SAPIENZA
UNIVERSITÀ DI ROMA

SCUOLA DOTTORALE VITO VOLTERRA

Microscopic Dynamics of Polar Active Systems: Inference Methods and Signatures of Irreversibility for Stochastic Models

FEDERICA FERRETTI

Advisor
IRENE GIARDINA

Collaborators
THIERRY MORA
ALEKSANDRA WALCZAK

2021, October

Federica Ferretti: *Microscopic Dynamics of Polar Active Systems: Inference Methods and Signatures of Irreversibility for Stochastic Models,* © 2021, October.

E-MAIL:
f.ferretti@uniroma1.it

ABSTRACT

Flocking denotes the spontaneous onset of collective motion in systems of self-propelled agents, of which groups of birds are a prototypical example. A pretty coherent corpus of theoretical models has been introduced over the past three decades to explain how this collective behavior arises from microscopic interactions, revealing that the emergence of polar order is a manifestation of the non-equilibrium character of the dynamics. The confront with experimental data allowed for the validation and refinement of those models, and, in some cases, even for the application of quantitative inference approaches.

In this thesis we employ standard methods from stochastic calculus to study problems related to the microscopic dynamics of such systems. Motivated by the availability of data collected by the CoBBS team, we firstly derive a novel Bayesian inference method for the inertial dynamics of flocks. Our inference scheme is based on a previously introduced model (the Inertial Spin Model), which is non-Markovian in the observed variables' space. This feature raises serious technical problems, when combined with discrete-time recordings, and is common to many stochastic dynamic systems. The analytical method we propose for the Inertial Spin Model applies in fact to a larger class of processes; examples are illustrated. We also exploit an analogy between the Renormalization Group and augmentation techniques used to infer partially observed SDEs to provide an alternative proof of the lack of finite-dimensional delay vector embeddings for stochastic dynamical systems.

The second focus of this thesis concerns the investigation of non-equilibrium effects in simple models for polar active matter. It is known that the emergence of polar order in systems of aligning self-propelled particles is due to the non-equilibrium character of the dynamics. We quantify the distance from equilibrium through the entropy production rate, which we measure from numerical simulations of interacting active Brownian particles. We investigate two kinds of short-ranged interaction rules, based on different notions of metrics. We find that the entropy production rate is maximal at the transition, while two equilibrium limits are reached in the deeply ordered (perfect flock) or completely disordered (ideal active gas) phase. We pivot on the entropy production rate to study how irreversibility constrains asymmetries in the steady state distribution of microstates. In the presence of pairwise forces, robust signatures of irreversibility are visible in the two-particle density, as confirmed by numerical simulations. On the contrary, in the presence of multi-particle interactions, irreversibility directly constrains only correlations among a higher number of particles. All these correlations are typically neglected in the derivation of hydrodynamic equations for polar active matter through kinetic approaches.

SOMMARIO

Con *flocking* si intende il fenomeno per cui sistemi di agenti autopropulsi si muovono spontaneamente nella stessa direzione: gli stormi di uccelli (*flocks*) sono un esempio tipico di tali sistemi. Molti modelli teorici si sono affastellati nell'utlimo trentennio per spiegare come questo comportamento collettivo possa emergere dalle interazioni microscopiche tra gli individui che compongono il gruppo, rivelando che esso è una specifica manifestazione del carattere di non-equilibrio di questi sistemi attivi. Il confronto coi dati sperimentali ha permesso di convalidare e raffinare questi modelli e, in alcuni casi, persino di tentare approcci quantitativi di inferenza statistica.

In questa tesi sono impiegati metodi standard del calcolo stocastico per studiare problemi legati alla dinamica microscopica di tali sistemi. Motivati dalla disponibilità di dati raccolti dal gruppo CoBBS, si è derivato in primis un nuovo metodo di inferenza Bayesiana per la dinamica inerziale degli stormi di uccelli. Il nostro schema di inferenza si basa su un modello precedentemente introdotto (Inertial Spin Model), che, a causa del suo carattere non-Markoviano, solleva difficoltà tecniche quando viene combinato con un'osservazione a tempi discreti. Questo fatto è comune a molti sistemi dinamici stocastici, e il metodo proposto si applica in realtà a una classe di processi più ampia, di cui sono illustrati degli esempi. Inoltre, è possibile sfruttare un'analogia tra alcune tecniche usate nei problemi di inferenza e il Gruppo di Rinormalizzazione, che mostra intuitivamente l'assenza di *delay vector embeddings* per osservazioni parziali di processi stocastici.

Il secondo argomento di ricerca di questo lavoro di tesi riguarda lo studio di effetti di non-equilibrio in semplici modelli di materia attiva polare. È noto che l'emergere di ordine polare in sistemi di particelle autoproulse è dovuto a una violazione della simmetria sotto inversioni temporali. Questa violazione è quantificata dal tasso di produzione di entropia, il quale è stato misurato in simulazioni numeriche di sistemi di particelle Browniane attive. Sono stati investigati due tipi di interazioni locali, basate su diverse nozioni di metrica. In entrambi i casi, la produzione di entropia è massima alla transizione, dove la motilità ha un impatto massimo sull'interazione tra le particelle, mentre due limiti di equilibrio sono raggiunti nella fase fortemente polarizzata o totalmente disordinata. Si è studiato inoltre l'effetto dell'irreversibilità sulla distribuzione di probabilità stazionaria dei microstati del sistema. Una produzione di entropia non nulla impone delle robuste asimmetrie, visibili nelle distribuzioni di due o più particelle. Si noti che le correlazioni associate a tali distribuzioni vengono tipicamente trascurate nelle teorie cinetiche usate per derivare equazioni idrodinamiche per materia attiva polare.

CONTENTS

1 INTRODUCTION

1.1 A LONG STORY SHORT

Traditionally, as physicists we hunt for a mathematical understanding of the natural phenomena that we observe in the world that surrounds us. Centuries of 'modern' scientific investigation have taught us how to proceed when confronted with a new empirical event to explain, but also pushed the boundaries of the 'still unexplained' progressively further. This trivial consequence of scientific research takes place at many different levels — theoretical, experimental, methodological — and is common to all areas of Physics. Among them, however, Statistical Mechanics occupies a peculiar position: as the natural framework of complexity, non-equilibrium and multi-scale problems, in recent times it broadly invaded, with its methods and models, fields that had long remained elusive from quantitative analysis but now appear as contiguous.

One very fortunate example is that of living matter. Any assembly of many living units interacting with each other belongs to this class. Examples range from tissues and clusters of migrating cells to colonies of bacteria, swarms of insects, or large animal groups — including human crowds — exhibiting collective motion. A crucial element is the multitude of components: while the study of most living systems at a deeper level of detail is still a challenge lacking a unifying picture, the macroscopic properties of large groups of interacting units have been successfully understood in the framework of *active matter*. Active matter conventionally designates systems composed of units which are able to transform free energy — internally stored or collected from the environment — into motion. Such definition specifies the peculiar way in which such systems are brought out of equilibrium, and finds obvious inspiration and application to the biological world.

After the seminal works by Vicsek and Toner and Tu in 1995, the field flourished and imposed itself as a standalone branch of research. Since then, theoretical physicists, attracted by the wealth of novel phenomena that activity makes possible, have borrowed intuition from Condensed Matter and Non-equilibrium Physics to adapt their beautiful formal constructions to this new field. Many descriptions at various levels have been proposed. The first effort concerns modeling the self-propulsion mechanism of single constituents: this is generally done through effective stochastic descriptions that aim at reproducing the persistence and diffusion properties of motion, as the final results of the metabolic, mechanical or chemical processes which allow the particle to self-propel and which are not taken explicitly into account. The simplest of these single-particle models have been enriched with interactions and successfully used to build microscopic agent-based models exhibiting emergent macroscopic properties, such as phase transitions, collective ordering, apparent criticality, formation of patterns. Many of these emergent phenomena have

been observed in real systems. Another level of description is the hydrodynamic one: here the relevant quantities which describe the behavior of large extended systems are continuous fields. General rules are available to derive continuous dynamical field theories either from first principles, on the basis of symmetries and asymptotic expansions, or from direct coarse-graining of microscopic models. The plurality of descriptions and the known connection between them have allowed to achieve a quite comprehensive understanding of the subject. Extensive and influential reviews added on over the years to track the evolution of this fast-developing field [191, 189, 158, 126, 48, 5, 92].

