Dynamical maximum entropy approach to flocking

Andrea Cavagna,[1,2,3] Irene Giardina,[1,2,3] Francesco Ginelli,[4] Thierry Mora,[5] Duccio Piovani,[2]
Raffaele Tavarone,[2] and Aleksandra M. Walczak[6]

[1]*Istituto Sistemi Complessi, Consiglio Nazionale delle Ricerche, UOS Sapienza, Rome, Italy*
[2]*Dipartimento di Fisica, Università Sapienza, Rome, Italy*
[3]*Initiative for the Theoretical Sciences, The Graduate Center, The City University of New York, New York, USA*
[4]*SUPA, Institute for Complex Systems and Mathematical Biology, King's College, University of Aberdeen, Aberdeen, UK*
[5]*Laboratoire de physique statistique, CNRS, UPMC and École normale supérieure, Paris, France*
[6]*Laboratoire de physique théorique, CNRS, UPMC and École normale supérieure, Paris, France*

We derive a new method to infer from data the out-of-equilibrium alignment dynamics of collectively moving animal groups, by considering the maximum entropy model distribution consistent with temporal and spatial correlations of flight direction. When bird neighborhoods evolve rapidly, this dynamical inference correctly learns the parameters of the model, while a static one relying only on the spatial correlations fails. When neighbors change slowly and the detailed balance is satisfied, we recover the static procedure. We demonstrate the validity of the method on simulated data. The approach is applicable to other systems of active matter.

I. INTRODUCTION

Flocking, the highly coordinated motion displayed by large groups of birds, has attracted much attention over the last 20 years as a prototypical example of out-of-equilibrium collective behavior. It has been suggested that flocking is an emergent phenomenon resulting from the mutual alignment of velocities between neighboring birds, much like the spontaneous symmetry breaking towards a magnetized state exhibited by ferromagnetic spins at low temperatures. This idea has been extensively studied from a theoretical view point [1–14], and recent advances in the imaging of large flocks of starlings, allowing for the three-dimensional (3D) reconstruction of large groups [15–17], have given empirical support to this picture. Interactions between individuals in the flock were shown to be topological and local [18], leading to the global ordering of flight orientations and scale-free correlation functions [19]. The analogy with ferromagnetic systems was made explicit by the quantitative inference of spin models from empirical data using the principle of maximum entropy [20,21]. These analyses have focused on the steady state behavior of flocks, by examining the flock configurations as drawn from a given statistical ensemble. This approach allows for an effective equilibrium-like description, without having to make detailed assumptions about the microscopic rules governing flock behavior. Yet it is an incomplete picture as it does not take into account the dynamical, out-of-equilibrum nature of the process.

The major difference between flocks and equilibrium spin systems is that birds are like active particles, constantly moving within the flock along the direction given by their "spin," exchanging local interaction partners, thus extending their effective interaction range, and also breaking detailed balance. This qualitative difference between equilibrium spins and out-of-equilibrium active particles can dramatically affect the thermodynamic properties of the system, including the existence of an ordered phase in two dimensions, and the value of the critical exponents [12,13]. One can thus interpret the parameters of static descriptions of flocks as a renormalized version of some underlying and unknown out-of-equilibrium dynamical model. This calls for a new theoretical approach to infer the correct dynamics of strongly out-of-equilibrium systems.

In this paper we propose a general framework for learning the features of the out-of-equilibrium dynamics of collective motion directly from data, while making minimal assumptions about the specific microscopic interaction rules. We choose to focus on the example of orientational order in bird flocks because of its central importance for collective animal behavior and its simple symmetries.

We generalize the principle of maximum entropy to account for correlations between birds at different times, in addition to same-time correlation functions. One of our main theoretical results is that maximizing the entropy under these constraints leads to an explicit stochastic dynamical equation for the flock. When bird orientations are strongly polarized, the procedure is equivalent to inferring a dynamical model of social forces with noise.

Explicit analytical expressions of the interaction parameters, as well as the likelihood of the data under the model, are derived as a function of global observables of the data. When the interaction network changes adiabatically, and interactions are symmetric, detailed balance is satisfied and our method recovers the results of the static approach [20] as expected. Even in this equilibrium case, the dynamical approach allows us to separate the contributions of interaction strength and noise to the alignment dynamics, while the static approach cannot.

To illustrate the usefulness and performance of our method on a well understood example, we apply our formulas to synthetic data generated by a topological Vicsek model (VM). We

show that its inferred interaction parameters are consistently better than the ones obtained in a static equilibrium framework, especially when the relative mobility between individuals is high, as predicted by the theory.

II. DYNAMICAL MAXIMUM ENTROPY

A. General maximum entropy approach

Maximum entropy distributions are the least constrained distributions that are consistent with certain selected key observables of the data. They usually map onto equilibrium statistical mechanics problems and do not involve any assumptions about the system under study, besides the choice of the relevant observables, which should be selected according to the fundamental symmetries of the underlying system. They have been particularly successful in describing collective and emergent phenomena in biological systems comprising many correlated degrees of freedom [22–27].

The principle of maximum entropy consists of finding the maximally disordered probability distribution consistent with carefully chosen observables of the data. In practice, given a stochastic variable **s**, and a set of observables $\{\mathcal{O}_\mu(\mathbf{s})\}$, with $\mu = 1, \ldots, K$, one looks for the model distribution P of maximum entropy

$$S[P] = -\sum_{\mathbf{s}} P(\mathbf{s}) \ln P(\mathbf{s}) \qquad (1)$$

that coincides with the data for the average values of each of the observables

$$\langle \mathcal{O}_\mu \rangle_{\text{data}} = \langle \mathcal{O}_\mu \rangle_P. \qquad (2)$$

Using the technique of Lagrange multipliers, one shows that the distribution takes the exponential form

$$P(\mathbf{s}) = \frac{1}{\mathcal{Z}(\{\lambda_\mu\})} \exp\left(-\sum_{\mu=1}^{K} \lambda_\mu \mathcal{O}_\mu(\mathbf{s})\right), \qquad (3)$$

where $\{\lambda_\mu\}$ are Lagrange multipliers that need to be set to satisfy Eq. (2), and $\mathcal{Z}(\{\lambda_\mu\})$ is a normalization factor enforcing $\sum_{\mathbf{s}} P(\mathbf{s}) = 1$. By analogy with the Boltzman distribution from equilibrium statistical mechanics, the sum inside the exponential may be interpreted as an energy.

