of the orientation degrees of freedom. The protocol is, however, not arbitrary: it must ensure that the *XY* Hamiltonian is on average stationary—a condition that implies a feedback between the state of the driven system and the driving protocol. A better understanding of the thermodynamic and information-theoretical meaning of the rewiring of the interaction network may help understand not only the microscopic origin of the symmetry breaking out of equilibrium, but also how to control a flocking system in any thermodynamic phase.

Since irreversibility consists in a symmetry breaking (with respect to the time-reversal transformation), it is not surprising that signatures of the out-of-equilibrium nature of the dynamics can be detected in asymmetries of the steady-state distribution of the system. When the collective dynamics emerges from reciprocal pairwise interactions, such signatures of irreversibility are manifested in the pair distribution, at the scale of the interaction radius. Asymmetries are washed away on larger scales, if the system is sufficiently homogeneous.

Our analysis suggests that irreversibility-related features can be exploited as a tool to infer relevant information about microscopic interaction mechanisms in active polar systems from the experimental data. Numerical simulations also show that the EPR is highest in the coexistence region of the metric additive Vicsek model, and peaks at the alleged critical point in the Voronoi Vicsek model, suggesting that the effects of irreversibility in polar active matter should be mostly visible in systems lying at the onset of collective motion. Candidate experimental models could be, for instance, moderate density actomyosin motility assays [62] or insect swarms [63].

Lastly, the numerically computed EPR exhibits a finite size scaling compatible with an extensive nature of the observable. It would be interesting to investigate how the EPR and the scaling we measure from agent-based microscopic models relates to the EPR of fluctuating coarse-grained theories and its scaling [64].

APPENDIX A: DERIVATION OF EPR FORMULAS

1. Heat dissipation

The process in Eq. (1)–(2) is Markovian, so the path probability associated with it reads

$$p[\mathbf{X}(t), \mathbf{\Theta}(t)] \propto \psi(\mathbf{X}(0), \mathbf{\Theta}(0))e^{-\mathcal{A}_I[\mathbf{X}(t),\mathbf{\Theta}(t)]}, \qquad \text{(A1)}$$

where $\psi(\mathbf{X}, \mathbf{\Theta})$ is the steady-state distribution of the N-body system, and

$$\mathcal{A}_I[\mathbf{X}(t), \mathbf{\Theta}(t)] = \lim_{\epsilon \to 0} \lim_{\Delta t \to 0} \sum_{s=0}^{N_s-1} \sum_{i=1}^{N} \frac{1}{\Delta t} \left\{ \frac{1}{\epsilon^2} \left[\mathbf{x}_i^{s+1} - \mathbf{x}_i^s - v_0 \mathbf{e}(\theta_i^s)\Delta t \right]^2 + \frac{1}{4D} \left[\theta_i^{s+1} - \theta_i^s + J \sum_j n_{ij}^s \sin\left(\theta_i^s - \theta_j^s\right)\Delta t \right]^2 \right\} \quad \text{(A2)}$$

is the Onsager-Machlup (OM) action [48] (in the Itô prescription; any other α-type prescription would lead to the same final expression for the EPR, since the process is additive). N_s indicate the number of steps of amplitude Δt in which the trajectory is discretized, and the limit is taken while keeping $N_s\Delta t = \tau$. The presence of the ϵ-dependent term in Eq. (A2) comes from the representation of $\delta[\dot{\mathbf{x}}_i - v_0\hat{\mathbf{e}}(\theta_i)]$ as the singular limit of a Gaussian of variance ϵ^2. This regularization corresponds to adding an infinitesimal translational diffusion noise to the deterministic Eq. (1), in order to make the diffusion matrix of the 2D process nonsingular.

Assuming that the system is in a steady state, the variation in the Shannon entropy of the initial state distribution vanishes, $\Delta\mathcal{S}_0 = \langle \log \psi(\mathbf{X}(0), \mathbf{\Theta}(0)) - \log \psi(\mathbf{X}^\dagger(0), \mathbf{\Theta}^\dagger(0)) \rangle_0 = 0$, and the entropy production of the irreversible process reduces to the *housekeeping entropy production* [44]: $\mathcal{S}(\tau) = \mathcal{S}^{hk}(\tau) = \mathcal{A}_I[\mathbf{X}(t), \mathbf{\Theta}(t)] - \mathcal{A}_I[\mathbf{X}^\dagger(t), \mathbf{\Theta}^\dagger(t)]$, where $\mathbf{X}^\dagger(t)$ and $\mathbf{\Theta}^\dagger(t)$ indicate the time-reversed trajectories.

Using Eqs. (A2) and the OM action of the time-reversed trajectory, we can compute the entropy production *rate* and obtain a first formula:

$$\dot{\mathcal{S}} = \frac{\langle \dot{q} \rangle}{D} = -\frac{J}{D} \sum_{ij} \langle \dot{\theta}_i \circ n_{ij} \sin(\theta_i - \theta_j) \rangle$$

$$= \lim_{N_s \to \infty} \lim_{\Delta t \to 0} \frac{1}{N_s\Delta t} \sum_{i=1}^{N} \sum_{s=0}^{N_s-1} \left\langle \frac{\theta_i^{s+1} - \theta_i^s}{\Delta t} \cdot \frac{F_i^s + F_i^{s+1}}{2} \right\rangle, \tag{A3}$$

where $F_i^s = F_i(\mathbf{X}^s, \mathbf{\Theta}^s) = -J \sum_j n_{ij}(\mathbf{X}^s) \sin(\theta_i^s - \theta_j^s)$ is the torque acting on the ith particle at time step s, and $\langle \cdot \rangle$ is the average over the ensemble of nonequilibrium stationary paths. Here the limit is taken in such a way that $N_s\Delta t \to \infty$. Equation (A3) coincides with the average stochastic heat [49] dissipated by the system into a heat bath at temperature D per

unit time, divided by D or, equivalently, to the work rate of the aligning torques.