At the same time, the theoretical progress has fostered (and has been fostered by) an intense experimental research activity. Major advances regarded the more tractable subclass of *inanimate* active matter. We can distinguish two main types of model systems in this subclass: on one hand, biological extracts like actin filaments, biopolymers and motor proteins [164, 181, 163]; on the other hand, suspensions of man-made artificial swimmers or active colloids, where the self-propulsion mechanism is systematically engineered by means of electrical [139, 18, 118], mechanical [64, 110], light-induced [143], or chemical forces [186, 16]. However, the ever increasing control over biology experiments has also allowed for the design of tunable *living* model systems, like E-coli, spermatozoa, flagellated cells [67, 197, 65] and, more recently, myxobacteria, epithelial and cancer cells [57, 182, 155, 161].

Experiments on animal groups, spanning much larger scales and involving much more complex organisms, have so far escaped this kind of systematization. For this reason, despite the original motivation for the introduction of the Vicsek model comes from the desire to find a minimalist description of the phenomenon of bird flocking, the quantitative matching between theoretical models and empirical data is a bit scarcer in this case. This is not only due to the lower amount of data, but also to the issue of reproducibility of experimental conditions (distinguishing between relevant and irrelevant details is a far from trivial task, bouncing back and forth between theory and observation) and to the fact that applying active matter models to animal systems requires a higher degree of reductionism compared to the previous examples. The constituents are indeed insects, birds, fish, mammals, all of them having elaborate cognitive powers and social functions and a big phase space of possible individual behaviors.

Nonetheless, some significant progress has been made in the last decades with the observation of unprecedentedly large groups of animals on the move. These observations proved that minimal active matter models are in fact able to capture the emergent features of such systems and to give a satisfactory quantitative account of them. When interested in macroscopic properties of a large sized system, we can then neglect the complexity of constituent elements, and describe their individual behavior through a handful of degrees of freedom — e.g. position and direction of motion in the classical Vicsek model. In some cases, experimental findings have also suggested a deep revision of existing models: for instance, the analysis of a large data-set of natural flocks of starlings allowed to understand that birds' alignment is due to topological rather than metric interaction rules [4]. Since then, a thorough in-

vestigation of the different properties of topological versus metric models occupied part of the theoretical scene for several years [48, 151, 39, 129].

In summary, times are mature to quantitatively match models and experiments even for animals. This understanding process obviously needs a close dialogue between the study of theoretical models and the analysis of empirical data. This PhD work focuses on two different projects whose general purpose fits in this broad picture.

The first project consists in the development of a statistical inference method for a flocking model with 'inertia'. Previous empirical findings revealed how the modelling of the dynamics of natural flocks of starlings requires this ingredient. Motivated by the availability of those data, we developed a maximum-likelihood scheme which can be efficiently used for inferring the damped stochastic dynamics of large collections of interacting particles. The main difficulty in the task comes from the presence of inertia, which raises general issues concerning Markovianity and phase space reconstruction for partially observed noise-driven systems. The origin of these problems can be tracked down to the lack of exact delay vector embeddings, a long-standing problem in parametric inference of dynamical systems, which marks the striking difference between stochastic and deterministic evolution. As a byproduct of our investigation, we proposed a physical interpretation of this well-known fact based on a formal analogy with the Renormalization Group construction.

The second project consists in a bottom-up investigation on the effects of irreversibility in Vicsek-style models. We combine stochastic thermodynamics, numerical simulations and kinetic theory to search for signatures of time reversal symmetry breaking (beyond the onset of collective motion) in two-dimensional polar active systems with alignment interactions. The flavor of this second project is to a greater extent theoretical, but we could identify quite general and robust features (i.e. not fully model-dependent) related to the non-equilibrium nature of the system, which it should be possible to check on various kinds of real data.

1.2 ORGANIZATION OF THE MANUSCRIPT

The dissertation is structured as follows. Some background concepts about dilute dry active matter which will be frequently recalled over all the manuscript are reviewed in the first section of Chapter 2. In the same Chapter, microscopic models of flocking are introduced and discussed. In particular, the Inertial Spin Model introduced by Cavagna et al. is reviewed in some detail. Chapter 3 and Chapter 4 are respectively dedicated to the two projects which have been developed as parts of this PhD thesis. Each chapter contains an introductory section about motivation and context of our work and a synthetic review of the employed methods, before presenting results from published and submitted papers [74, 73] (in Chapter 3) and original not-yet-published material (Chapter 4). I will close each of these two chapters discussing unaccomplished tasks and questions that are left unanswered. A Summary concludes the manuscript.

2 FLOCKING MODELS

Active matter is a broad umbrella. It comprises various types of many-body systems, whose common characteristic is being made of self-propelled particles (SPP). A useful standard classification is based on the three following aspects [126, 48]: (i) momentum conservation, (ii) type of interaction, and (iii) density of the system. The class of interest for the work presented in the following Chapters is that of *polar dilute dry* active matter, where (i) total momentum is not conserved because the fluid which surrounds the particles is not included in the description, (ii) interactions consists in ferromagnetic alignment of the directional degrees of freedom of the constituents of the system, and (iii) the diluteness assumption consents to neglect repulsive interactions.

Here I will try to briefly revise how microscopic models of flocking are built and motivate the introduction of generalized flocking models that include memory effects in the alignment mechanism, which represent our original motivation for the work presented in Chapter 3.

2.1 MICROSCOPIC MODELS OF FLOCKING

2.1.1 Self propelled agents

It is clear that the interest of active matter is not that of giving a detailed and realistic account of the self-propulsion mechanism which maintains the system in a non-equilibrium condition: a schematic description of it through persistent random walks is often more than enough. In spite of this simplification, the problem of modelling self-propelled agents is interesting per se, and several types of persistent random walks have been introduced over the years to model swimming protozoa, crawling cells, Brownian motors, and many other natural and artificial self-propelled particles in the category.

The distinctive trait of active particles is their ability to inject energy into the system, rather than simply dissipate it. Active Brownian Particle (ABP) models are modified equations for Brownian motion which attain an effective description for this property through the introduction of non-conservative non-dissipative forces [160]. A typical ABP model is of the form

$$\dot{\mathbf{r}} = \mathbf{v}, \qquad m\dot{\mathbf{v}} = \mathbf{F}_{n.c.} - \nabla U(\mathbf{r}) + \mathbf{f}, \tag{1}$$

with $\mathbf{f}$ a Gaussian zero-mean random process, with second moment

$$\langle f_\alpha(t) f_\beta(t') \rangle = 2D_{\alpha\beta}\delta(t - t'), \tag{2}$$

and $\mathbf{F}_{n.c.} = -\gamma(\mathbf{r}, \mathbf{v})\mathbf{v}$ a non-conservative force, acting for simplicity in the direction of motion. While for a standard Brownian particle — which can only dissipate energy through $\mathbf{F}_{n.c.}$ — we have $\gamma(\mathbf{r}, \mathbf{v}) > 0$, in order to model active systems we must take $\gamma(\mathbf{r}, \mathbf{v}) < 0$, in at least a region of the phase space. This creates an increase in the energy of the particle, rather than a dissipative compensation to the uptake of thermal energy. Hence $\mathbf{F}_{n.c.}$ is not only non-conservative, but also non-dissipative, and it models the effect of self-propulsion on the system.

A well-known parametrization of $\gamma(\mathbf{r}, \mathbf{v})$ for ABP models is the Rayleigh-Helmholtz parabolic friction coefficient [160]

$$\gamma(\mathbf{r}, \mathbf{v}) = \gamma(\mathbf{v}) = -a + b|\mathbf{v}|^2 = b(|\mathbf{v}|^2 - v_0^2) \tag{3}$$

where $v_0^2 = a/b$, for $a, b > 0$ — a negative a would correspond to a passive model. Given Eq. (3), with $U(\mathbf{r}) = 0$, the stationary probability density of $\mathbf{v}$ is the typical crater-like distribution associated to the double-well potential:

$$P_0(v) \propto \exp \frac{1}{D}\left[\frac{a}{2}|v|^2 - \frac{b}{4}|v|^4\right]. \tag{4}$$

Simplifications of the Rayleigh-Helmholtz models have also been studied in the literature, from the linear variant of Schienbein and Gruler [165], where $\gamma(\mathbf{v}) = \gamma_0(1 - v_0/|\mathbf{v}|)$, to the constant-speed limit case, where $P_0(\mathbf{v}) \propto \delta(|\mathbf{v}| - v_0)$. This last model can be obtained from the Rayleigh-Helmholtz model in the $b/D \to \infty$ limit. When $|\mathbf{v}|$ is constant, it is convenient to work in polar coordinates, where the ABP model reduces to a diffusive stochastic equation for the angular variables of the particle's velocity. In this case, $\gamma(\mathbf{r}, \mathbf{v})$ can also be viewed as a Lagrange multiplier implementing the speed constraint. Another interesting parametrization of $\gamma(\mathbf{r}, \mathbf{v})$ is that of the energy depot model [70], which allows for a mechanistic interpretation of the conversion of internally stored energy into kinetic energy.