Conveniently, the Lagrange multipliers that match the mean value of the observables are also those that maximize the likelihood of the data given the exponential form (3). Given M data points $\mathbf{s}^1, \ldots, \mathbf{s}^M$, the log-likelihood of the data reads

$$\ln \mathcal{P}(\{\lambda_\mu\}) \equiv \ln \prod_{a=1}^{M} P(\mathbf{s}^a)$$
$$= -\sum_{a=1}^{M} \sum_{\mu=1}^{K} \lambda_\mu \mathcal{O}_\mu(\mathbf{s}^a) - M \ln \mathcal{Z}(\{\lambda_\mu\}). \qquad (4)$$

Maximizing the log-likelihood with respect to the parameters $\{\lambda_\mu\}$ implies

$$\frac{\partial \ln \mathcal{P}(\{\lambda_\mu\})}{\partial \lambda_\mu} = M\left[-\frac{\partial \ln \mathcal{Z}}{\partial \lambda_\mu} - \langle \mathcal{O}(\mathbf{s})\rangle_{\text{data}}\right] = 0,$$
$$\qquad (5)$$
$$M[\langle \mathcal{O}(\mathbf{s})\rangle_P - \langle \mathcal{O}(\mathbf{s})\rangle_{\text{data}}] = 0.$$

By virtue of this equivalence, the correct maximum entropy distribution can be obtained by maximizing the expression of the log-likelihood with respect to the parameters.

Note that the variable **s** can be any stochastic variable. It may be a configuration of the system, or a full time trace of configurations, in which case the observables may span different times.

B. Static maximum entropy

In the following we will focus on the particular example of bird flocks, although many of our results on the dynamics should hold for other choices of systems and observables. For each particular situation, the choice of observables is dictated by the symmetries of the problem, as well as the relevant order parameters that one wishes to capture.

When polar order is present as in bird flocks, a natural choice of observables to be constrained by the data are the equal time pairwise correlation functions between the birds' orientations: $\langle s_i s_j \rangle$, where s_i is a d-dimensional unit vector denoting the flight direction of bird i, with $i = 1, \ldots, N$. (Throughout the paper the inner products over the physical space are implicit.) These correlations were found to exhibit scale-free behavior in natural flocks [19] and characterize the collective nature of flocking. The maximum entropy distribution $P(\mathbf{s})$ for the orientations can then be computed by maximizing the entropy $S[P] = -\sum_{\mathbf{s}} P(\mathbf{s}) \ln P(\mathbf{s})$, while constraining the equal-time correlations to their experimental values. The result is the stationary probability distribution for the equilibrium heterogeneous Heisenberg model [20]

$$P(\mathbf{s}) = \frac{1}{Z} \exp\left(\frac{1}{2} \sum_{i \neq j} J_{ij}^{\text{stat}} s_i s_j\right), \qquad (6)$$

where **s** is a shorthand for $(s_1, s_2, \ldots, s_N)$ and Z a normalization constant. The interaction parameters J_{ij}^{stat} are Lagrange multipliers that need to be tuned so that the probability distribution (6) matches the empirical correlation functions $\langle s_i s_j \rangle$. Using three-dimensional, single-individual-resolution data of large bird flocks, this class of models was shown to recapitulate quantitatively the ordering properties of real flocks [20].

C. Dynamical approach: Maximum entropy on trajectories

A limitation of the static approach is that infinitely many dynamical models may give rise to the same steady-state distribution, most of which break detailed balance. In fact, the change of neighborhoods causes the interaction network to vary in time, keeping the system constantly out of equilibrium. Here we extend the maximum entropy framework to account for the nonequilibrium nature of flocking, to what is often called the principle of maximum caliber [28]. We consider the set of entire trajectories $(\mathbf{s}^1, \mathbf{s}^2, \ldots, \mathbf{s}^T)$, where the superscript index denotes time points separated by δt. We then look for the distribution $P(\mathbf{s}^1, \ldots, \mathbf{s}^T)$ that maximizes the entropy while reproducing some given experimental observables. Since we want to capture the dynamics, in addition to equal-time correlation functions, we also constrain the correlation functions between two consecutive time points $\langle s_i^{t+1} s_j^t \rangle$. Doing

so yields the following form of the probability distribution over *trajectories*:

$$P(\mathbf{s}^1, \ldots, \mathbf{s}^T) = \frac{1}{\hat{Z}} \exp(-\mathcal{A}), \tag{7}$$

where $\hat{Z}$ is a normalization factor, and the "effective action" (or minus log-likelihood) reads

$$\mathcal{A} = -\frac{1}{2} \sum_t \sum_{i \neq j} \left(J_{ij;t}^{(1)} s_i^t s_j^t + J_{ij;t}^{(2)} s_i^{t+1} s_j^t \right). \tag{8}$$

There now are two sets of time-dependent coupling parameters, for synchronous and consecutive times.

III. EQUIVALENCE WITH MODELS OF SOCIAL FORCES

A. Flock dynamics as a Markov process

Because the action only involves cross-terms between consecutive times, the likelihood of a flock trajectory be written as a sum of terms corresponding to the transition probabilities $P(\pi'|\pi)$ betweenx successive time points

$$P(\mathbf{s}^1, \ldots, \mathbf{s}^T) = P(\mathbf{s}^1) \prod_{t=1}^{T-1} P(\mathbf{s}^t | \mathbf{s}^{t-1}), \tag{9}$$

which can be rewritten in terms of the action as

$$\mathcal{A} + \ln \hat{Z} = -\ln P(\mathbf{s}^1) + \sum_t \mathcal{L}_t(\mathbf{s}^{t+1}, \mathbf{s}^t), \tag{10}$$

where

$$\mathcal{L}_t(\mathbf{s}^{t+1}, \mathbf{s}^t) \equiv -\ln P(\mathbf{s}^{t+1} | \mathbf{s}^t) \tag{11}$$

may be interpreted as a Lagrangian density in the path integral formalism.

Let us check that this Markovian decomposition is possible. Identifying the two expressions of $\mathcal{A}$ (8) and (10), we may write $\mathcal{L}_t$ in the form

$$\mathcal{L}_t(\mathbf{s}', \mathbf{s}) = -\frac{1}{2} \sum_{ij} \left(J_{ij;t}^{(2)} s_i' s_j + J_{ij;t}^{(1)} s_i s_j \right) - K_t(\mathbf{s}') + K_{t-1}(\mathbf{s}), \tag{12}$$

with the constraint that, for all $\mathbf{s}$, the transition probability be normalized

$$1 = \sum_{\mathbf{s}'} \exp[-\mathcal{L}_t(\mathbf{s}', \mathbf{s})], \tag{13}$$

which entails

$$K_{t-1}(\mathbf{s}) = \ln \sum_{\mathbf{s}'} \exp\left[\frac{1}{2} \sum_{ij} \left(J_{ij;t}^{(2)} s_i' s_j + J_{ij;t}^{(1)} s_i s_j \right) + K_t(\mathbf{s}') \right]. \tag{14}$$

Equation (14) defines a descending recursion, by which K_t is calculated from the next time point.

Note that the process is Markovian because we chose to constrain consecutive-time correlation functions. Non-Markov forms would arise if we constrained more complex multitime observables.

B. Spin-wave approximation

In general, the integral in Eq. (14) cannot be calculated analytically, and K_t does not have a simple quadratic form as a function of $\mathbf{s}$.