2. Local equilibrium

An alternative formula to Eq. (A3) is obtained by eliminating the spatial d.o.f.'s and projecting the process to those phase space directions where the irreversible current has nonzero components. In this reduced space, the diffusion matrix is invertible and no regularization is needed. The effect of the eliminated d.o.f.'s is taken into account by recognizing them as parameters driving the system through a transformation that does not modify the average value of the effective XY Hamiltonian (4). Having rephrased the problem in this way, we can follow [47], and proceed by discretizing the trajectories of the angular d.o.f.'s and of the protocol parameters as follows:

$$\text{direct}: \quad \mathbf{\Theta}_0 \xrightarrow{\mathbf{X}_1} \mathbf{\Theta}_1 \xrightarrow{\mathbf{X}_2} \mathbf{\Theta}_2 \cdots \mathbf{\Theta}_{M-1} \xrightarrow{\mathbf{X}_M} \mathbf{\Theta}_M,$$

$$\text{time-reversed}: \quad \mathbf{\Theta}_0 \xleftarrow{\mathbf{X}_1} \mathbf{\Theta}_1 \xleftarrow{\mathbf{X}_2} \mathbf{\Theta}_2 \cdots \mathbf{\Theta}_{M-1} \xleftarrow{\mathbf{X}_M} \mathbf{\Theta}_M.$$

Let us now assume the local detailed balance condition [47] (justified *a posteriori* by numerical results—no *a priori* assumption of time scale separation which is valid in every region of the phase diagram can be made):

$$\frac{P_c(\mathbf{\Theta}_n; \mathbf{X}_{n+1})P(\mathbf{\Theta}_n \xrightarrow{\mathbf{X}_{n+1}} \mathbf{\Theta}_{n+1})}{P_c(\mathbf{\Theta}_{n+1}; \mathbf{X}_{n+1})P(\mathbf{\Theta}_{n+1} \xrightarrow{\mathbf{X}_{n+1}} \mathbf{\Theta}_n)} = 1, \qquad \text{(A4)}$$

where

$$P_c(\mathbf{\Theta}; \mathbf{X}) = \frac{1}{Z(\mathbf{X})} e^{-\beta\mathcal{H}_{XY}(\mathbf{\Theta}; \mathbf{n}(\mathbf{X}))} \tag{A5}$$

is the equilibrium Boltzmann distribution of an XY spin system on a fixed network with connectivity matrix $\mathbf{n}(\mathbf{X})$ and $\beta = D^{-1}$. Using hypothesis (A4) to compute the entropy

production of the discretized Markov process,

$$
\begin{aligned}
\mathcal{S} &= \left\langle \log \frac{P_c(\boldsymbol{\Theta}_0; \mathbf{X}_0)}{P_c(\boldsymbol{\Theta}_M; \mathbf{X}_M)} + \log \frac{\prod_{n=0}^{M-1} P(\boldsymbol{\Theta}_n \xrightarrow{\mathbf{X}_{n+1}} \boldsymbol{\Theta}_{n+1})}{\prod_{n=0}^{M-1} P(\boldsymbol{\Theta}_{n+1} \xrightarrow{\mathbf{X}_{n+1}} \boldsymbol{\Theta}_n)} \right\rangle \\
&= \left\langle \log \prod_{n=0}^{M-1} \frac{P_c(\boldsymbol{\Theta}_n; \mathbf{X}_n)}{P_c(\boldsymbol{\Theta}_n; \mathbf{X}_{n+1})} \right\rangle,
\end{aligned}
\tag{A6}
$$

and taking the continuous limit yields for the EPR,

$$
\dot{\mathcal{S}} = \beta \left\langle \sum_{i<j} \dot{n}_{ij} \circ \left[\frac{\partial \mathcal{H}_{XY}(\boldsymbol{\Theta}; \mathbf{n})}{\partial n_{ij}} - \frac{\langle \partial \mathcal{H}_{XY}(\boldsymbol{\Theta}; \mathbf{n}) \rangle_{c,\mathbf{n}}}{\partial n_{ij}} \right] \right\rangle,
\tag{A7}
$$

where $\langle \cdot \rangle_{c,\mathbf{n}}$ is the equilibrium over the canonical ensemble, for fixed $\mathbf{n}$. The same result is obtained by applying the standard rules of stochastic calculus to Eq. (A3). Recalling that this expression involves a time average (eliminated under the assumption of ergodicity), we can perform an integration by parts resulting in

$$
\dot{\mathcal{S}} = -\frac{J}{2D} \sum_{ij} \langle \dot{n}_{ij} \circ \cos(\theta_i - \theta_j) \rangle \\
+ \lim_{\tau \to \infty} \frac{\langle \mathcal{H}_{XY}(\tau) - \mathcal{H}_{XY}(0) \rangle}{\tau}.
\tag{A8}
$$

Let us now assume that the system is in a steady state, where the average XY Hamiltonian does not vary with time (or at most sublinearly). Equation (A8) then reduces to the rate of work which fictitious reshuffling forces make on the system:

$$
\dot{\mathcal{S}} = \frac{\langle \dot{w}_{\text{resh}} \rangle}{D} = -\frac{J}{2D} \sum_{ij} \langle \dot{n}_{ij} \circ \cos(\theta_i - \theta_j) \rangle.
\tag{A9}
$$

Let us remark that these fictitious forces are not independent of the system's state: due to this crucial dependency between the external protocol and the state of the system we can have a nonzero EPR.

3. Arbitrary dimension

The generalization of the active XY model to the $d > 2$ case is the active $O(n)$ ferromagnet, whose dynamics is described by

$$
dx_i^\alpha = v_i^\alpha dt,
\tag{A10}
$$

$$
dv_i^\alpha = P_i^{\alpha\beta} \left(-J \sum_j n_{ij} v_j^\beta + v_0 \xi_i^\beta \right),
\tag{A11}
$$

where $\langle \xi_i^\alpha(t) \xi_i^\beta(t') \rangle = \delta_{ij} \delta_{\alpha\beta} \delta(t - t')$ and $P_i^{\alpha\beta}(t) = \delta^{\alpha\beta} - v_i^\alpha(t) v_i^\beta(t) / v_0^2$ is the orthogonal projector to $\mathbf{v}_i(t)$ that is required for normalization purposes. For $d > 2$ the process is multiplicative: the prescription to integrate Eqs. (A10)–(A11) is the Stratonovich one.

The corresponding generalization of the EPR formulas (A3) and (A9) to the d-dimensional case reads

$$
\dot{\mathcal{S}} = \frac{\langle \dot{q} \rangle}{Dv_0^2} = \frac{J}{Dv_0^2} \sum_{ij} \langle \dot{v}_i^\alpha(t) \circ P_i^{\alpha\beta}(t) n_{ij}(t) v_j^\beta(t) \rangle,
\tag{A12}
$$

$$
\dot{\mathcal{S}} = \frac{\langle \dot{w} \rangle}{dt} = -\frac{J}{2Dv_0^2} \sum_{ij} \langle \dot{n}_{ij}(t) \circ v_i^\alpha(t) v_j^\alpha(t) \rangle.
\tag{A13}
$$

4. Equilibrium limits

The obtained EPR formulas reveal the existence of several equilibrium limits for the Vicsek model. Two of them obviously correspond to $v_0 \to 0$, where activity is suppressed and the model corresponds to an equilibrium XY ferromagnet, and $J \to 0$. From inspection of Eq. (A3) a third equilibrium limit can be deduced, for $D \to \infty$, since torques and phase increments are bounded. In this limit, the interaction term in Eq. (2) becomes negligible, compared to the noise, and the system behaves as in the noninteracting equilibrium case. Because of the equivalence between Eqs. (A3) and (A9), this means that reshuffling is suppressed as the rotational diffusion coefficient increases. Particles tend indeed to swirl around their positions, being stuck in the vicinity of their own neighborhood for a long time (see Fig. 5).