Beyond ABPs, two other successful models for self-propelled agents are run and tumble particles (RTPs) and active Ornstein-Uhlenbeck particles (AOUPs). The dynamics of RTPs is, by definition, made up of runs at constant velocity $\mathbf{v}$, and sudden random re-orientations of the swimming direction, occurring at a rate α. The long-time, long-distance behavior of a pure running and tumbling particle is an effective random walk with diffusivity $D_0 = v^2/\alpha d$, where d is the spatial dimension and $v = |\mathbf{v}|$. The large scale dynamics of RTPs is therefore equivalent to that of ABPs with fixed speed and isotropic diffusion coefficient, at least in the absence of anisotropies or direct interactions [174]. This equivalence holds even when we assume that the microscopic parameters v, α and D (the isotropic diffusion coefficient of the ABP) depend on external variables, like the local density sensed by the particle [30]. These models have acquired a paradigmatic role, as it is now well-established that both RTPs and ABPs with a density-dependent swim speed can explain the motility-induced phase separation [174].

AOUPs provide an alternative way of modeling the self-propulsion mechanism through overdamped Brownian motion with colored Gaussian noise:

$$\dot{\mathbf{r}} = -\nabla U(\mathbf{r}) + \mathbf{f}, \qquad \tau\dot{\mathbf{f}} = -\mathbf{f} + \sqrt{2D}\boldsymbol{\xi}. \tag{5}$$

The self-propulsion force $\mathbf{f}$ follows an Ornstein-Uhlenbeck process [82], with

$$\langle f_\alpha(t) f_\beta(t')\rangle = \delta_{\alpha\beta}\frac{D}{\tau}e^{-|t-t'|/\tau}. \tag{6}$$

The parameter τ quantifies the persistence of motion in this case. Since color is not accompanied by any memory kernel in the evolution equation for $\mathbf{r}$, the process does not satisfy the second fluctuation-dissipation theorem [109] and is therefore a non-equilibrium process. This type of model has been largely studied in the literature, both analytically and numerically, achieving a quite exhaustive characterization of its fundamental statistical properties [130, 27]: the Gaussian nature of the noise makes it more analytically tractable than ABPs and especially RTPs.

Of course, more ingredients can be added to the minimal models mentioned above, like translational diffusion (through an additional white noise term in equation for the coordinate $\mathbf{r}$), or a dependency on time or on other degrees of freedom of the microscopic coefficients (this is the case of the energy depot model). However, we are interested here in the simplest among these models, as the goal is to increase their complexity by including inter-particle interactions.

2.1.2 Vicsek-like models

ABPs with alignment interactions

The class of SPP models which have been most largely used in polar active matter is that of constant speed ABPs — although empirical observations revealed that speed fluctuations are far from trivial in real animal systems [40, 10]. The hard constraint on $|\mathbf{v}|$ prevents speed fluctuations and allows us to work directly with angular variables, thus reducing the dynamics to a lower-dimensional manifold. The inter-particle interaction is expressed by means of aligning torques acting on the orientation vectors, which, in $d = 2$, are parametrized by a single angular variable: $\mathbf{e}(\theta) = (\cos\theta, \sin\theta)$. The dynamics results into a stochastic underdamped rotation of particles' orientations at an angular velocity $\dot{\theta}$, and particles' transport at constant speed v_0:

$$\dot{\mathbf{r}}_i = v_0 \mathbf{e}(\theta_i), \quad \dot{\theta}_i = -\sum_j J_{ij} \sin(\theta_i - \theta_j) + \sqrt{2D}\xi_i, \quad i = 1, \dots, N. \tag{7}$$

The J_{ij}'s are the entries of the coupling matrix, defining the strength and nature of interaction — here assumed purely ferromagnetic, with $J_{ij} \geq 0 \ \forall i, j$. D is the rotational diffusivity and $\langle \xi_i(t)\xi_j(t')\rangle = \delta_{ij}\delta(t - t')$.

A generalization of the model in $d \geq 2$ is obtained using $\gamma(\mathbf{r}, \mathbf{v})$ as a Lagrange multiplier in Eq.(1):

$$\dot{\mathbf{r}}_i = v_0 \mathbf{e}(\boldsymbol{\theta}_i), \quad \dot{\mathbf{v}}_i = \left(-\sum_j J_{ij}\mathbf{v}_j + \sqrt{2D}\boldsymbol{\xi}_i \right)_{\perp}, \quad i = 1, \dots, N. \tag{8}$$

where now $\boldsymbol{\xi}_i$ is a vector of isotropic white noises: $\langle \xi_{i,\alpha}(t)\xi_{j,\beta}(t')\rangle = \delta_{\alpha\beta}\delta_{ij} \cdot \delta(t - t')$ and $\boldsymbol{\theta}_i$ is a set of $d - 1$ angles specifying the direction of the particle i. The $\perp$ symbol denotes the projection operation onto the plane perpendicular to the velocity of each bird $\mathbf{v}_i$. The SDE (8) must be integrated following the Stratonovich convention: because of the projection operation, the process is indeed multiplicative [1].

1 Alternatively, one can transform Eq. (8) into an Itô SDE by adding a parallel drift term equal to $-(d-1)D\mathbf{v}_i$ to the second equation, which guarantees that $\frac{d|\mathbf{v}_i|^2}{dt} = 0$ when the derivative is computed using the rules of Itô calculus.

The original Vicsek model

One motivation for the choice of constant speed ABP models in polar active matter is that, historically, the first model introduced by Vicsek et al. in 1995 [192] was defined through the following set of agent-based update rules for the positions and orientations of the particles:

$$\mathbf{r}_i(t+1) = \mathbf{r}_i(t) + \mathbf{v}_i(t+1), \tag{9}$$

$$\mathbf{v}_i(t+1) = v_0 \mathcal{R}_\eta \Theta\Big[\sum_{j\in\mathcal{S}_i} \mathbf{v}_j(t)\Big], \tag{10}$$

where Θ is the normalization operator, and $\mathcal{S}_i$ is the set of neighbors of particle i by which this is influenced. The set of neighbors $\mathcal{S}_i$ includes, in the original version of the model, all the particles which are placed within a circle of radius R from the i-th one. Alternative choices for the parametrization of $\mathcal{S}_i$ are possible: we can classify them into two major categories, namely metric and topological, depending on the notion of distance (metric or metric-free) which is employed to evaluate whether pairs of birds directly interact. Finally, the noise operator $\mathcal{R}_\eta$ implements a uniformly distributed random rotation of its argument in the interval $[-2\pi\eta, 2\pi\eta]$.

Another interesting point concerns the distinction between additive and non-additive interactions. In the first case, the strength of alignment forces increases with the local particle density, while in the latter case the relative strength of alignment to noise is roughly independent of local density. The Vicsek model in Eq. (10) falls in this second class as it displays non-additive interactions. An equivalent parametrization in terms of a continuous-time model of the kind of Eq. (8) should involve a coupling matrix of the form $J_{ij} = w_i\theta(R - r_{ij})$, with $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ and $w_i \propto 1/n_i$, where $n_i = \sum_j \theta(R - r_{ij}) = |\mathcal{S}_i|$ is the total number of neighbors by which bird i is influenced. It is evident that non-additive interactions are not pairwise, since they depend on the global configuration of the system, nor symmetric, hence not satisfying an action-reaction principle.

Conversely, additive interactions are realized in Langevin-Vicsek-like models like (8) if we take $w_i = \text{const}$. In this case, the feedback between alignment and local density fluctuations is enhanced, and dense finite-size clusters tend to form. Average momentum conservation, symmetry and pairwise nature of the interaction are preserved. At the agent-based level, an equivalent way of incorporating the feedback with density fluctuations is through a so-called 'vectorial' noise [49]: instead of adding an angular random variable to the normalized average of neighbors' velocities, a vector random variable $\mathbf{f}_i$ (independent of n_i) can be added to $\sum_{j\in\mathcal{S}_i} \mathbf{v}_j(t)$ and the resulting vector can be later normalized: $\Theta\Big[\sum_{j\in\mathcal{S}_i} \mathbf{v}_j(t) + \mathbf{f}_i\Big]$.