However, when flight orientations are highly polarized (as in the case of starling flocks [19]), one can use the spin-wave (SW) approximation [29] to explicitly rewrite the action as a sum of terms that are quadratic in the spin-wave variables. Specifically, we denote $s_i = \pi_i + n\sqrt{1 - (\pi_i)^2}$, where n is an arbitrary unit vector close to the average flight direction of the flock, and π_i is the perpendicular component of the orientation $\pi_i n = 0$. (When there is no ambiguity we drop the time superscript.) When the flock is highly polarized, we have $\pi_i^2 \ll 1$, and we may expand at small π_i.

Let us assume a quadratic form for K_t:

$$K_t(\mathbf{s}) = \frac{1}{2} \sum_{ij} K_{ij;t} s_i s_j + U_t. \tag{15}$$

The integral in Eq. (14) can be expanded at small π:

$$\begin{aligned} K_{t-1}(\boldsymbol{\pi}) = {} & \frac{1}{2} \sum_{ij} J_{ij;t}^{(1)} \left(1 + \pi_i \pi_j - \pi_i^2 \right) \\ & + U_t + \frac{1}{2} \sum_{ij} \left(J_{ij;t}^{(2)} + K_{ij;t} \right) - \frac{1}{4} \sum_{ij} J_{ij;t}^{(2)} \pi_j^2 \\ & + \ln \int d\boldsymbol{\pi}' \exp\left[-\frac{1}{2} \sum_{ij} A_{ij;t} \pi_i' \pi_j' \right. \\ & \left. + \frac{1}{2} \sum_{ij} J_{ij;t}^{(2)} \pi_i' \pi_j \right], \end{aligned} \tag{16}$$

with

$$A_{ij;t} = -K_{ij;t} + \delta_{ij} \sum_k K_{ik;t} + \frac{1}{2} \delta_{ij} \sum_k J_{ik;t}^{(2)}. \tag{17}$$

This Gaussian integral can be calculated exactly. Doing so, and expanding the left-hand side of Eq. (14) at small π, yields a simple matrix expression for the recursion (only valid for nondiagonal terms)

$$\mathbf{K}_{t-1} = \mathbf{J}_t^{(1)} + \frac{1}{4} \mathbf{J}_t^{(2)\dagger} \mathbf{A}_t^{-1} \mathbf{J}_t^{(2)}. \tag{18}$$

We can now replace the expression of K_t in Eq. (12), and thus rewrite the transition probability in terms of π in a Gaussian form

$$\begin{aligned} \mathcal{L}_t(\boldsymbol{\pi}', \boldsymbol{\pi}) = {} & -\frac{d-1}{2} \ln\left(\frac{\det \mathbf{A}_t}{(2\pi)^N} \right) \\ & + \frac{1}{2} (\boldsymbol{\pi}' - \mathbf{M}_t \boldsymbol{\pi})^\dagger \mathbf{A}_t (\boldsymbol{\pi}' - \mathbf{M}_t \boldsymbol{\pi}), \end{aligned} \tag{19}$$

with

$$\mathbf{M}_t = \frac{1}{2} \mathbf{A}_t^{-1} \mathbf{J}_t^{(2)}. \tag{20}$$

C. Equivalence with generalized Vicsek models

The Gaussian form of the transition probabilities Eq. (19), corresponds to a spin-wave dynamics described by the follow-

ing stochastic equation:

$$\pi_i^{t+1} = \sum_j M_{ij;t} \pi_j^t + \epsilon_i^t, \tag{21}$$

with ϵ^t being a random, isotropic Gaussian noise perpendicular to n, of zero mean and covariance

$$\langle \epsilon^t (\epsilon^{t'})^\dagger \rangle = (d-1) \mathbf{A}_t^{-1} \delta_{t,t'}, \tag{22}$$

where $\delta_{t,t'}$ is the Kronecker delta. [Note that the $(d-1)$ factor, here and in previous equations, corresponds to the dimensionality of the perpendicular component π.]

Equation (21) can be interpreted as follows. At each time, individual i computes its new orientation from a weighted average over the orientation of other individuals, including itself, at the previous time point with weights encoded in the matrix $\mathbf{M}_t$. $\mathbf{M}_t$ defines a well-balanced weighted average, as it satisfies

$$\sum_j M_{ij;t} = 1. \tag{23}$$

[This identity can be checked by rewriting it in a matrix form: $\frac{1}{2} \mathbf{A}_t^{-1} \mathbf{J}_t^{(2)} \mathbf{u} = \mathbf{u}$ where $\mathbf{u}$ is a vector of ones, $u_i = 1$. Proving Eq. (23) is therefore equivalent to showing $\mathbf{J}_t^{(2)} \mathbf{u} = 2 \mathbf{A}_t \mathbf{u}$, which follows from the definition of $\mathbf{A}_t$ in Eq. (17).]

Noise ϵ^t added to this average in Eq. (21) determines the level of error in the alignment. Without it, all individuals would be perfectly aligned. This model may be viewed as the spin-wave expansion of a generalized Vicsek model [1] with arbitrary weights and noise.

Equations (21) and (22), which we have just derived for the perpendicular component π of the motion, hold almost the same for the flight direction s itself. Starting from the update equation

$$s_i^{t+1} = \theta \left[\sum_j M_{ij;t} s_j^t + \eta_i^t \right], \tag{24}$$

where $\theta(x) = x / \|x\|$ is the normalization operator, and expanding in the spin-wave approximation ($\pi_i \ll 1$), one recovers Eq. (21) with $\epsilon_i^t = \eta_i^t - (n \cdot \eta_i^t) n$ the perpendicular component of the vectorial noise η.

IV. INFERENCE FROM DATA

A. Model parametrization

As we argued earlier, tuning the parameters to match the correlation functions is equivalent to maximizing the likelihood, Eq. (7), or equivalently maximizing the log-likelihood $-\sum_t \mathcal{L}_t$, with which we will work from now on. To maximize the likelihood with respect to the two equivalent sets of parameters $\{\mathbf{J}_t^{(1)}, \mathbf{J}_t^{(2)}\}$ or $\{\mathbf{M}_t, \mathbf{A}_t\}$, we would need to observe a large number of random realizations of the same flock dynamics. This is impossible in practice due to limited data compared to the prohibitively large number of potential configurations of the bird positions that one would need to sample.

To overcome this problem, we need to introduce some additional assumptions about the interaction network and the form of the noise to simplify the parameter space and the

number of observables. These simplifications come naturally in the Markovian description parametrized by $\mathbf{M}_t$ and $\mathbf{A}_t$. From a biological standpoint, it is reasonable to assume that birds treat information from each interacting neighbor (the precise definition of "neighborhood" being left unspecified for the moment) equally, while keeping the memory of their own direction. Mathematically this translates into

$$M_{ij} = (1 - J\delta t n_i)\delta_{ij} + J\delta t n_{ij}, \tag{25}$$

where $n_{ij} = 1$ if j is one of i's neighbors, and 0 otherwise, and $n_i = \sum_j n_{ij}$ is the global number of neighbors interacting with bird i. (For ease of notation we omit the t index, even though n_{ij} depends on t.) The scalar parameter J now measures the alignment interaction strength. The errors made by different birds when trying to align with their neighbors can be assumed to be of the same amplitude and independent of each other, so that noise is uncorrelated and $\mathbf{A}$ is proportional to the identity

$$A_{ij} = [1/(2\delta t T)]\delta_{ij}. \tag{26}$$

Here T is a squared noise amplitude (the out-of-equilibrium equivalent of a temperature) that sets the level of disorder in the system.