Another equilibrium limit, which is independent of the parametrization of n_{ij}, corresponds to $D \to 0$. In the strongly polarized phase the system approaches the behavior of a passive ferromagnet in the comoving reference frame. The relevant (asymmetric) contribution of reshuffling is suppressed faster than D in this limit, resulting in an effective equilibrium behavior. An approximate argument for this fact is provided in Sec. A 5.

Two other nontrivial equilibrium limits can be deduced for the two variants of the Vicsek model considered in the main text, by varying the control parameter ρ. In the additive metric case (Model I) the $\rho \to \infty$ limit corresponds to an equilibrium limit, since the system approaches an effective mean field configuration. In the topological case (Model II) the opposite limit, $\rho \to 0$, is an equilibrium one.

Enhancement of reshuffling close to the transition is witnessed by the decay of the reconstructed autocorrelation function of the connectivity matrix:

$$
C_{\text{net}}(t) = \frac{1}{N(N-1)} \sum_{ij} \langle n_{ij}(t_0 + t) n_{ij}(t_0) \rangle_{t_0},
\tag{A14}
$$

where $\langle \cdot \rangle_{t_0}$ denotes a time average over multiple starting times. We show the quantity (A14) computed for Model II in Fig. 5. The fastest decay is observed at the alleged critical point ($D_c \simeq 2$). All the curves, for varying D, are well fitted by the functional form:

$$
C_{\text{net}}(t) \sim M(1 + ct^a)^{-d},
\tag{A15}
$$

which was empirically introduced in [65] to measure the average fraction of nonchanging neighbors after a time delay t in real flocks of birds. We fitted Eq. (A15) on numerical data by taking as fixed parameters $d = 2$ and $M = 6$ (the average degree of a Voronoi vertex in a planar graph is fixed by the Euler formula). Results from the fit are shown in Fig. 5(g) where parameters are plotted parametrically versus the average EPR of the system.
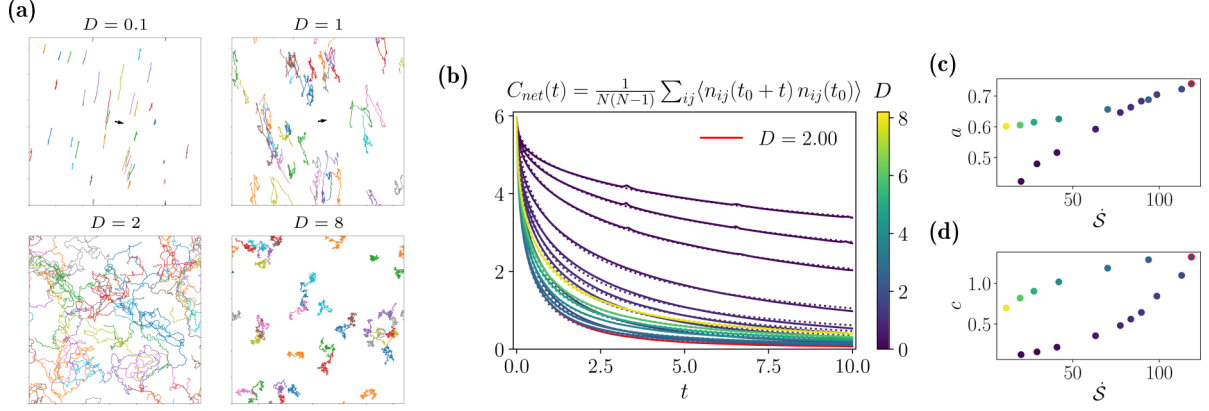
FIG. 5. (a) Subsample of trajectories of diffusing particles in different regimes. All the trajectories span the same time interval. In the symmetry broken phase the trajectories are shown in the reference frame of the center of mass; mutual diffusion occurs mainly in the transverse direction with respect to collective motion (black arrow). At $D = 2$, where reshuffling is mostly efficient, the flock is disordered but the motion of the particles is still persistent. Persistence is reduced as the rotational diffusion coefficient D is increased. In all cases, $N = 1024$, $\rho_0 = 1$. (b) Autocorrelation function of the adjacency matrix of the flock, $C_{\text{net}}(t)$, defined as in Eq. (A14). The color map refers to D values. The maximum point of the EPR, $D = 2$, is marked in red and corresponds to the curve with the fastest decay. Dashed lines are the fitted curves from Eq. (A15). [(c),(d)] Parametric plot of fitted parameters, a and c, vs EPR. The figure shows a positive correlation both for the effective diffusion coefficient c and exponent a. Close to the transition point (marked by the red dot) reshuffling is the most efficient, and the system is in the farthest condition from an equilibrium one.

5. Scaling with D

Let us focus on Eq. (A9) in the two equilibrium limits $D \to 0$ and $D \to \infty$. The parameter D enters in the formula through $\cos(\theta_i - \theta_j)$ (alignment of bird pairs that are changing their status of neighbors), $\dot{n}_{ij}$ (reshuffling rate), and the D^{-1} prefactor. Let us suppose that $\cos(\theta_i - \theta_j) \sim C(l)$, with C the (full) spatial correlation function of birds' velocities, which we assume isotropic, and l the interaction radius (in Model I) or the average distance of a pair of birds which are leaving each other's Voronoi shell (in Model II). In this paragraph the symbol $\sim$ must be read as "approximately proportional to." Let us assume that the typical reshuffling rate, $\tau_{\text{resh}} \sim \dot{n}_{ij}^{-1}$, is related to the time needed for a particle to travel the same reference distance l:

$$\langle |\Delta \mathbf{x}(\tau_{\text{resh}})|^2 \rangle \sim l^2. \tag{A16}$$

In the low-D regime the system is deeply ordered and we can adopt the spin-wave approximation. This approximation consists in linearizing the equations of motion by taking into account only the transverse fluctuations of the velocity variables (or spin wave excitations) $v_0 \boldsymbol{\pi}_i$, where $\boldsymbol{\pi}_i \perp \mathbf{V} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{v}_i$. The magnitude of these Gaussian fluctuations scales as $|\boldsymbol{\pi}_i| \sim D^{1/2}$. The corresponding low-$D$ expansion of the correlation function is $C(l) \sim 1 - |\boldsymbol{\pi}|^2 \sim 1 - D$. In Model II $\sum_{ij} \dot{n}_{ij} = 0$ because the average degree of a vertex in a Delaunay triangulation of the plane (Voronoi tessellation's dual) is exactly 6 [66], so a nonzero contribution to the EPR is only expected to come from the connected velocity correlation. In Model I a similar argument holds on average, if the system is in a stationary condition. Hence we shall replace $C(l)$ with $C_c(l) \sim D$ and estimate τ_{resh} from Eq. (A16). Since the system is strongly ordered, mutual displacement occurs mainly in the transverse space to the collective direction of motion, and we can assume that l is typically small enough to

work in the ballistic regime. The condition

$$\langle \Delta \mathbf{x}_\perp^2(t) \rangle \sim |\boldsymbol{\pi}|^2 t^2 \sim l^2 \tag{A17}$$

identifies a reshuffling timescale $\tau_{\text{resh}} \sim D^{-1/2}$, so that $\dot{\mathcal{S}} \sim D^{1/2}$.