Despite these details might look as irrelevant for the study of the emergent phenomena, the differences between the above-mentioned classes of interactions do matter and affect the behaviour of the systems at the macroscopic scale.

The order-disorder transition

The most striking feature of the Vicsek model is the spontaneous transition from a disordered isotropic phase to a polar ordered phase, where all the particles move in the same direction and true long-range order is realized, even in $d = 2$. This

transition occurs as the noise amplitude is lowered (with respect to the average strength) or the average density is increased (for a fixed interaction range) [192], and it is now clear that it is better described as a first order phase transition, where coexistence of phases is realized in a wide region of the phase diagram, rather than resorting to analogies to the XY model [176, 49]. Indeed, an equilibrium XY spin system cannot exist in the ordered phase in $d = 2$, as only quasi-long range order is established [133]. In contrast, order is truly long-ranged in the Vicsek model: thanks to the presence of activity, the symmetry broken phase is stabilized by a feedback mechanism, according to which local misalignment enhances reshuffling, which allows for a more efficient transmission of the alignment message, which in turn suppresses fluctuations [48]. As a result, orientational fluctuations are not capable of disrupting order in a polarized flock, even in $d = 2$, since they have the effect of increasing the impact of activity.

The presence or absence of the 'coexistence phase' and the way it manifests itself depends on the type of microscopic interaction. For the classical (metric non-additive) Vicsek model, where coexistence is realized, numerical studies have revealed that, as the system approaches the transition from the ordered phase, most of the particles organize into dense travelling bands which move against a dilute background in a perpendicular direction to the band itself. In very large systems, another coexistence regime — dubbed cross-sea phase — has also been observed: this is characterized by the emergence of a complex pattern looking similar to the superposition of non-interacting travelling waves at a selected angle of intersection [112, 48]. Different kinds of patterns do form when interactions are additive [51, 198]. In this case, even in large-sized systems, bands are replaced by polar clusters which do not span the linear size of the system but may nevertheless contain a fraction of particles close to 1.

A different story is for topological interactions. It has been long believed that the transition to collective motion was of second order for Vicsek-like models with an alignment mechanism based on a topological notion of distance [87]. Traditional numerical results involve models where the metric-free interaction is implemented either selecting a fixed number of nearest neighbors (called k-nearest neighbor models) or the first shell of neighbors in a Voronoi tessellation. In fact, recent numerical results on flocks of larger sizes suggest the opposite thing, i.e. that the transition is of first order, at least when the topological interaction is realized through a k-nearest neighbor Vicsek model [129]. The first order nature of the transition is observed in numerical simulations thanks to the emergence of travelling bands which denote the presence of strong density fluctuations mediating the onset of order. The fact has been formally explained in [129] with the introduction of an effective fluctuating field theory fuelled by a density-dependent multiplicative white noise. To the best of our knowledge, if the topological interaction involves Voronoi neighbors rather than the k nearest neighbors, only transition phenomena compatible with a second order description have been observed so far. Hence Voronoi Vicsek systems can be treated as critical, at least up to moderate sizes.

A glimpse of the Toner-Tu theory

For the sake of completeness, let us recall here the Toner-Tu equations, which will be useful to understand why Vicsek-like models and their large-scale predictions proved inappropriate to describe real flocks of starlings. The Toner-Tu theory involves two continuous fields, namely the density field $\rho(\mathbf{x}, t)$ and the polarization field $\mathbf{v}(\mathbf{x}, t)$, which evolve according to the following fluctuating hydrodynamic equations:

$$\partial_t \rho = -\nabla \cdot (\rho \mathbf{v}), \tag{11}$$

$$\begin{aligned} &\partial_t \mathbf{v} + \lambda(\mathbf{v} \cdot \nabla)\mathbf{v} + \lambda_2(\nabla \cdot \mathbf{v})\mathbf{v} \\ &+ \lambda_3 \nabla(|\mathbf{v}|^2) = (\alpha - \beta|\mathbf{v}|^2)\mathbf{v} - \nabla P + D_T \nabla^2 \mathbf{v} + D_B \nabla(\nabla \cdot \mathbf{v}) + D_2(\mathbf{v} \cdot \nabla)^2 \mathbf{v} + \mathbf{f}. \end{aligned} \tag{12}$$

Here $P = \sum_n \sigma_n (\rho - \rho_0)^n$ is the pressure, with ρ_0 the average density of the system, and $\mathbf{f}$ is a Gaussian white noise source, satisfying:

$$\langle f_\alpha(\mathbf{x}, t) f_\beta(\mathbf{x}', t') \rangle = \Delta \delta_{\alpha\beta} \delta(\mathbf{x} - \mathbf{x}') \delta(t - t'). \tag{13}$$

This field-theoretical description introduced by Toner and Tu in [188] helped to understand that the ferromagnetic analogue was not the right framework to interpret the phenomenon of flocking: the passage from disorder to order is not a direct transition, unlike in the theory of magnetism, but it is mediated by the density field. The wealth of phenomena arising from the Toner-Tu theory originates from the double role of $\mathbf{v}$ as the order parameter and as the velocity of mass transport in the continuity equation. The crucial coupling between ρ and $\mathbf{v}$, allows, among many other phenomena, for a wavelike propagation of perturbations in the flock, which will be discussed in the following section.

2.2 THE INERTIAL SPIN MODEL

2.2.1 Evidence of inertial dynamics

Because of its minimalist appeal and the richness of emergent phenomena, the Vicsek model (together with the corresponding continuous theory by Toner and Tu) imposed itself as the major paradigm for active matter. Despite this fact, a deeper examination of the model, in comparison with the available data, reveals that it is unable to provide a realistic account of the behavior of bird flocks on their natural time and size scales. The main evidence of this shortage comes from the experimental observation of turning groups of starlings [3, 37, 34].

Dispersion relation

Turns are an example of decision making process which occurs in the group not as a homogeneous response to an external stimulus, but thanks to bird-to-bird propagation of information. It has been observed that, irrespective of the flock size, turns

start as perturbations having a localized origin in space (Fig. 1.b). The perturbation consists of a change in the flight direction of the birds which is marked by a peak in the radial acceleration profile of each individual (Fig. 1.d). Measuring the temporal displacements of these peaks for every pair of birds, Attanasi et al. [3] built a set of mutual turning delays, which they used to rank the individuals in a flock from the starter of the turn (ranked 0) to the last one that turns (ranked $N-1$). The ranking curve in Fig. 1.a can be converted into the curve shown in Fig. 1.c, where the spatial distance traveled by the turning information, $x(t)$, is plotted as a function of the absolute turning time, t. Far from the region where border effects are relevant, $x(t)$ exhibits a linear dependency on time: the fitted slope c_2 represents the speed of propagation of the directional information in the center-of-mass reference frame. The value of the speed c_2 extrapolated from the data has been shown to be much larger than the typical velocity of birds. This result means that information transport in the flock is *not* due to the absolute motion of the constituents, whose relative positions in the center-of-mass reference frame does not change much (and in any case at a much smaller velocity than the fitted c_2).

The linear relation $x(t) = c_2 t$, combined with the independence of c_2 from the typical birds' speed, is a fact which does not find proper explanation within the Vicsek or Toner-Tu theory. First of all, numerical simulations of the Vicsek-like stochastic model in Eq. (8) cannot reproduce collective turns and radial acceleration profiles as those observed in real flocks [37]. This first argument is compelling to prove the inadequacy of the microscopic discrete model, and it holds even for small-sized flocks. Yet, it is true that this schematic description of the dynamics leaves out, by construction, most of the details which characterize a biological system. Shall we expect this discrepancy to be due to this fact? The answer depends on the predictive power we expect the model to have for the system under investigation. The presence (or absence) of propagative phenomena on intermediate scales is, at least qualitatively, something which carries information of general character about the behavior of a fluid or the properties of a condensed matter system; by analogy, even in active models, transport phenomena should not depend on the little details which are left out (cfr. [190]).

The hydrodynamic theory by Toner and Tu offers us the possibility to work directly at the macroscopic scale. The model in Eqs. (11)–(12) can indeed be thought of as the coarse-grained description of several agent-based models whose common trait is the pairwise alignment of the particles' velocity vectors [150, 7, 8, 151, 100]. From the study of the power spectra of the density and polarity fluctuations in the symmetry-broken phase of the Toner-Tu theory, it can be observed that the dispersion relation of sound modes always has a nonzero real part [2] [187]:

$$\omega_{\pm}(\mathbf{q}) = c_{\pm}(\theta_{\mathbf{q}})q - iq_{\perp}^{z} f_{\pm}\left(\frac{q_{\parallel} l_0}{(q_{\perp} l_0)^{\zeta}}\right), \tag{14}$$

2 The fact that $\Re\,\omega_{\pm} \neq 0$ (for $q \neq 0$) occurs even at the level of the linearized theory (around the polar homogeneous solution). The perturbative renormalization group analysis of [187] impacts only the imaginary part, yielding the critical exponents z (ruling the critical dynamics of collective modes) and ζ (anisotropy critical exponent).