B. Continuous time limit and equivalence with static maximum entropy at equilibrium

The scaling in δt ensures a well-defined continuous limit when $\delta t \to 0$, described by a Langevin equation. When $\delta t \to 0$, Eq. (21) becomes

$$\frac{d\boldsymbol{\pi}}{dt} = -J\mathbf{\Lambda}\boldsymbol{\pi} + \boldsymbol{\xi}(t), \tag{27}$$

where $\Lambda_{ij} = n_i \delta_{ij} - n_{ij}$, and $\xi_i(t)$ are independent and identically distributed (i.i.d.) Gaussian white noises with $\langle \xi_i(t)\xi_i(t') \rangle = 2T(d-1)\delta(t-t')$, with $\delta(x)$ Dirac's delta function.

This dynamical description can be reconciled with the static inference [20] in the special case of equilibrium dynamics, which is realized when n_{ij} is symmetric and constant in time. In this case, the spins can be described for $\delta t \to 0$ by a stationary distribution with the same form as in Eq. (6) and the steady-state couplings take the simple equilibrium value [20]

$$J_{ij}^{\text{stat}} = \frac{J}{T} n_{ij}. \tag{28}$$

We refer the reader to Appendix A for details of the derivation.

Even in this limit case where dynamical and static maximum entropy are consistent, the dynamical inference allows us to separate the coupling strength J from the temperature T, which the static inference cannot.

C. Likelihood maximization

The Lagrangian (19) can be rewritten as an explicit function of the relevant observables

$$\mathcal{L}_t(\boldsymbol{\pi}^{t+1} | \boldsymbol{\pi}^t) = -\frac{d-1}{2} \ln\left(\frac{\det \mathbf{A}_t}{(2\pi)^N}\right) + \frac{1}{2}\text{Tr}(\mathbf{C}_{t+1}\mathbf{A}_t^\dagger)$$
$$- \frac{1}{2}\text{Tr}(\mathbf{J}_t^{(2)}\mathbf{G}_t^\dagger) + \frac{1}{8}\text{Tr}(\mathbf{J}_t^{(2)\dagger}\mathbf{A}_t^{-1}\mathbf{J}_t^{(2)}\mathbf{C}_t^\dagger), \tag{29}$$

TABLE I. Empirical correlation functions used in the text.

C_s^1	$(1/N)\sum_i (\pi_i^{t+1})^2$	C_{int}	$(1/Nn_c)\sum_{ij} n_{ij}\pi_i^t\pi_j^t$
C_s	$(1/N)\sum_i (\pi_i^t)^2$	C'_{int}	$(1/Nn_c^2)\sum_{ijk} n_{ij}n_{ik}\pi_j^t\pi_k^t$
G_s	$(1/N)\sum_i \pi_i^{t+1}\pi_i^t$	G_{int}	$(1/Nn_c)\sum_{ij} n_{ij}\pi_i^{t+1}\pi_j^t$
$\tilde{C}_s$	$(1/Nn_c)\sum_i n_i(\pi_i^t)^2$	$\hat{C}_s$	$(1/Nn_c^2)\sum_{ij} (n_i\pi_i^t)^2$
$\tilde{G}_s$	$(1/Nn_c)\sum_i n_i\pi_i^{t+1}\pi_i^t$	$\tilde{C}_{\text{int}}$	$(1/Nn_c^2)\sum_{ij} n_i n_{ij}\pi_i^t\pi_j^t$

where $\mathbf{C}_t = \boldsymbol{\pi}^t(\boldsymbol{\pi}^t)^\dagger$ and $\mathbf{G}_t = \boldsymbol{\pi}^{t+1}(\boldsymbol{\pi}^t)^\dagger$.

Taking the specific form of Eqs. (25) and (26) for $\mathbf{M}_t$ and $\mathbf{A}_t$ we obtain a formula for $\mathcal{L}_t$ that only depends on two parameters, the interaction strength J and the "effective temperature" T

$$\mathcal{L}_t = \frac{d-1}{2}\ln(2T\delta t) - 2Jn_c\delta t[\tilde{C}_s - C_{\text{int}} - \tilde{G}_s + G_{\text{int}}]$$
$$+ (Jn_c\delta t)^2[\hat{C}_s - 2\tilde{C}_{\text{int}} + C'_{\text{int}}] + C_s^1 + C_s - 2G_s, \quad (30)$$

with $n_c = (1/N)\sum_i n_i$. The number of independent observables appearing in $\mathcal{L}_t$ is also drastically reduced to a handful of empirical integrated pair correlation functions defined in Table I. These correlations can be evaluated over pairs of consecutive configurations, or averaged over the entire sequence if we work with time-independent parameters and steady-state dynamics.

Maximizing the log-likelihood with respect to J and T, $\partial\mathcal{L}_t/\partial T = 0$ and $\partial\mathcal{L}_t/\partial J = 0$, yields simple analytical expressions for the parameters as a function of the empirical correlation functions

$$J = \frac{1}{n_c}\frac{\Omega + (d-1)T_0}{C'_{\text{int}} + \hat{C}_s - 2\tilde{C}_{\text{int}}}, \quad (31)$$

$$T = T_0 + \frac{C_s^1 - C_s}{2(d-1)\delta t} - \frac{Jn_c\delta t}{2(d-1)}\left(\frac{\tilde{C}_s - \tilde{G}_s}{\delta t} + \Omega\right), \quad (32)$$

where

$$T_0 = \frac{C_s - G_s}{\delta t(d-1)}, \quad \Omega = \frac{G_{\text{int}} - C_{\text{int}}}{\delta t}. \quad (33)$$

Appendix B contains detailed formulas for calculating the maximum likelihood after replacing Eqs. (31) and (32) in Eq. (30).

The leading-order temperature T_0 is the derivative of a self-correlation function, and obeys the standard fluctuation-dissipation relationship found in equilibrium dynamics. The term Ω is related to the dynamics of the network. In particular, at steady state $dC_{\text{int}}/dt = 0$ implies $\Omega \propto \sum_{ij}\pi_i\pi_j dn_{ij}/dt$.

When the system is at steady state, the inference formulas Eqs. (31) and (32) reduce to the static inference formula, derived in a slightly different form in [20]

$$\frac{Jn_c}{T} \approx \frac{d-1}{\hat{C}_s - \tilde{C}_{\text{int}}}. \quad (34)$$

A detailed derivation is given in Appendix C.