In the high-D regime thermal noise dominates over alignment interactions. The limit model is an ideal gas of free ABPs, whose mean-squared displacement is known [67]:

$$\langle \Delta \mathbf{x}(t)^2 \rangle = \frac{2v_0^2}{D} \left[t - \frac{1}{D}(1 - e^{-Dt}) \right] \sim l^2. \tag{A18}$$

From Eq. (A18) we deduce (in both ballistic and diffusive limits) that the reshuffling rate scales as $\tau_{\text{resh}}^{-1} \sim D^{-1}$. In this disordered phase, the full velocity correlation function corresponds to the connected one, which exhibits a sharp, short-range decay: $C_c(l) \sim e^{-l/\xi(D)}$, where $\xi(D)$ is a finite correlation length, of the order of l, which scales as D^0 as $D \to \infty$. Therefore, the resulting scaling for the EPR is $\dot{\mathcal{S}} \sim D^{-2}$.

6. Single-particle decomposition

Let us focus on Model I, where phase coexistence is realized. We consider the statistics of single-particle contributions to the EPR from Eqs. (A3) and (A9), respectively, corresponding to the heat rate per particle, divided by D:

$$\dot{s}_i = \frac{\dot{q}_i}{D} = -\frac{J}{D} \sum_j \langle \dot{\theta}_i \circ n_{ij} \sin(\theta_i - \theta_j) \rangle, \tag{A19}$$

and work of reshuffling forces per particle, divided by D:

$$\dot{s}_i = \frac{\dot{w}_i}{D} = -\frac{J}{2D} \sum_j \langle \dot{n}_{ij} \circ \cos(\theta_i - \theta_j) \rangle. \tag{A20}$$

The goal of this analysis is to unveil possible correlations between dissipation and spatial segregation of each
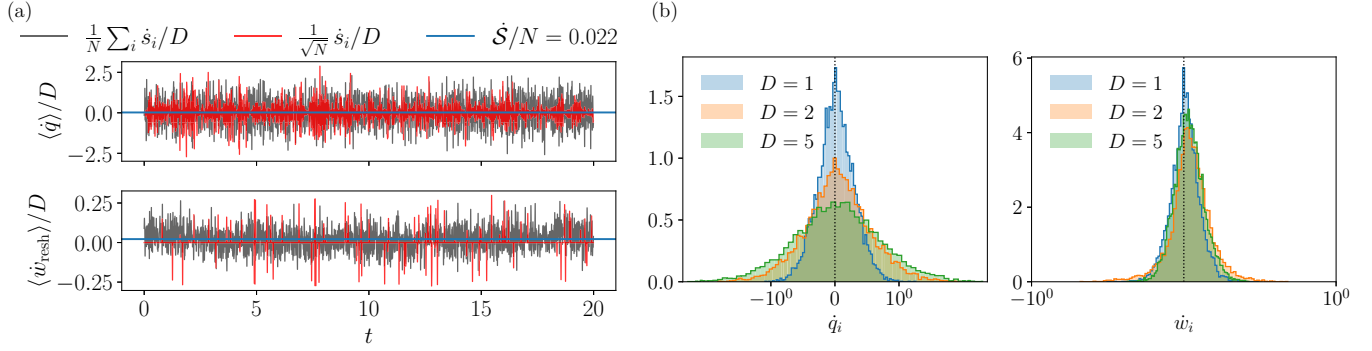
FIG. 6. (a) Time series of the average EPR of the flock (gray) and of the contribution of a sample single particle, properly rescaled (red), computed as in Eq. (A19) (top) or as in Eq. (A20) (bottom). The two time series have comparable fluctuations, especially when the stochastic heat is considered (top). This fact shows that all the particles contribute in the same way to the global EPR. (b) Probability density of the single-particle stochastic heat [left, from Eq. (A19)] or irreversible work of fictitious reshuffling forces [right, from Eq. (A20)]. The histogram is unimodal, even when particles are organized into polar clusters, in clear contrast with the MIPS phenomenology [59].

self-propelled particle, as it has been observed in scalar active matter.

In contrast to MIPS models [21,22,59], however, dissipation seems to occur in a nonlocalized way in the system. Two findings have been made in the context of scalar active matter: first of all, the analysis of both coarse-grained models and microscopic agent-based models for MIPS has shown that the breakdown of the time-reversal symmetry occurs at the interfaces, whereas in the bulk of both the dense and dilute phases the EPR density vanishes [21,22]. A second analysis on a different type of microscopic model (ABP with repulsive interactions, in contrast to active Ornstein-Uhlenbeck particles) has revealed that the work of the self-propulsion force per single particle assumes a different value depending on the particle's position inside the system [59]. The probability distribution of the active work per particle is assumed to obey a large deviation principle and the associated rate function is found to be nonconvex. The values corresponding to the minima of the rate function are the typical active work values of particles belonging to the bulk of the two phases; work values in the nonconvex region are typical of particles at the interface.

If we represent the EPR contribution per particle, using either Eq. (A19) or Eq. (A20), we see that violations of the time-reversal symmetry are not concentrated at the interface; on the contrary, the bulk contributes significantly and not with a definite sign [68]. If we also consider the time series of single-particle EPR contributions, we see that they largely fluctuate, taking both positive and negative values. The typical amplitude of fluctuations is comparable to that of the rescaled average EPR of the system, $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} s_i$ [Fig. 6(a)]. Finally, the histograms of the time-averaged single particle contributions, shown in Fig. 6(b), seem to have a unimodal shape, corroborating the idea that in polar flocks there is no correspondence between the particle's energetic role and its position in the flock.

APPENDIX B: TIME REVERSAL SYMMETRY BREAKING IN LANGEVIN SYSTEMS

In this Appendix we analyze general consequences deriving from the irreversibility condition on ABP models with alignment. We start by recalling some results presented in [61] and move from them to the following:

(1) Derive a new formula for the entropy production of nonequilibrium stochastic process driven by additive noise [Eq. (B14)]

(2) Show that the irreversibility condition induces on any ABP-based model of flocking an explicit asymmetry in the steady-state distribution of the system's microstates

(3) Generalize this last result to any Langevin system with completely irreversible drift.