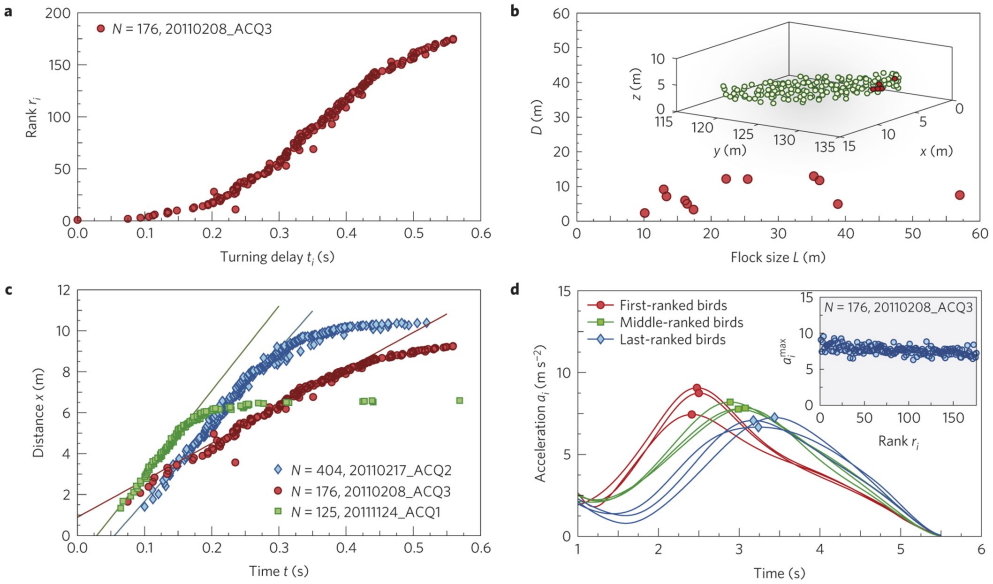

Figure 1: Figure and caption from [3]. **a**: The rank r of each bird, i.e. its order in the turning sequence, is plotted versus its absolute turning delay t, i.e. the delay with respect to the first bird to turn **b**: The maximum mutual distance D between the top 5 birds in the rank does not increase with the linear size of the flock, L, indicating that the first birds to turn are actually close to each other in space. The result does not change if we use a different number of top birds, as long as this number is much smaller than the flock's size. Inset: the actual position of the top 5 birds (red) within a real flock. **c**: The distance x travelled by the information in a time t is proportional to the radius of the sphere containing the first $r(t)$ birds in the rank, namely $x(t) = [r(t)/\rho]^{1/3}$. The speed of propagation, c_s, is the slope of the linear regime of $x(t)$ for early and intermediate times (solid lines are linear fits). **d**: The intensity of the peak of the radial acceleration, $a^{\max}$, (filled symbols) decreases very weakly in passing from the first to the last turning birds. Inset: plot of $a_i^{\max}$ versus the rank r_i for each bird. Hence, information propagates through the flock with negligible attenuation.

where $c_\pm(\mathbf{q}) = \frac{v_0(1+\lambda)}{2}\cos(\theta_\mathbf{q}) \pm \sqrt{\frac{v_0^2(1-\lambda)^2}{4}\cos^2(\theta_\mathbf{q}) + \sigma_1\rho_0\sin^2(\theta_\mathbf{q})}$. The parameter $\rho_0 = \langle\rho(\mathbf{x},t)\rangle$ is the average density; $v_0 = |\langle\mathbf{v}(\mathbf{x},t)\rangle|$ is not the average velocity of the whole flock $\frac{1}{N}|\sum_{i=1}^N \mathbf{v}_i|$, but a parameter which gives us an indication of the symmetry breaking [187]; l_0 is a characteristic length of the system, needed to have a dimensionless argument for the scaling function $f_\pm(x)$. Finally, $\theta_\mathbf{q}$ is the angle formed by the wavevector $\mathbf{q}$ with the collective direction of motion, and the symbols $\perp$ and $\parallel$ indicate its transverse and longitudinal components with respect to the same reference direction.

From an inspection of Eq. (14) we can detect the following elements:

- Attenuation of the wavelike perturbation. The Toner-Tu dispersion relation allows for wave propagation, but, since the imaginary part is nonzero, it is also accompanied by anisotropic damping. At long wavelengths ($q \to 0$), propagation dominates and becomes almost isotropic. However, localized perturbations like those observed in real flocks are better described by wavepackets which also include larger q modes. These modes get more fastly damped,

according to Eq. (14). In the limit $q \to \infty$, the scaling function scales as $f_{\pm}(x) \sim x^{z/\zeta}$, and the high-wavenumber turning waves get mostly damped in the longitudinal direction ($z/\zeta > 1$). In contrast, a very small attenuation was indifferently observed in the radial acceleration profiles of real flocks (see Fig. 1.b)

- Dependency on v_0. In the Toner-Tu theory the presence of a nonzero $\Re\omega_{\pm}$ is due to the coupling between density and velocity fields — with the latter playing the role of local order parameter and transport velocity at the same time — and to the lack of Galileian invariance. Precisely, the pressure force term $\nabla P(\rho)$, in combination with mass conservation, is responsible of transverse propagation (in the reference frame of the fluid), while the convective term $\lambda(\mathbf{v}\cdot\nabla)\mathbf{v}$ is affecting longitudinal propagation. The sum of the two terms can be split in the following way:

$$T_1 + T_2 = [\nabla P + (\mathbf{v}\cdot\nabla)\mathbf{v}] \; + \; [(\lambda - 1)(\mathbf{v}\cdot\nabla)\mathbf{v}]\,. \tag{15}$$

  This rewriting highlights the density-velocity coupling and the advective role of $\mathbf{v}$, on one hand (T_1), and the lack of Galileian invariance, on the other hand (T_2). The term T_1 is also present in standard fluids, where $\lambda = 1$; its contribution to $\Re\omega_{\pm}$ just amounts to the transverse propagation of compression waves and the trivial transport of perturbations due to the motion of the fluid itself. The T_2 term, on the contrary, is peculiar of active fluids and is the candidate term to explain the experimental observations within the framework of the Toner-Tu theory. Let us recall the dispersion relation (14): here $c_{\pm}$ refers to an external reference frame, while the experimental c_2 is measured in the flock's frame. In order to compare the two, we should then subtract from $c_{\pm}$ the mean cruising velocity $v_0\cos(\theta_{\mathbf{q}})$ along the considered direction. Let us focus on the longitudinal one, $\theta_{\mathbf{q}} = 0$: in this case the speed of density waves is $c_+(0) = v_0$, and the speed of orientation waves is $c_-(0) = \lambda v_0$. The two coincide in a standard, Galileian invariant, fluid, and both become zero in the reference frame of the moving fluid. In an active fluid, the speed at which the turning perturbation travels depends on how 'strongly' the Galileian invariance is broken, i.e. on how efficient reshuffling of bird's relative positions is, at the microscopic level. In real flocks the relative positions of birds do not change much, even during turns, and the reconstructed adjacency matrix is, to an excellent degree of approximation, constant [1, 136]. The measured velocity c_2 should therefore be very small, in contrast to experimental values of 20-40 $\mathrm{m}\cdot\mathrm{s}^{-1}$ (starlings' typical speeds are 7-12 $\mathrm{m}\cdot\mathrm{s}^{-1}$).

To sum up, experiments provide evidence that the propagation of information in a turning flock mustn't be explained through Eq. (14): an alternative model is needed.

Collective Turning

How should this alternative model be built? A hint came from another experimental evidence, namely the observation that birds did turn on intersecting equal-radius paths rather than on parallel paths [3, 1, 34].

There are two ways in which a many-particle system can perform a rotation: as a rigid body, where all the material points in the assembly rotate with the same angular velocity with respect to an external fixed point, or along paths with equal radius of curvature (see Fig. 2.a for a schematic representation). In a rigid turn, it is possible to identify a co-rotating frame where the relative positions of the particles are unaltered; to an external observer, this appears as a parallel-path turn. Moreover, since the angular velocity is the same for all the particles in the system, their speed must be different, and increase with the distance from the pole. Both of these facts are not observed in real flocks of starlings, whose relative positions do not change in the observer's reference frame (Fig.2.b) and all the trajectories have approximately the same radius of curvature, as computed from the radial acceleration profiles (in Fig.1.b), allowing all the birds to fly without sharp speedups.