To apply Eqs. (31) and (32) to data, one still needs to specify the neighboring matrix n_{ij}. In the absence of prior information, one of the simplest possibilities is to assume that each bird interacts with the first n_c neighbors [20]. An alternative choice would be to define neighbors according to a metric rule, each bird interacting with neighbors within a given distance r_c. In both cases an extra parameter is introduced, either the "topological" interaction range n_c or the metric range r_c, that can also be inferred by likelihood maximization. Another scheme is to define neighbors through a Voronoi tassellation [30], as in the topological VM [14]. The likelihoods between different neighborhood definitions may also be compared to find the one closest to optimality.

V. TEST ON ARTIFICIAL DATA

We tested our dynamical inference method on synthetic data generated from a slight generalization of the topological VM on a two-dimensional torus of linear size $L = 32$ with $N = 1024$ particles

$$\theta_i^{t+\delta t} = \text{Arg}\left[s_i^t + J_V\delta t\sum_j n_{ij}s_j^t\right] + \sqrt{\delta t}\,\xi_i^t, \quad (35)$$

$$r_i^{t+\delta t} = r_i^t + v_0\,\delta t\, s_i^{t+\delta t}, \quad (36)$$

where $s_i = (\cos\theta_i, \sin\theta_i)$, and $\text{Arg}(s)$ is the angle of vector s. The delta-correlated angular noise ξ_i^t is uniformly distributed in $[-\eta\pi, +\eta\pi]$, corresponding to an effective temperature $T_V = (\eta\pi^2)/6$ for $\delta t \to 0$. The Voronoi adjacency matrix n_{ij} has a nonuniform degree n_i, of mean $n_V = 6$. A spin-wave expansion of Eq. (35) leads to an expression of the form of Eqs. (21) to (26), with $J \approx J_V/(1 + J_V n_V\delta t)$ (see Appendix D). The degree of neighbor mixing is characterized by a single mixing parameter $\mu = \langle 1/(Nn_c)\sum_{ij}|dn_{ij}/dt|\rangle$, which quantifies how fast birds exchange neighbors. We performed simulations with time step $\delta t = 0.01$ in three regimes with slow, medium, and fast neighbor mixing ($\mu = 0.18, 0.35, 0.76$; $v_0 = 0.5, 1.0, 2.0$; $J_V = 1.0, 1.0, 0.1$; and $\eta = 0.3, 0.2, 0.12$, respectively), all of which display the same level of polarization $N^{-1}\|\sum_i s_i\| \approx 0.97$.

We then applied the inference procedure described in Eqs. (31) and (32) to the synthetic dataset generated by the simulations. In the inference we tried the choices for n_{ij} discussed above: the n_c nearest-neighbor (NN) topological rule, the metric rule where $n_{ij} = 1$ within a metric range r_c (and 0 outside), and the Voronoi rule (actually used to generate the data). The correlation functions were averaged over 10^3 different configurations in the stationary state, sampled from a single run at 100 time unit intervals, ensuring independent sampling.

The likelihood as a function of n_c can be computed with the NN rule using Eqs. (31), (32), and (30). The result is shown in the inset of Fig. 1 for the high mixing regime. Its maximum n_c^* corresponds to the most likely interaction range, from which the optimal J^* and T^* are computed via Eqs. (31) and (32). Figure 1 shows that the new dynamical procedure performs systematically better than the static approach described in [20] in predicting the mean interaction range n_c. The error made by the static inference is larger when neighbor mixing is higher and the dynamics is strongly out of equilibrium. That is because in the high-mixing case, the *effective* number of interacting neighbors, as inferred by the static approach,
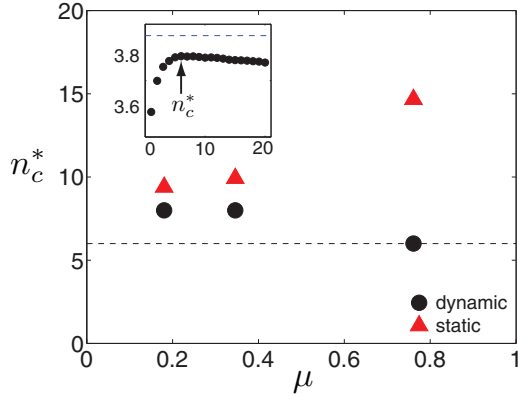
FIG. 1. (Color online) Comparison between dynamical and static inference. Data were generated using Voronoi neighborhood. The inference was performed by using either a Voronoi rule or a nearest-neighbor (NN) topological rule, parametrized by the number n_c of interacting neighbors. Main panel: The inferred number of interacting neighbors n_c^* is shown as a function of the mixing rate μ. Circles: dynamical inference; triangles: static inference. The dashed line marks the real average value $n_V = 6$. Static inference badly overestimates the number of interacting neighbors at large mixing, while dynamical inference does a much better job. Inset: Dynamical normalized log-likelihood $-\mathcal{L}_t/N$ as a function of n_c for the NN topological rule (circles). The maximum of this function gives the NN value of n_c^* reported in the main panel. The Voronoi likelihood (dashed line) is larger than the NN one, revealing that Voronoi was the actual generating rule. Data are for high mixing.

includes neighbors visited in the recent past in addition to the current ones, and thus is larger than the true n_c. By contrast, our new dynamical inference based on NN interactions overcomes this issue and hence performs very well, considering that the model used for the inference incorrectly assumes a constant n_c. Not surprisingly, the likelihood computed with the (correct) Voronoi topology is larger than with the (incorrect) NN one. The temperature T is well inferred in both cases (8% error), while the alignment strength J is well recovered when assuming Voronoi neighbors (3% error), and approximately with an NN topology (20% error).

If we apply the dynamical inference assuming a metric rule, but on the configurations generated by the Voronoi topological rule, we obtain significantly lower likelihoods, and a wrong $n_c^* \sim 3$ (see Fig. 2). This indicates that a topological rule is more consistent with the data than a metric one, as it should.

Hence the dynamical method not only gives us the correct interaction parameters, but also distinguishes the rule used to build the interaction network. The method achieves this by exploiting the different ways in which spatial density fluctuations translate into neighbors fluctuations. In the Voronoi network (the generating one), the number of neighbors n_i of each point fluctuates weakly around its mean value of 6 and in the NN topological case n_i does not fluctuate at all. With the metric rule, on the contrary, n_i has *very* large fluctuations, directly linked to the VM giant density fluctuations [14]. The large fluctuations of n_i make the metric correlation functions of Table I very different from their correct (Voronoi) value.
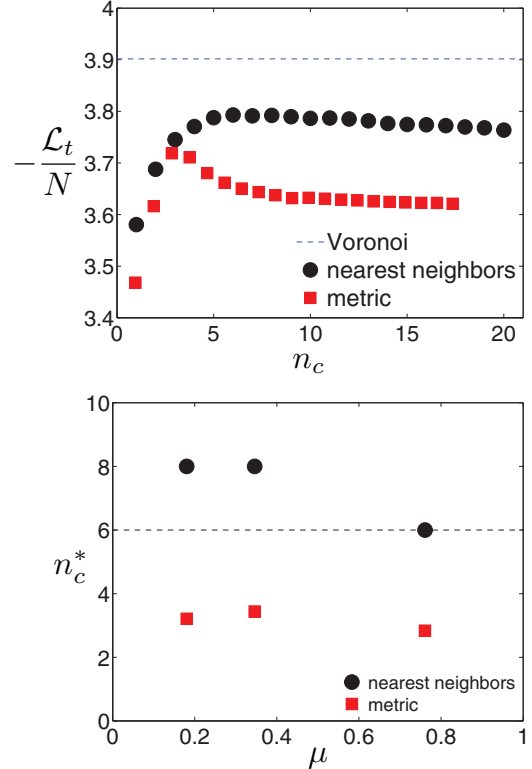