1. General results

Let us consider a general stochastic additive process $\mathbf{z}(t)$ described by a Langevin equation with drift term $\mathbf{A}(\mathbf{x})$ and diffusion term $\mathbf{D}$. Let $\psi(\mathbf{z})$ be the steady-state distribution that solves the stationary Fokker-Planck equation: $\mathcal{L}_{FP}\psi(\mathbf{z}) = -\mathbf{\nabla} \cdot \mathbf{j}(\mathbf{z}) = 0$, where

$$\mathbf{j}(\mathbf{z}) = \mathbf{A}(\mathbf{z})\psi(\mathbf{z}) - \mathbf{D}\mathbf{\nabla}\psi(\mathbf{z}) \quad (B1)$$

is the probability current. This current is standardly decomposed into a *reversible* and *irreversible* part [44] as

$$\mathbf{j}^{\text{rev}}(\mathbf{z}) = \mathbf{A}^{\text{rev}}(\mathbf{z})\psi(\mathbf{z}), \quad (B2)$$

$$\mathbf{j}^{\text{irr}}(\mathbf{z}) = \mathbf{A}^{\text{irr}}(\mathbf{z})\psi(\mathbf{z}) - \mathbf{D}\mathbf{\nabla}\psi(\mathbf{z}), \quad (B3)$$

where

$$A_{\alpha}^{\text{rev}}(\mathbf{z}) = \tfrac{1}{2}[A_{\alpha}(\mathbf{z}) - \epsilon_{\alpha}A_{\alpha}(\boldsymbol{\epsilon}\mathbf{z})], \quad (B4)$$

$$A_{\alpha}^{\text{irr}}(\mathbf{z}) = \tfrac{1}{2}[A_{\alpha}(\mathbf{z}) + \epsilon_{\alpha}A_{\alpha}(\boldsymbol{\epsilon}\mathbf{z})], \quad (B5)$$

with $\boldsymbol{\epsilon}$ denoting the time-reversal operator. For now, we assume that $\boldsymbol{\epsilon}$ acts linearly on the state variable $\mathbf{z}$ (precisely, diagonally with $\epsilon_{\alpha} = \pm 1 \,\forall\alpha$).

A necessary and sufficient condition for detailed balance to hold is that the two following conditions are verified (in addition to stationarity, $\mathbf{\nabla} \cdot \mathbf{j} = 0$) [69]:

$$\mathbf{j}^{\text{irr}}(\mathbf{z}) = 0 \quad \text{and} \quad D_{\alpha\beta}(\mathbf{z}) = \epsilon_{\alpha}\epsilon_{\beta}D_{\alpha\beta}(\boldsymbol{\epsilon}\mathbf{z}). \quad (B6)$$

Therefore the condition of irreversibility implies $\mathbf{j}^{\text{irr}}(\mathbf{z}) \neq 0$, when $\mathbf{D}$ is independent of $\mathbf{z}$. This condition is sufficient to guarantee the positivity of the entropy production in a proper

NESS, i.e., in the absence of external drivings, where the total entropy production reduces to the housekeeping entropy production:

$$\dot{\mathcal{S}}^{hk} = \int d\mathbf{z}\, \psi(\mathbf{z})\, V_\alpha^{\mathrm{irr}}(\boldsymbol{\epsilon}\mathbf{z})[\mathbf{D}^{-1}(\boldsymbol{\epsilon}\mathbf{z})]_{\alpha\beta} V_\beta^{\mathrm{irr}}(\boldsymbol{\epsilon}\mathbf{z}). \quad \text{(B7)}$$

In (B7) $\mathbf{V}(\mathbf{z}) = \mathbf{j}(\mathbf{z})/\psi(\mathbf{z})$ is the phase space velocity [44], which is decomposed as $\mathbf{V}^{\mathrm{rev}}(\mathbf{z}) + \mathbf{V}^{\mathrm{irr}}(\mathbf{z})$ after (B2), and (B3). Dal Cengio *et al.* [61] identified Eq. (B7) with a second expression for the EPR given in [44],

$$\dot{\mathcal{S}} = \int d\mathbf{z}\, \psi(\mathbf{z})\, V_\alpha^{\mathrm{irr}}(\mathbf{z})[\mathbf{D}^{-1}(\mathbf{z})]_{\alpha\beta} V_\beta^{\mathrm{irr}}(\mathbf{z}), \quad \text{(B8)}$$

to derive the following constraint:

$$\sum_{\alpha\beta} D^{-1}{}_{\alpha\beta} V_\alpha^{\mathrm{irr}}(\mathbf{z}) V_\beta^{\mathrm{irr}}(\mathbf{z}) = \sum_{\alpha\beta} D^{-1}{}_{\alpha\beta} V_\alpha^{\mathrm{irr}}(\boldsymbol{\epsilon}\mathbf{z}) V_\beta^{\mathrm{irr}}(\boldsymbol{\epsilon}\mathbf{z}). \quad \text{(B9)}$$

Equation (B9) is valid almost surely if the diffusion matrix is invertible and $\mathbf{z}$-independent. Using the definition of $\mathbf{V}^{\mathrm{irr}}$ and the property $A_\alpha^{\mathrm{irr}}(\mathbf{z}) = \epsilon_\alpha A_\alpha^{\mathrm{irr}}(\boldsymbol{\epsilon}\mathbf{z})$, Eq. (B9) implies

$$\sum_\alpha \left[A_\alpha^{\mathrm{irr}}(\mathbf{z}) + D_{\alpha\beta} \partial_\beta \phi_+(\mathbf{z}) \right] \partial_\alpha \phi_-(\mathbf{z}) = 0 \quad \text{(B10)}$$

almost surely. The functions $\phi_\pm$ are defined as the T-symmetric and T-antisymmetric parts of the quasipotential $\phi(\mathbf{z}) = -\log \psi(\mathbf{z})$:

$$\phi_+(\mathbf{z}) = \tfrac{1}{2}[\phi(\mathbf{z}) + \phi(\boldsymbol{\epsilon}\mathbf{z})], \quad \phi_-(\mathbf{z}) = \tfrac{1}{2}[\phi(\mathbf{z}) - \phi(\boldsymbol{\epsilon}\mathbf{z})]. \quad \text{(B11)}$$

For the sake of simplicity, let us assume $\mathbf{D}$ is diagonal. All the results can be generalized to the nondiagonal case, if $\mathbf{D}$ is symmetric and positive definite. We remark that it is not possible to merely invoke the diagonalizability of $\mathbf{D}$ through a change of basis, because we need all the coordinates of $\mathbf{z}$ to have a definite parity under time reversal.