The equal-radius type of rotation cannot be attained using the standard generator of spatial rotations, i.e. the angular momentum, which would generate a parallel-path turn. An equivalent generator acting on the internal space of velocities, $\mathbf{v}_i$, while keeping the particles' positions unaltered, is required. This quantity, called *spin* in [3] and [37], is the generator of the rotational symmetry characterizing the classical XY or Heisenberg systems (see Fig. 2.c). In analogy to the definition of angular momentum in the external space of positions,

$$\mathbf{l} = \mathbf{r} \times \mathbf{p}, \quad \text{with} \quad \mathbf{p} = m\dot{\mathbf{r}}, \tag{16}$$

the *spin* $\mathbf{s}$ is defined as:

$$\mathbf{s} = \mathbf{v} \times \mathbf{p}_\mathbf{v}, \quad \text{with} \quad \mathbf{p}_\mathbf{v} = \chi\dot{\mathbf{v}}, \ |\mathbf{v}| = 1. \tag{17}$$

The definition of the new momentum $\mathbf{p}_\mathbf{v}$ in the internal space of velocities brings in the definition of the analogous of the inertial mass m, i.e. a generalized inertia χ. The generalized inertia χ is unrelated to the mass of the bird, but it rather accounts for the resistance opposed by the particles to instantaneous changes of their orientations.

2.2.2 Model description

The experimental evidence described above has been rationalized with the introduction of a new dynamical model, called *inertial spin model* (ISM). It consists of a minimal modification of the continuous-time Vicsek model in Eq. (8), obtained through the introduction of the above-mentioned generalized inertia χ, which transforms the stochastic process from first order (in the variables $\mathbf{v}_i$) to second order.

The temporal evolution of the system in $d = 3$ is given by the following set of Langevin equations:

$$\dot{\mathbf{r}}_i = \mathbf{v}_i; \tag{18}$$

$$\dot{\mathbf{v}}_i = -\frac{1}{\chi}\mathbf{v}_i \times \mathbf{s}_i; \tag{19}$$

$$\dot{\mathbf{s}}_i = -\frac{\eta}{\chi}\mathbf{s}_i + \mathbf{v}_i \times \frac{J}{v_0^2}\sum_j n_{ij}(\{\mathbf{r}_i\})\mathbf{v}_j + \sqrt{2D}\boldsymbol{\xi}_{i\perp}. \tag{20}$$

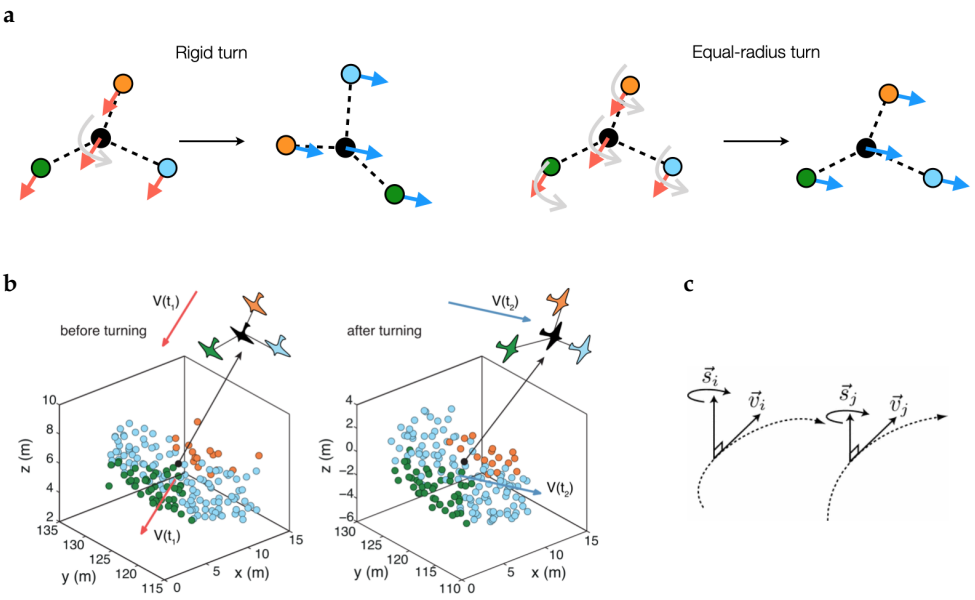

Figure 2: **a**: On the first row, a schematic representation of parallel path versus equal radius turns is shown. **b**: Reorientation of the flock during the turn. The figure shows the change in the relative positions of birds in the reference frame of the turning flock. The reference bird in the center is pictured in black, birds flying in front of it at the start of the turn are green, those on the sides are light blue, the ones behind are orange. After the turn, the color code does not correspond to the same relative positions. Green and orange birds are now in the sides, blue birds are at the head and the back of the flock. Figure adapted from [1]. **c**: Schematic representation of velocity, spin, and trajectory for two particles, from [37].

Here the $\boldsymbol{\xi}_i$'s are independent isotropic white noises $\langle \xi_{i\alpha}(t)\xi_{i\beta}(t')\rangle = \delta_{ij}\delta_{\alpha\beta}\cdot\delta(t-t')$ and $\perp$ denotes again the perpendicular projection with respect to $\mathbf{v}_i$, for each $i = 1,\ldots,N$.

We focus on the dynamics of the orientational degrees of freedom, Eqs. (19)–(20), neglecting for a while the coupling to the positions $\mathbf{r}_i$. These equations can be read as the stochastic analogue of Newton's law for the rotational dynamics in the internal space of velocities. According to Newton's law, the angular acceleration of each particle in this internal space is proportional to the torque exerted on the particle through the generalized inertia χ. The torque is determined in this case by the contribution of a social force, which aligns the orientation of the bird to the surrounding ones, plus those of a linear damping force and of a gaussian noise[3].

Exploiting the fact that $|\mathbf{v}_i| = v_0\ \forall i$, we can rewrite Eqs. (19)–(20) as a second order SDE:

$$\chi\ddot{\mathbf{v}}_i = \left(-\eta\dot{\mathbf{v}}_i + J\sum_j n_{ij}\mathbf{v}_j + v_0\sqrt{2D}\boldsymbol{\xi}_i\right)_\perp - \chi\left|\frac{\dot{\mathbf{v}}_i}{v_0}\right|^2\mathbf{v}_i\,, \qquad i = 1,\ldots,N. \tag{21}$$

Eq. (21) can also be derived from first principles, starting from the conservative dynamics of the interacting system, immersing it into an effective heat bath which

3 The random torque and the damping torque must satisfy a fluctuation-dissipation relation if we want to recover an equilibrium limit that reproduces the Heisenberg or XY ferromagnet for vanishing motility ($v_0 \to 0$).

introduces a viscous drag and a fluctuating force source, and enforcing the hard constraint $|\mathbf{v}_i| = v_0 \ \forall i$. The perpendicular projection with respect to the direction of motion of the i-th bird and the centripetal acceleration term keep the particles' speed fixed to a given value. Again, the multiplicative process must be integrated applying the Stratonovich convention. We can go from Eq. (21) to Eqs. (19)–(20) using the definition of the spin (17) — see [6] for a more detailed discussion.

By finally coupling the dynamics in the external space of positions to the dynamics of the orientational degrees of freedom, the model becomes active. The same parametrizations of the adjacency matrix n_{ij} that can be adopted in the overdamped case can also be adopted here. The inertial spin model represents indeed just a Newtonian generalization of the continuous-time Vicsek model, which is recovered in the overdamped limit $\chi/\eta^2 \to 0$. In the absence of interactions, the difference between the two models would be purely dynamical, and only visible at intermediate time and size scales. How inertia mingles with non-equilibrium interactions and to what extent this modifies the non-equilibrium steady state or the critical dynamics of the system is less clear, but the topic has attracted considerable interest recently [141, 36, 38, 96].

In any case, we can conclude that a second order dynamics is associated to the presence of two time scales for each mode $\mathbf{q}$, even at equilibrium, in the absence of any coupling with the density field. The relative magnitude of these two time scales is what mainly determines how a mode is propagated or dissipated and, at a global level, the damping regime in which the system operates. For a more exhaustive discussion of this issue, we refer to Refs. [37, 34]. A detailed derivation of the equations of motion of the ISM can be found in [3, 34].