FIG. 2. (Color online) Upper panel: Comparison of the normalized log-likelihood for the nearest-neighbor and metric rules, as a function of n_c. For the metric case, for increasing values of r_c, the empirical $n_c = (1/N) \sum_i n_i$ is shown. The dashed line corresponds to the log-likelihood calculated with the Voronoi rule. Lower panel: Inferred interaction range n_c^* for the nearest-neighbor and metric cases, as a function of the mixing parameter μ.

VI. CONCLUSION

In summary, we have derived a dynamical maximum entropy method to infer the alignment dynamics of highly ordered animal groups from just two consecutive snapshots, using simple analytical formulas. Tests on synthetic data confirm the validity of our method. Our approach is very general and makes minimal, symmetry-based assumptions on the structure of the dynamics under investigation. In existing inference methods, proposed for schooling fish [31–33], the rules of motion were learned from small groups of fish by assuming specific forms for the turning forces caused by the proximity of the tank wall [31,33] as well as by the orientation [33] or position [31,32] of other fish. By contrast, in our approach the dynamics is entirely derived from the choice of observables that we choose to constrain. In addition, our inference procedure is straightforward and does not require complex multivariable optimization.

Our work emphasizes the need for a dynamical inference approach to out-of-equilibrium active matter systems, especially when there is no *a priori* knowledge of the time scales in the system, which is usually the case when dealing with experimental data. The analytical formulas we derive require the flock to be well polarized for the SW approximation to be valid. Even when this is not the case, integral expressions for the Markovian transition probabilities still exist [Eq. (14)], and

may be evaluated numerically, similarly to the case of Ising spins or spiking neurons where related approaches have been proposed [34–36]. Such a regime would be useful for studying weakly polarized groups such as midges [37]. The method also assumes homogeneous agents, a hypothesis that can be relaxed by treating differently some classes of agents, like those on the border as in [20]. It is also presented in the absence of external perturbations, which could be incorporated by adding local or global fields to the action, Eq. (8).

The effective alignment forces that we derive from the maximum entropy principle are not built into the approach, but rather follow from the local correlative structure of the data. In cases where polar order does not stem from alignment forces, the correlation functions between birds may not be the right choice of observables, and the approach should be adapted accordingly by adding other observables to be constrained. Ultimately, when applying the method to real data, the choice of observables should be validated by testing the predictions of the model on observables that were not used to construct it, such as higher-order correlation functions, as was done for example in [20].

Other forces than alignment forces may be at play in real systems, such as cohesion forces in flocks, which tend to attract birds towards the rest of the group. One could be worried that they might interfere with the inference of the alignment dynamics. However, in topological models moderate cohesion forces mostly prevent the diffusive spreading of birds' positions (otherwise observed in cohesionless models in open space [14]) keeping the mixing rate μ finite, and do not have much of an effect on the bulk dynamics. At first approximation they can be neglected in the bulk and we expect our inference method to be robust to their presence. In fact, it was shown that the alignment parameters of a system simulated with cohesion forces were correctly inferred by the static inference procedure, even though the model explicitly ignored cohesion [20].

Our approach is applicable to many systems where collective motion is observed, including moving animal groups [38], bacterial colonies [39,40], motility assays [41], collective motion of epithelial cells [42], vibrated polar disks [43,44], colloids [45], or nematic order [46]. The method should be tailored to each case. For example, the symmetry between the front and back can be broken by parametrizing the interaction network to account for different behaviors with respect to front and back neighbors as in [47]. In the case of nematic order, local correlation functions should be replaced by local nematic-order parameters. "Pursuit" behaviors such as those observed in [31,32] could be captured by constraining correlation functions between the orientation and relative positions between individuals. As we have already stressed, each system has its natural symmetries, and the choice of relevant observables to be constrained should also reflect prior biological or physical insight about the considered system.

Throughout this work we have assumed that δt is equal to (or smaller than) the real update time lag, namely the biological time scale. This may not be true for some datasets, as the sampling time of the experimental equipment is likely to be larger than the neural update time actually used by animals. This is certainly the case for the starling data of [20]. When

this happens, the experimental time series is a coarse-grained version of the real dynamics, so that the present method would probably provide a time-renormalized value of the interaction parameters. It would therefore be important to generalize our equations to deal with this issue. Other generalizations include the extension to more than two consecutive times (with the risk of accumulating tracking errors), which would allow one to infer dynamics characterized by linear, not diffusive, dispersion relations [48].

APPENDIX A: CONTINUOUS-TIME, ADIABATIC LIMIT, AND STATIC MAXIMUM ENTROPY

Here we show that, when the interaction network is symmetric and evolves adiabatically, the continuous-time limit of the dynamical maximum entropy stochastic process admits a steady state that is given by the static maximum entropy distribution.

We start from the Langevin equation (27)

$$\frac{d\boldsymbol{\pi}}{dt} = -J\boldsymbol{\Lambda}\boldsymbol{\pi} + \boldsymbol{\xi}(t), \tag{A1}$$

where $\Lambda_{ij} = n_i \delta_{ij} - n_{ij}$, and $\xi_i(t)$ are i.i.d. Gaussian white noises with $\langle \xi_i(t)\xi_i(t') \rangle = 2T(d-1)\delta(t-t')$, where $\delta(x)$ is Dirac's delta function.

When $\boldsymbol{\Lambda}$ varies slowly with time, Eq. (A1) can be formally integrated as follows:

$$\boldsymbol{\pi}(t) = \int_{-\infty}^{t} dt' \, e^{-J\boldsymbol{\Lambda}(t-t')}\boldsymbol{\xi}(t'). \tag{A2}$$

If, in addition, $\boldsymbol{\Lambda}$ is symmetric, the system reaches some equilibrium steady state. More precisely, the collective mode that is parallel to $\mathbf{u}$, which corresponds to the average direction of the flock $(1/N)\sum_i \pi_i$, and follows an unconstrained random walk, as it corresponds to a the zero mode of $\boldsymbol{\Lambda}$, $\boldsymbol{\Lambda}\mathbf{u} = 0$. All the other modes that are orthogonal to $\mathbf{u}$ are bounded by a restoring force. The steady-state distribution of $\boldsymbol{\pi}$ is therefore Gaussian, with $C_{ij} = \mathrm{Cov}(\pi_i, \pi_j)$ satisfying

$$J\boldsymbol{\Lambda}\mathbf{C} = (d-1)T\left(\mathbf{1} - \frac{\mathbf{u}\mathbf{u}^{\dagger}}{N}\right), \tag{A3}$$

where $\mathbf{1}$ is the identity matrix.