We start from an explicit rewriting of the entropy production rate from Eq. (B8):

$$\dot{\mathcal{S}} = \int d\mathbf{z}\, e^{-\phi(\mathbf{z})} \sum_\alpha D_{\alpha\alpha}^{-1} \left[A_\alpha^{\mathrm{irr}}(\mathbf{z}) + D_\alpha \partial_\alpha \phi(\mathbf{z}) \right]^2. \quad \text{(B12)}$$

Exploiting the symmetries of $\mathbf{A}^{\mathrm{irr}}$, ϕ_+ and ϕ_-, we can rewrite Eq. (B12) as

$$\begin{aligned}
\dot{\mathcal{S}} = &\int d\mathbf{z}\, 2\cosh[\phi_-(\mathbf{z})] e^{-\phi_+(\mathbf{z})} \\
&\times \sum_\alpha D_{\alpha\alpha}^{-1} \left\{ \left[A_\alpha^{\mathrm{irr}}(\mathbf{z}) + D_{\alpha\alpha} \partial_\alpha \phi_+(\mathbf{z}) \right]^2 + D_{\alpha\alpha}[\partial_\alpha \phi_-(\mathbf{z})]^2 \right\} \\
&+ \int d\mathbf{z}\, 4\sinh[\phi_-(\mathbf{z})] e^{-\phi_+(\mathbf{z})} \\
&\times \sum_\alpha \left[A_\alpha^{\mathrm{irr}}(\mathbf{z}) + D_{\alpha\alpha} \partial_\alpha \phi_+(\mathbf{z}) \right] \partial_\alpha \phi_-(\mathbf{z}).
\end{aligned} \quad \text{(B13)}$$

The second line of Eq. (B13) is zero due to Eq. (B10), hence

$$\begin{aligned}
\dot{\mathcal{S}} = &\int d\mathbf{z}\, \cosh[\phi_-(\mathbf{z})] e^{-\phi_+(\mathbf{z})} \\
&\times \sum_\alpha \left\{ D_{\alpha\alpha}^{-1} \left[\left[A_\alpha^{\mathrm{irr}}(\mathbf{z}) + D_{\alpha\alpha} \partial_\alpha \phi_+(\mathbf{z}) \right]^2 + [\partial_\alpha \phi_-(\mathbf{z})]^2 \right] \right\} \geqslant 0.
\end{aligned} \quad \text{(B14)}$$

Equality is realized in Eq. (B14) iff both of the following conditions are verified:

$$A_\alpha^{\mathrm{irr}}(\mathbf{z}) + D_\alpha \partial_{\alpha\alpha} \phi_+(\mathbf{z}) = 0 \quad \forall \alpha \quad \text{(B15)}$$

and

$$\partial_\alpha \phi_-(\mathbf{z}) = 0 \quad \forall \alpha \iff \phi_-(\mathbf{z}) = 0. \quad \text{(B16)}$$

Notice that ϕ_- cannot be a constant function different from zero.

Breakdown of detailed balance imposes that at least one of the terms in Eqs. (B15)–(B16) is nonzero, along with the constraint (B10). If all the coordinates of the state variable are T-even, then $\phi_-(\mathbf{z}) = 0$ and violation of Eq. (B15) is enforced. In the presence of T-odd coordinates, $\phi_-(\mathbf{z})$ can be different from zero (see Appendix B 3).

2. Application to Langevin-Vicsek model

The results presented above are valid for general Langevin stochastic processes, provided that they are additive with an invertible diffusion matrix. Let us now focus on our system of ABPs in Eqs. (1)–(2), which models a Langevin-Vicsek flock.

Since we work in the absence of translational diffusion, the $\mathbf{D}$ matrix is noninvertible, irrespectively of whether we decide to study the process in the $(\mathbf{X}, \boldsymbol{\Theta})$ phase space or in the $(\mathbf{X}, \mathbf{V})$ phase space, where $\mathbf{V} = e^{i\boldsymbol{\Theta}}$. Moreover, the process is nonadditive in the second case. In the first case the $\boldsymbol{\Theta}$ variables are not T-even or T-odd under time reversal, but the time-reversal operator acts by shifting them of an amount π. Hence the hypotheses that led to the derivation of the above results are not valid for the polar system of interest. Nonetheless, we can still derive from the irreversibility condition a condition on the asymmetry of the steady-state distribution $\psi(\mathbf{z})$.

Let us apply the same definition of quasipotential $\phi(\mathbf{z})$ and of its decomposition into T-symmetric and T-antisymmetric parts given in (B11). Our goal is to show (by contradiction) that irreversibility implies $\phi_-(\mathbf{z}) \neq 0$. Let us start from the stationary FPE,

$$v_0 \sum_i \mathbf{e}(\theta_i) \cdot \nabla_i \psi = J \sum_{ij} n_{ij} \partial_{\theta_i}[\sin(\theta_i - \theta_j)\psi] + D \sum_i \partial_{\theta_i \theta_i}^2 \psi, \quad \text{(B17)}$$

and derive the corresponding PDE for the quasipotential ϕ:

$$\begin{aligned}
v_0 \sum_i \mathbf{e}(\theta_i) \cdot \nabla_i \phi = &J \sum_{ij} n_{ij} \cos(\theta_i - \theta_j) \\
&+ J \sum_{ij} n_{ij} \sin(\theta_i - \theta_j) \partial_{\theta_i} \phi \\
&+ D \sum_i \partial_{\theta_i \theta_i}^2 \phi + D \sum_i (\partial_{\theta_i} \phi)^2. \quad \text{(B18)}
\end{aligned}$$

Let us apply the time-reversal operator to both the r.h.s. and l.h.s. of Eq. (B18) (this operation corresponds to a change of variable from the state variable $\mathbf{z}$ to $\boldsymbol{\epsilon}\mathbf{z}$) and split the

quasipotential into its T-symmetric and T-antisymmetric parts to rewrite a set of two coupled stationary equations:

$$v_0 \sum_i \mathbf{e}(\theta_i) \cdot \nabla_i \phi_- = J \sum_{ij} n_{ij} \cos(\theta_i - \theta_j)$$
$$+ J \sum_{ij} n_{ij} \sin(\theta_i - \theta_j) \partial_{\theta_i} \phi_+$$
$$+ D \sum_i \partial^2_{\theta_i \theta_i} \phi_+$$
$$+ D \sum_i (\partial_{\theta_i} \phi_+)^2 + D \sum_i (\partial_{\theta_i} \phi_-)^2,$$

$$\text{(B19)}$$

$$v_0 \sum_i \mathbf{e}(\theta_i) \cdot \nabla_i \phi_+ = J \sum_{ij} n_{ij} \sin(\theta_i - \theta_j) \partial_{\theta_i} \phi_-$$
$$+ D \sum_i \partial^2_{\theta_i \theta_i} \phi_- + 2D \sum_i (\partial_{\theta_i} \phi_+)(\partial_{\theta_i} \phi_-).$$

$$\text{(B20)}$$

We now assume that $\phi_-(\mathbf{z}) = 0$ in Eq. (B19). The resulting equation for $\psi = e^{\phi_+}$ is of the form

$$0 = J \sum_{ij} n_{ij} \partial_{\theta_i} [\sin(\theta_i - \theta_j)\psi] + D \sum_i \partial^2_{\theta_i \theta_i} \psi. \quad \text{(B21)}$$

This equation is solved by the Boltzmann distribution that we would have in the absence of self-propulsion:

$$\psi(\mathbf{z}) = f(\mathbf{X}) \exp[-\mathcal{H}(\mathbf{z})] \iff \phi_+(\mathbf{z}) = \frac{\mathcal{H}(\mathbf{z})}{D} + c(\mathbf{X}),$$

$$\text{(B22)}$$

with $c(\mathbf{X})$ an arbitrary constant. Uniqueness of the stationary solution of a Fokker-Planck equation is guaranteed under rather general smoothness hypotheses on the solution and drift term [70]. Let us notice that we crucially exploited the symmetry of n_{ij} to write the steady-state solution as Eq. (B22). We can conclude that the nonequilibrium condition is contradicted: the hypothesis $\phi_-(\mathbf{z}) = 0$ leads to an absurdum.