2.2.3 The spin wave approximation

To study the model in the ordered phase, it is useful to consider a linearization of the equations of motion (19)–(20) around the perfectly ordered solution $\mathbf{v}_i = v_0\mathbf{n}$ $\forall i$, where $\mathbf{n}$ is the collective direction of motion of the flock. This approximation is called, in analogy to ferromagnets [69], spin wave approximation (SWA). We will use it in the following for the analysis of the experimental data.

In order to derive it, it is sufficient to decompose each vector $\mathbf{v}_i$ into its longitudinal and transverse components with respect to $\mathbf{n}$:

$$\frac{\mathbf{v}_i}{v_0} = v_i^L\mathbf{n} + \boldsymbol{\pi}_i \quad \text{where} \quad v_i^L = \sqrt{1 - |\boldsymbol{\pi}_i|^2} \simeq 1 - \frac{1}{2}|\boldsymbol{\pi}_i|^2. \tag{22}$$

We assume here, without lack of generality, that $|\mathbf{v}_i| = v_0 = 1$. Thanks to this assumption, the perpendicular projection of any vector $\mathbf{a}$ in $d = 3$ can be obtained as:

$$\mathbf{a}_\perp = -\mathbf{v}_i \times (\mathbf{v}_i \times \mathbf{a}). \tag{23}$$

Using Eqs. (22)–(23), one can evaluate all the terms appearing in Eq. (21), at the desired order in $|\boldsymbol{\pi}_i|$.

Let us focus on time derivatives: we notice that, in addition to $\dot{\boldsymbol{\pi}}$ and $\ddot{\boldsymbol{\pi}}$, there must also have terms containing $\dot{\mathbf{n}}$ and $\ddot{\mathbf{n}}$. We assume however that the direction

of collective motion is constant: $\dot{\mathbf{n}} = \ddot{\mathbf{n}} = 0$. This is legitimate in the limit $N \to \infty$, when the wandering of the order parameter is suppressed. If, on the contrary, one wants to take the effect of the rotation of $\mathbf{n}$ into account, apparent forces emerge because the chosen reference frame is non-inertial. Neglecting apparent forces enables us to segregate *on-plane* (i.e. perpendicular to $\mathbf{n}$) and *off-plane* (i.e. parallel to $\mathbf{n}$) contributions, and completely disentangle the corresponding equations. One can then consider the equations in the $\boldsymbol{\pi}$-plane only:

$$\frac{d^2\boldsymbol{\pi}_i}{dt^2} + \eta\frac{d\boldsymbol{\pi}_i}{dt} + J\Lambda_{ij}\boldsymbol{\pi}_j = \hat{P}\boldsymbol{\xi}_{i\perp} + O(|\boldsymbol{\pi}|^3), \tag{24}$$

where $\Lambda_{ij} = n_c\delta_{ij} - n_{ij}$ and $\hat{P}$ is the projection operator onto the plane perpendicular to the collective velocity. The velocity fluctuations $\boldsymbol{\pi}_i$ play in this case the same role as spin excitations in Dyson's SWA, since they become the new degrees of freedom and are subject to a linear interaction.

At this stage, what remains to explicitly evaluate is $\hat{P}\boldsymbol{\xi}_{i\perp}$. Since $\boldsymbol{\xi}_{i\perp}$ lives in the plane perpendicular to $\mathbf{v}_i$, the perpendicular component to the plane spanned by $\mathbf{n}$ and $\mathbf{v}_i$ is left unchanged by the projection operator, while the other one is contracted with a factor $\cos\theta_i = \mathbf{v}_i \cdot \mathbf{n}$. As a result:

$$\langle \hat{P}\boldsymbol{\xi}_i(t) \cdot \hat{P}\boldsymbol{\xi}_i(s)\rangle = 2(1+\cos^2\theta_i)\frac{T\eta}{\chi^2}\delta(t-s). \tag{25}$$

The second moment of each noise term is then rescaled, with respect to the original one, by a factor:

$$\frac{1}{2}(1+\cos^2\theta_i) = \frac{1}{2}\left(1+\left(v_i^L\right)^2\right) = 1 - \frac{1}{2}|\boldsymbol{\pi}_i|^2 \simeq v_i^L. \tag{26}$$

We can reabsorb this rescaling factor into the temperature parameter T/χ, defining a new spin wave temperature[4]

$$\tilde{T}_{SWA}/\chi = \Phi T/\chi, \tag{27}$$

where $\Phi = |\frac{1}{N}\sum_{i=1}^N \mathbf{v}_i|$ is the polarization of the flock. In the low temperature case, where $|\boldsymbol{\pi}| \ll 1$, the correction to the temperature parameter in Eq.(27) is of a lower order than the terms which have been neglected in the deterministic part of Eq. (24). Hence it is consistent to include this first correction to the spin-wave temperature through this simple effective rescaling. The resulting equations are then

$$\frac{d^2\boldsymbol{\pi}_i}{dt^2} + \eta\frac{d\boldsymbol{\pi}_i}{dt} + J\Lambda_{ij}\boldsymbol{\pi}_j = \tilde{\boldsymbol{\xi}}_i, \tag{28}$$

where $\tilde{\boldsymbol{\xi}}_i = \hat{P}\boldsymbol{\xi}_{i\perp}$ are skew independent two-dimensional white noises, with effective amplitude $2\eta\tilde{T}_{SWA}/\chi$ for each coordinate.

The dispersion relation

Since the main motivation for the introduction of the ISM comes from the observation that the turning disturbance cannot be propagated in flocks following the dispersion relation of the Toner-Tu theory, it is worth studying the dispersion relation

4 In principle, a different temperature for each bird could be defined, using the bird's longitudinal velocity v_i^L, according to Eq. (26). However, it is convenient to work with a single average temperature $\tilde{T}_{SWA}/\chi$ to let the model keep a formal equilibrium structure (when not coupled to the position dynamics).

associated to the ISM in its ordered phase. This was the object of [44], considering an on-lattice variant of the ISM, and of [42, 195], where continuous models were considered.

The on-lattice variant is particularly simple, since there is no advection nor reshuffling. In this case the density field is homogenoeus and non-evolving, and the dynamics of the polarity excitations only results from the Landau-Ginzburg coarse-graining of the equilibrium linear model in Eq. (29). This is a plain Gaussian theory, whose propagator is well-known. When particles lie on the sites of a fixed network, the model can be analytically solved even at the microscopic level. The linearized dynamics is easily diagonalized in the basis of eigenvectors of the discrete Laplacian Λ_{ij}[5]:

$$\frac{d^2\hat{\boldsymbol{\pi}}_k}{dt^2} + \eta\frac{d\hat{\boldsymbol{\pi}}_k}{dt} + J\lambda_k\hat{\boldsymbol{\pi}}_k = \tilde{\boldsymbol{\xi}}_k, \tag{29}$$

whose dispersion relation reads $\omega_k = i\gamma \pm \sqrt{-J\lambda_k - \gamma^2/4}$. In a cubic lattice with periodic boundary conditions and lattice spacing a, $\lambda_{k(n,m)} = -n_c\frac{4\pi^2}{N}(n^2+m^2)$, with $n, m = 1, \ldots, L/a$, where L is the linear size of the system, and L/a is the number of sites per side.

To reintroduce self-propulsion, it is necessary to restart from Eqs. (18)–(20) and derive the hydrodynamic theory associated to them. In addition to the density and velocity fields, the local spinning velocity of the active fluid — corresponding to $\mathbf{s}$ in the microscopic model — must be treated as a slow-varying field. Full continuous equations are derived in [195] following the Smoluchowski approach. In contrast, in [42] non-linearities are kept to a minimum, and Galileian invariance is assumed to hold.

The dispersion relation resulting from this simpler phenomenological model is analytically tractable and firstly revealed the existence of a different kind of propagating spin waves. The key idea emerging from [42] is that $\Re\omega(k)$ receives contribution from both first sound (compression waves due to density fluctuations, which propagate in the transverse direction) and second sound (propagating waves which would pass the turning information throughout the flock even in the absence of density-velocity coupling). This idea was confirmed by the more accurate analysis in [195].

It is interesting to notice that first and second sound have opposite properties. First sound is damped for large wave-numbers k, whereas it propagates, in a strongly anisotropic way, close to the hydrodynamic limit $k \to 0$. Conversely, second sound modes are damped at small k, whereas they propagate at large k, with a speed which is isotropic in the reference frame of the moving flock. At intermediate wavelengths and in a Galileian invariant system, wavelike perturbations either propagate as a combination of first and second sound modes, or do not propagate at all, depending on the ratio between second and first sound speeds [42].