Remarkably, in the spin-wave approximation, this distribution is the same as the one obtained by the principle maximum entropy constrained by the static correlation functions

$$P(\mathbf{s}) = \frac{1}{Z}\exp\left(\frac{1}{2}\sum_{i \neq j} J_{ij}^{\mathrm{stat}} s_i s_j\right), \tag{A4}$$

with

$$J_{ij}^{\text{stat}} = \frac{J}{T} n_{ij}. \tag{A5}$$

One can check this by expanding Eq. (A4) at small π, after setting n to be the average direction of the flock, so that $\sum \pi_i = 0$, and

$$P(\boldsymbol{\pi}) \propto \delta\left(\sum_i \pi_i\right) \exp\left(-\frac{J}{2T} \sum_{ij} \Lambda_{ij} \pi_i \pi_j\right). \tag{A6}$$

By virtue of Gaussian integration rules, this distribution has the same covariance as Eq. (A3), and therefore is identical.

APPENDIX B: LIKELIHOOD MAXIMIZATION

In this Appendix we give detailed expressions for maximizing the likelihood. The likelihood reads:

$$\frac{\mathcal{L}}{N} = \frac{d-1}{2} \ln 2T\delta t + \frac{\hat{\mathcal{L}}}{4T\delta t}, \tag{B1}$$

where

$$\hat{\mathcal{L}} = C_s^1 + C_s - 2\alpha\tilde{C}_s + \alpha^2\hat{C}_s + 2\alpha(C_{\text{int}} - \alpha\tilde{C}_{\text{int}})$$
$$+ \alpha^2 C_{\text{int}}' - 2\alpha G_{\text{int}} - 2(G_s - \alpha\tilde{G}_s), \tag{B2}$$

and $\alpha = Jn_c\delta t$.

There are three parameters to optimize over: the interaction strength J, the interaction range n_c, and the "temperature" T which sets the strength of noise. This last one is simply given by the condition $\partial\mathcal{L}/\partial T = 0$, which yields

$$T = \frac{\hat{\mathcal{L}}}{2(d-1)\delta t}. \tag{B3}$$

At this optimum value of T, we have

$$\frac{\mathcal{L}}{N} = \frac{d-1}{2} \{\ln[\hat{\mathcal{L}}/(d-1)] + 1\}. \tag{B4}$$

Minimizing $\hat{\mathcal{L}}$, $\partial\hat{\mathcal{L}}/\partial\alpha$, then yields the optimum value of α:

$$\alpha = \frac{C_{\text{int}} - \tilde{C}_s + \tilde{G}_s - G_{\text{int}}}{2\tilde{C}_{\text{int}} - C_{\text{int}}' - \hat{C}_s}. \tag{B5}$$

At this optimum, one has $\hat{\mathcal{L}} = C_s^1 + C_s - 2G_s + \tilde{\mathcal{L}}$, where

$$\tilde{\mathcal{L}} = \frac{(C_{\text{int}} - \tilde{C}_s + \tilde{G}_s - G_{\text{int}})^2}{2\tilde{C}_{\text{int}} - C_{\text{int}}' - \hat{C}_s} \tag{B6}$$

is the only term that depends on the interaction matrix n_{ij}. Therefore, to find the optimum interaction range n_c in the case of the nearest-neighbor model, one just needs to minimize $\tilde{\mathcal{L}}(n_c)$.

APPENDIX C: CONSISTENCY WITH THE STATIC APPROACH

Here we show how the dynamical inference equations derived above reduce to the static inference formulas of [20] in the appropriate limits.

To recover the static inference equations, we start by rewriting the dynamical inference equations, Eqs. (B3)

and (B5), explicitly

$$J = \frac{1}{n_c} \frac{\Omega + (d-1)T_0}{C_{\text{int}}' + \hat{C}_s - 2\tilde{C}_{\text{int}}}, \tag{C1}$$

$$T = T_0 + \frac{C_s^1 - C_s}{2(d-1)\delta t} - \frac{J n_c\delta t}{2(d-1)} \left(\frac{\tilde{C}_s - \tilde{G}_s}{\delta t} + \Omega\right), \tag{C2}$$

with $n_c = (1/N)\sum_i n_i$ and

$$T_0 = \frac{C_s - G_s}{\delta t(d-1)}, \quad \Omega = \frac{G_{\text{int}} - C_{\text{int}}}{\delta t}. \tag{C3}$$

When the system is at steady state, we have $C_s^1 \approx C_s$ and $\Omega \approx (2Nn_c)^{-1} \sum_{ij} \pi_i \pi_j \frac{dn_{ij}}{dt}$ [directly from the definitions in Table I of the main text and Eq. (C3)]; the second term in Eq. (C2) cancels and $T \approx T_0$ for small δt. If we further assume that data were actually generated by *exactly* the class of models we are trying to infer (which may not be the case in general, as we are looking at effective descriptions), we have exactly $T = T_0$. If, in addition, the neighbor changes are slow, then $\Omega \approx 0$ and Eq. (A3) implies $\tilde{C}_{\text{int}} \approx C_{\text{int}}'$. Equation (C1) thus gives

$$\frac{Jn_c}{T} \approx \frac{d-1}{\hat{C}_s - \tilde{C}_{\text{int}}}, \tag{C4}$$

which is the result of the static inference. Note, however, that in addition to recovering the alignment strength, the dynamical inference procedure allows us to separate the interaction coupling J from the temperature T.

APPENDIX D: SPIN-WAVE EXPANSION OF THE TOPOLOGICAL VICSEK MODEL

As described in the main text, to test our dynamical inference method we generated synthetic data with the topological VM defined by

$$\theta_i^{t+\delta t} = \text{Arg}\left[s_i^t + J_V\delta t \sum_j n_{ij}s_j^t\right] + \sqrt{\delta t}\, \xi_i^t, \tag{D1}$$

$$r_i^{t+\delta t} = r_i^t + v_0\,\delta t\, s_i^{t+\delta t}. \tag{D2}$$

In this section, we show that Eq. (D1) is in fact equivalent in the spin-wave limit to an update equation of the same kind as Eqs. (21), (25), and (26). To this aim, it is convenient to rewrite Eq. (D1) in the following equivalent form:

$$s_i^{t+\delta t} = \frac{s_i^t + J_V\delta t \sum_j n_{ij}s_j^t}{\left\| s_i^t + J_V\delta t \sum_j n_{ij}s_j^t \right\|} + \sqrt{\delta t}\, \epsilon_i^t, \tag{D3}$$

$$r_i^{t+\delta t} = r_i^t + v_0\,\delta t\, s_i^{t+\delta t}, \tag{D4}$$

where ϵ_i is a delta-correlated noise perpendicular to s_i with variance $2(d-1)T_V$ [i.e., whose effect is the same as the angular noise appearing in Eq. (D1)].