In Langevin-Vicsek models, the condition $\phi_-(\mathbf{z}) \neq 0$ corresponds to requiring that

$$\psi(\mathbf{X}, \mathbf{V}) \neq \psi(\mathbf{X}, e^{i\pi}\mathbf{V}); \quad \text{(B23)}$$

i.e., irreversibility constrains an explicit symmetry breaking under rotations in the internal space of velocities. A possible way to measure of the degree of irreversibility of the process is to quantify the asymmetry of the steady-state pdf in Eq. (B23). However, reconstructing the N-body probability density is clearly an out-of-reach task, both in numerical simulations or real experiments. Having some knowledge of the aligning interaction potential $\mathcal{H}_{XY}(\mathbf{\Theta}; \mathbf{n}(X))$ allows us to visualize the asymmetry on a much lower-dimensional space, and to predict how the second law of thermodynamics ($\dot{\mathcal{S}} \geqslant 0$) constrains the realization of such asymmetries.

3. Langevin processes with irreversible drift

The result obtained in the previous section can be extended to any Langevin process where $\mathbf{A}^{\text{rev}}(\mathbf{z}) = 0$ and the state variable coordinates have a definite parity under time reversal. It is sufficient to write the stationary Fokker-Planck equation

$$\nabla \cdot \mathbf{A} + \mathbf{A} \cdot \nabla \phi + D\nabla^2 \phi + D(\nabla \phi)^2 = 0 \quad \text{(B24)}$$

and perform a change of variable, identifying $\mathbf{z}$ as $\boldsymbol{\epsilon} \mathbf{z}'$. Using the decomposition of $\mathbf{A}$ and ϕ into their T-symmetric and T-antisymmetric components, and combining (respectively with a positive and negative sign) the resulting equation with (B24), one can obtain the two following equations:

$$\nabla_O \cdot \mathbf{A}^{\text{irr}}_O + \mathbf{A}^{\text{irr}}_E \cdot \nabla_E \phi_- + \mathbf{A}^{\text{irr}}_O \cdot \nabla_O \phi_+$$
$$+ D\nabla^2 \phi_+ + D(\nabla \phi_+)^2 + D(\nabla \phi_-)^2 = 0, \quad \text{(B25)}$$

$$\nabla_E \cdot \mathbf{A}^{\text{irr}}_E + \mathbf{A}^{\text{irr}}_E \cdot \nabla_E \phi_+ + \mathbf{A}^{\text{irr}}_O \cdot \nabla_O \phi_-$$
$$+ D\nabla^2 \phi_- + 2D\nabla \phi_+ \cdot \nabla \phi_- = 0. \quad \text{(B26)}$$

The subscripts in the equations above indicate the parity of the state variable coordinates, which can be split as $\mathbf{z} = (\mathbf{z}_E, \mathbf{z}_O)$, such that $\boldsymbol{\epsilon}\mathbf{z} = (\mathbf{z}_E, -\mathbf{z}_O)$. Assuming that $\phi_-(\mathbf{z}) = 0$ (absurdum), the sum of (B25) and (B26) yields

$$\nabla \cdot (\mathbf{A}^{\text{irr}}\psi) + D\nabla^2 \psi = 0, \quad \text{(B27)}$$

where we have identified $\psi = e^{\phi_+}$. Providing natural boundary conditions to the FPE (B24), we deduce from (B27) that $\mathbf{A}^{\text{irr}}\psi + D\nabla\psi = 0$. This condition, together with $\phi_-(\mathbf{z}) = 0$, implies detailed balance. Therefore irreversibility implies, in a system where $\mathbf{A}^{\text{rev}} = 0$ and the state variables coordinates are either T-odd or T-even, that $\phi_-(\mathbf{z}) \neq 0$.

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

[2] H. Chaté, Annu. Rev. Condens. Matter Phys. **11**, 189 (2020).

[3] J. Toner, Y. Tu, and S. Ramaswamy, Ann. Phys. **318**, 170 (2005).

[4] T. F. F. Farage, P. Krinninger, and J. M. Brader, Phys. Rev. E **91**, 042310 (2015).

[5] M. Fruchart, R. Hanai, P. B. Littlewood, and V. Vitelli, Nature (London) **592**, 363 (2021).

[6] L. P. Dadhichi, J. Kethapelli, R. Chajwa, S. Ramaswamy, and A. Maitra, Phys. Rev. E **101**, 052601 (2020).

[7] M. J. Bowick, N. Fakhri, M. C. Marchetti, and S. Ramaswamy, Phys. Rev. X **12**, 010501 (2022).

[8] S. A. M. Loos and S. H. L. Klapp, Sci. Rep. **9**, 2491 (2019).

[9] H. Tasaki, Phys. Rev. Lett. **125**, 220601 (2020).

[10] E. Fodor, C. Nardini, M. E. Cates, J. Tailleur, P. Visco, and F. van Wijland, Phys. Rev. Lett. **117**, 038103 (2016).

[11] S. C. Takatori and J. F. Brady, Phys. Rev. E **91**, 032117 (2015).

[12] L. Dabelow, S. Bo, and R. Eichhorn, Phys. Rev. X **9**, 021009 (2019).

[13] P. Pietzonka and U. Seifert, J. Phys. A: Math. Theor. **51**, 01LT01 (2018).

[14] C. Battle, C. P. Broedersz, N. Fakhri, V. F. Geyer, J. Howard, C. F. Schmidt, and F. C. MacKintosh, Science **352**, 604 (2016).

[15] Q. Yu and Y. Tu, arXiv:2205.13149 (2022).

[16] D. Loi, S. Mossa, and L. F. Cugliandolo, Phys. Rev. E **77**, 051111 (2008).

[17] G. Szamel, Phys. Rev. E **90**, 012111 (2014).

[18] D. Levis and L. Berthier, Europhys. Lett. **111**, 60006 (2015).

[19] C. Maggi, N. Gnan, M. Paoluzzi, E. Zaccarelli, and A. Crisanti, Commun. Phys. **5**, 55 (2022).

[20] S. Dal Cengio, D. Levis, and I. Pagonabarraga, Phys. Rev. Lett. **123**, 238003 (2019).

[21] C. Nardini, E. Fodor, E. Tjhung, F. van Wijland, J. Tailleur, and M. E. Cates, Phys. Rev. X **7**, 021007 (2017).