In the large polarization regime we can perform a spin-wave expansion $s_i = \pi_i + n\sqrt{1 - \pi_i^2}$, where n is a vector representing the global direction of motion and π_i is the component of the direction s_i perpendicular to n. We can now expand the normalization at the right-hand side in Eq. (D3)

with respect to π_i^2 to get

$$\left\| s_i^t + J_V \delta t \sum_j n_{ij} s_j^t \right\| = 1 + \delta t J_V n_i + \mathrm{O}(\pi^2),\tag{D5}$$

where $n_i = \sum_j n_{ij}$. Equation (D3) then leads to the following update equation for the $\{\pi_i\}$:

$$\pi_i^{t+\delta t} = \frac{\pi_i^t + \delta t J_V \sum_j n_{ij} \pi_j^t}{1 + \delta t J_V n_i} + \sqrt{\delta t} \epsilon_i = \left(1 - \delta t \frac{J_V}{1 + \delta t J_V n_i} \right) \pi_i^t + \delta t \frac{J_V}{1 + \delta t J_V n_i} \sum_j n_{ij} \pi_j^t + \sqrt{\delta t} \epsilon_i.\tag{D6}$$

When δt is small, we can disregard the fluctuations in n_i and Eq. (D6) is of the same form of Eq. (21) with the parametrization defined in Eqs. (25) and (26) and

$$J = \frac{J_V}{1 + \delta t J_V n_V}.\tag{D7}$$

---

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

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

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

[4] E. Bertin, M. Droz, and G. Grégoire, Phys. Rev. E **74**, 022101 (2006).

[5] E. Bertin, M. Droz, and G. Grégoire, J. Phys. A **42**, 445001 (2009).

[6] T. Ihle, Phys. Rev. E **83**, 030901 (2011).

[7] H. Chaté, F. Ginelli, G. Grégoire, and F. Raynaud, Phys. Rev. E **77**, 046113 (2008).

[8] P. Szabó, M. Nagy, and T. Vicsek, Phys. Rev. E **79**, 021908 (2009).

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

[10] J. Toner, Phys. Rev. E **86**, 031918 (2012).

[11] S. Ramaswamy, Annu. Rev. Condens. Matter Phys. **1**, 323 (2010).

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

[13] J. Toner and Y. H. Tu, Phys. Rev. E **58**, 4828 (1998).

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

[15] A. Cavagna, I. Giardina, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, and V. Zdravkovic, Animal Behaviour **76**, 217 (2008).

[16] A. Cavagna, I. Giardina, A. Orlandi, G. Parisi, and A. Procaccini, Animal Behaviour **76**, 237 (2008).

[17] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, and V. Zdravkovic, Animal Behaviour **76**, 201 (2008).

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

[19] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, and M. Viale, Proc. Natl. Acad. Sci. USA **107**, 11865 (2010).

[20] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. M Walczak, Proc. Natl. Acad. Sci. USA **109**, 4786 (2012).

[21] W. Bialek, A. Cavagna, I. Giardina, T. Mora, O. Pohl, E. Silvestri, M. Viale, and A. Walczak, arXiv:1307.5563v1 [Proc. Natl. Acad. Sci. USA (to be published)].

[22] E. Schneidman, M. J Berry, R. Segev, and W. Bialek, Nature (London) **440**, 1007 (2006).

[23] J. Shlens, G. D. Field, J. L. Gauthier, M. I. Grivich, D. Petrusca, A. Sher, A. M. Litke, and E. J. Chichilnisky, J. Neurosci. **26**, 8254 (2006).

[24] M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, and T. Hwa, Proc. Natl. Acad. Sci. USA **106**, 67 (2009).

[25] T. Mora, A. M. Walczak, W. Bialek, and C. G. Callan, Proc. Natl. Acad. Sci. USA **107**, 5405 (2010).

[26] M. Santolini, T. Mora, and V. Hakim, arXiv:1302.4424v1.

[27] T. Mora and W. Bialek, J. Stat. Phys. **144**, 268 (2011).

[28] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, Rev. Mod. Phys. **85**, 1115 (2013).

[29] F. Dyson, Phys. Rev. **102**, 1217 (1956).

[30] G. Voronoi, J. Reine Angew. Math. **133**, 97 (1907).

[31] J. E. Herbert-Read, A. Perna, R. P. Mann, T. M. Schaerf, D. J. T. Sumpter, and A. J. W. Ward, Proc. Natl. Acad. Sci. USA **108**, 18726 (2011).

[32] Y. Katz, K. Tunstrøm, C. C. Ioannou, C. Huepe, and I. D. Couzin, Proc. Natl. Acad. Sci. USA **108**, 18720 (2011).

[33] J. Gautrais *et al.*, PLoS Comput. Biol. **8**, e1002678 (2012).

[34] O. Marre, S. El Boustani, Y. Frégnac, and A. Destexhe, Phys. Rev. Lett. **102**, 138101 (2009).

[35] Y. Roudi and J. Hertz, Phys. Rev. Lett. **106**, 048702 (2011).

[36] J. C. Vasquez, O. Marre, A. G. Palacios, M. J. Berry Ii, and B. Cessac, J. Physiol. (Paris) **106**, 120 (2012).

[37] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, S. Melillo, L. Parisi, O. Pohl, B. Rossaro, E. Shen, E. Silvestri, and M. Viale, arXiv:1307.5631v1.

[38] J. K. Parrish and W. M. Hamner, *Animal Groups in Three Dimensions* (Cambridge University Press, Cambridge, England, 1997).

[39] H. P. Zhang, A. Be'er, E.-L. Florin, and H. L. Swinney, Proc. Natl. Acad. Sci. USA **107**, 13626 (2010).

[40] X. Chen, X. Dong, A. Be'er, H. L. Swinney, and H. P. Zhang, Phys. Rev. Lett. **108**, 148101 (2012).

[41] Y. Sumino, K. H. Nagai, Y. Shitaka, D. Tanaka, K. Yoshikawa, H. Chaté, and K. Oiwa, Nature (London) **483**, 448 (2012).

[42] N. Sepúlveda, L. Petitjean, O. Cochet, E. Grasland-Mongrain, P. Silberzan, and V. Hakim, PLoS Comput. Biol. **9**, e1002944 (2013).

[43] J. Deseigne, O. Dauchot, and H. Chaté, Phys. Rev. Lett. **105**, 098001 (2010).

[44] C. A. Weber, T. Hanke, J. Deseigne, S. Léonard, O. Dauchot, E. Frey, and H. Chaté, Phys. Rev. Lett. **110**, 208001 (2013).

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

[46] F. Ginelli, F. Peruani, M. Bär, and H. Chaté, Phys. Rev. Lett. **104**, 184502 (2010).

[47] P. Romanczuk, I. D. Couzin, and L. Schimansky-Geier, Phys. Rev. Lett. **102**, 010602 (2009).

[48] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, T. S. Grigera, A. Jelić, S. Melillo, L. Parisi, O. Pohl, E. Shen, and M. Viale, arXiv:1303.7097v1.