[22] D. Martin, J. O'Byrne, M. E. Cates, E. Fodor, C. Nardini, J. Tailleur, and F. van Wijland, Phys. Rev. E **103**, 032607 (2021).

[23] E. Crosato, M. Prokopenko, and R. E. Spinney, Phys. Rev. E **100**, 042613 (2019).

[24] J. O'Byrne, Y. Kafri, J. Tailleur, and F. van Wijland, Nat. Rev. **4**, 167 (2022).

[25] Ø. L. Borthne, É. Fodor, and M. E. Cates, New J. Phys. **22**, 123012 (2020).

[26] N. Kyriakopoulos, F. Ginelli, and J. Toner, New J. Phys. **18**, 073039 (2016).

[27] L. P. Dadhichi, A. Maitra, and S. Ramaswamy, J. Stat. Mech.: Theory Exp. (2018) 123201.

[28] B. Partridge and C. F. Lee, Phys. Rev. Lett. **123**, 068002 (2019).

[29] C. Maggi, M. Paoluzzi, A. Crisanti, E. Zaccarelli, and N. Gnan, Soft Matter **17**, 3807 (2021).

[30] F. Caballero, C. Nardini, and M. E. Cates, J. Stat. Mech.: Theory Exp. (2018) 123208.

[31] L. Chen, J. Toner, and C. F. Lee, New J. Phys. **17**, 042002 (2015).

[32] J. Toner and Y. Tu, Phys. Rev. Lett. **75**, 4326 (1995).

[33] Y. Tu, J. Toner, and M. Ulm, Phys. Rev. Lett. **80**, 4819 (1998).

[34] A. Cavagna, L. D. Carlo, I. Giardina, T. S. Grigera, S. Melillo, L. Parisi, G. Pisegna, and M. Scandolo, arXiv:2107.04432 (2021).

[35] A. P. Solon, Y. Fily, A. Baskaran, M. E. Cates, Y. Kafri, M. Kardar, and J. Tailleur, Nat. Phys. **11**, 673 (2015).

[36] Y. Fily, Y. Kafri, A. P. Solon, J. Tailleur, and A. Turner, J. Phys. A: Math. Theor. **51**, 044003 (2017).

[37] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, Phys. Rev. Lett. **75**, 1226 (1995).

[38] C. Maes, J. Stat. Phys. **95**, 367 (1999).

[39] C. Maes, F. Redig, and A. V. Moffaert, J. Math. Phys. **41**, 1528 (2000).

[40] O. Chepizhko, D. Saintillan, and F. Peruani, Soft Matter **17**, 3113 (2021).

[41] P. Rahmani, F. Peruani, and P. Romanczuk, Commun. Phys. **4**, 206 (2021).

[42] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini *et al.*, Proc. Natl. Acad. Sci. USA **105**, 1232 (2008).

[43] M. Camperi, A. Cavagna, I. Giardina, G. Parisi, and E. Silvestri, Interface Focus **2**, 715 (2012).

[44] R. E. Spinney and I. J. Ford, Phys. Rev. E **85**, 051113 (2012).

[45] U. Seifert, Eur. Phys. J. B **64**, 423 (2008).

[46] U. Seifert, Rep. Prog. Phys. **75**, 126001 (2012).

[47] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998).

[48] L. Onsager and S. Machlup, Phys. Rev. **91**, 1505 (1953).

[49] K. Sekimoto, J. Phys. Soc. Jpn. **66**, 1234 (1997).

[50] Computational Geometry Algorithms Library, https://www.cgal.org.

[51] G. Grégoire and H. Chaté, Phys. Rev. Lett. **92**, 025702 (2004).

[52] F. Ginelli and H. Chaté, Phys. Rev. Lett. **105**, 168103 (2010).

[53] A. Peshkov, S. Ngo, E. Bertin, H. Chaté, and F. Ginelli, Phys. Rev. Lett. **109**, 098101 (2012).

[54] D. Martin, H. Chaté, C. Nardini, A. Solon, J. Tailleur, and F. Van Wijland, Phys. Rev. Lett. **126**, 148001 (2021).

[55] S. Shankar and M. C. Marchetti, Phys. Rev. E **98**, 020604(R) (2018).

[56] T. Mora, A. M. Walczak, L. Del Castello, F. Ginelli, S. Melillo, L. Parisi, M. Viale, A. Cavagna, and I. Giardina, Nat. Phys. **12**, 1153 (2016).

[57] G. Rassolov, L. Tociu, É. Fodor, and S. Vaikuntanathan, J. Chem. Phys. **157**, 054901 (2022).

[58] S. Ro, B. Guo, A. Shih, T. V. Phan, R. H. Austin, D. Levine, P. M. Chaikin, and S. Martiniani, arXiv:2105.12707 (2021).

[59] P. Chiarantoni, F. Cagnetta, F. Corberi, G. Gonnella, and A. Suma, J. Phys. A: Math. Theor. **53**, 36LT02 (2020).

[60] R. Kürsten and T. Ihle, Phys. Rev. E **104**, 034604 (2021).

[61] S. D. Cengio, D. Levis, and I. Pagonabarraga, J. Stat. Mech.: Theory Exp. (2021) 043201.

[62] R. Suzuki, C. A. Weber, E. Frey, and A. R. Bausch, Nat. Phys. **11**, 839 (2015).

[63] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, S. Melillo, L. Parisi, O. Pohl, B. Rossaro, E. Shen, E. Silvestri *et al.*, Phys. Rev. Lett. **113**, 238102 (2014).

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

[65] A. Cavagna, S. M. D. Queirós, I. Giardina, F. Stefanini, and M. Viale, Proc. R. Soc. B **280**, 20122484 (2013).

[66] *Stochastic Geometry and Its Applications*, edited by S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, Wiley Series in Probability and Statistics (John Wiley & Sons, New York, 2013), Chap. 9, pp. 343–410, doi: 10.1002/9781118658222.ch09.

[67] P. Romanczuk, M. Bär, W. Ebeling, B. Lindner, and L. Schimansky-Geier, Eur. Phys. J.: Spec. Top. **202**, 1 (2012).

[68] See Supplemental Material at https://link.aps.org/supplemental/10.1103/PhysRevE.106.034608 for videos showing single particle contributions to the total entropy production in different regions of the phase diagram. Simulation parameters are specified in the file names S#_movie_work_T(*D value*)_N(*particle number*) show the single particle contribution to the EPR as computed from Eq. (8); S#A_movie_T(*D value*)_N(*particle number*) show the single particle contribution to the EPR as computed from Eq. (7). Color code red and blue correspond to positive and negative entropy production respectively.

[69] C. Gardiner, *Stochastic Methods*, Springer Series in Synergetics Vol. 13 (Springer-Verlag, Berlin, 2009).

[70] E. C. Zeeman, Nonlinearity **1**, 115 (1988).