5 Notice that $\Lambda_{ij} = n_c\delta ij - n_{ij}$, as defined above, can be identified with the discrete Laplacian only if n_{ij} is symmetric. The origin of the name is clear if we consider the coarse-grained approximation of a regular cubic lattice in d dimensions, with a the lattice spacing and $n_c = 2d$ nearest neighbors per site. After the coarse-graining, $\Lambda_{ij} \to n_c a^2\nabla^2$. Generally speaking, the $\mathbf{\Lambda}$ matrix represents the extension of the concept of Laplacian for a regular graph, where each vertex has the same degree n_c [12].

When reshuffling is reinstated in the flock and Galileian invariance is broken, the dispersion relation becomes too complicated to be studied in the whole $\mathbf{q}$ space, and the existence of propagating modes, in any working regime, cannot be ruled out. However, as we have said before, there is convincing evidence that the contribution coming from the breakdown of Galileian invariance alone could not explain experimental observations: second sound seems to be the crucial element [34].

3 STATISTICAL INFERENCE FOR THE DYNAMICS OF FLOCKS

In this Chapter I will present a novel maximum likelihood inference method for second order stochastic differential equations, which has been developed for the study of birds' dynamics. The presentation is preceded by an introduction about parametric inference methods for stochastic dynamical problems, and by a review of previous results obtained from the analysis of experimental data of natural flocks of European starlings.

3.1 PARAMETRIC INFERENCE FOR STATIC AND DYNAMIC PROBLEMS

3.1.1 Non-parametric vs Parametric inference

The goal of statistical inference is to optimally reconstruct from empirical data-sets a parameter-free model that could have generated them. In order to tackle the problem, several approaches are available, which can be distinguished into two main classes, namely *parametric* and *non-parametric* inference methods. In parametric approaches the goal is just to infer, based on observations, the parameters of a given model. In contrast, in the non-parametric case, the observed data — whether stationary samples or time series — are not asked to follow any specified distribution.

Non-parametric methods therefore employ standard statistical tools or general analytical techniques which do not require prior information on the underlying model. In this context, the simplest and most common idea is to use density estimators:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{30}$$

where $f(x)$ is the probability density function we want to reconstruct, n is the number of data points, $\{X_i\}$ is the set of measured values, h is a smoothing parameter, and $K(x)$ is a kernel function that must satisfy the condition $\int dx K(x) = 1$. For a standard histogram, h is the bin size and $K(x) = \mathbb{1}_{[-1/2,1/2)}(x)$. Alternative kernels can be used, for instance to give a larger weight to the focal point X_i.

The method of density estimators naturally applies to static problems, where the data-set consists of multiple observations of a variable — or a set of variables — describing the stationary state of a system. No equilibrium assumption is implied: $f(x)$ can represent either an equilibrium ensemble or a non-equilibrium steady state. However, the simple method described above requires the sampling not to be too sparse in the phase space, a condition which makes the reconstruction feasible only for low-dimensional systems. Even when each coordinate of the state variable is schematized as a discrete random variable which can take only m different values,

we need a number of observations $M \gg m^D$, where D indicates the phase space dimension. Hence the lower bound on the sample size can grow very fast, no matter how small m is (a recurrent example is the inverse Ising problem [138], with $m = 2$ and $D = N$ is typically very large). The situation rapidly gets out of control when m is large, as it is the case when the coordinates of the state variable are continuous and we try to approximate their distribution.

In summary, non-parametric inference is exposed to the curse of dimensionality. A first way to escape it is to adopt parametric or semi-parametric inference approaches: in the following, we will focus on them. An alternative strategy is to adopt effective low-dimensional descriptions of the complex system of interest. Many complex systems of diverse nature indeed exhibit a separation of scales, allowing for the identification of slow and fast variables and for the introduction of reduced Markov models, where the effect of fast or integrated variables is reabsorbed into noise terms. For a review of inference approaches designed for this kind of dynamic systems, we refer to [80].

3.1.2 The Bayesian setting

In this section, we will put our attention on parametric inference. The fundamental ingredients we have in this setting are:

- a set of experimental observations $\bar{\mathbf{X}} = (\mathbf{X}^1, \mathbf{X}^2 \dots \mathbf{X}^M)$ of the observable $\mathbf{X}$, which can have a large but finite number of scalar components, corresponding to an M-uple of probabilistic events;

- a parameter vector $\boldsymbol{\theta}$, treated as a random variable, with an associated p.d.f. $P(\boldsymbol{\theta})$.

This framework is known as *inverse problem*. In contrast to the *direct problem* of Statistical Physics, which consists in deducing predictions about observable quantities from a known model that describes the system, the inverse problem takes as a starting point empirical observations and aims at extracting a parameter-free model from them.

The inference problem is conveniently cast in a probabilistic framework: with this, we do not refer to any statistical mechanical description of the system under study, but to the fact that a *degree of belief* — in quantitative terms, a probability distribution — is associated to the result of the inference procedure itself. The first advantage of parametric methods over non-parametric ones is therefore the possibility to compare different generative models and to assess how probable a certain description of the data is. In this way, the parametric Bayesian setting allows us to accept or reject a model at some 'significance level' [122]. The disadvantage is that assumptions on the structure of the generative model are inevitably needed. We denote by I (understood in the following) the amount of prior information carried by these assumptions: a model is therefore specified by the pair $(\boldsymbol{\theta}, I)$.

Given I, the inference problem is fully solved when the posterior distribution of model parameters is found. Using Bayes' theorem,

$$P(\boldsymbol{\theta}|\bar{\mathbf{X}}, I) = \frac{P(\bar{\mathbf{X}}|\boldsymbol{\theta}, I)\rho(\boldsymbol{\theta}, I)}{P(\bar{\mathbf{X}}, I)} = \frac{P(\bar{\mathbf{X}}|\boldsymbol{\theta}, I)\rho(\boldsymbol{\theta}, I)}{\sum_{\boldsymbol{\theta}} P(\bar{\mathbf{X}}|\boldsymbol{\theta}, I)\rho(\boldsymbol{\theta}, I)}, \tag{31}$$

the task boils down to assigning a prior distribution to the parameters $\rho(\boldsymbol{\theta}, I)$ and computing the likelihood $P(\bar{\mathbf{X}}|\boldsymbol{\theta}, I)$. Knowing $P(\boldsymbol{\theta}|\bar{\mathbf{X}}, I)$, we can associate a probability 'score' to any set of parameter values, given the data. The best output is then identified by the maximum point of the posterior (maximum a posteriori estimator):

$$\hat{\boldsymbol{\theta}}_{MAP} = \arg\max_{\boldsymbol{\theta}} P(\boldsymbol{\theta}|\bar{\mathbf{X}}, I). \tag{32}$$

Even if dropped in our notation, it is evident that $\hat{\boldsymbol{\theta}}_{MAP}$ is in fact a function of the data $\bar{\mathbf{X}}$ and of I.

Maximum likelihood

Usually the probability distribution $\rho(\boldsymbol{\theta})$ is not known and it is rather hard, if not impossible, to acquire it. One general rule consists in taking a uniform prior[1]. With this choice, the maximization of the posterior coincides with the maximization of the likelihood, and the maximum a posteriori estimator $\hat{\boldsymbol{\theta}}_{MAP}$ is replaced by the so-called maximum likelihood (ML) estimator:

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} P(\boldsymbol{X}|\boldsymbol{\theta}) = \arg\min_{\boldsymbol{\theta}} -\log P(\mathbf{X}|\boldsymbol{\theta}). \tag{33}$$

An important property of the ML estimator is that in the limit of infinitely numerous samples, it is *consistent*, i.e. it converges in probability to the value of $\boldsymbol{\theta}$ being estimated. Precisely, for any $\epsilon > 0$, consistency requires that:

$$\lim_{M\to\infty} \Pr(|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}| > \epsilon) = 0. \tag{34}$$

Notice that far from the limit, for finite-size data samples, the estimator can be biased.

Another legitimate question is whether the solution of the problem (33) is unique. A sufficient condition is the convexity of the minus-log-likelihood (as a function of the parameters), but this is in general not guaranteed. Its convexity may depend on the features of the problem, as well as on the quality of the data, which enter as given parameters in this inverse problem setting. At equilibrium, however, uniqueness of the solution is typically ensured [138]. The state of a physical system in thermal equilibrium is described by the Boltzmann ensemble:

$$P(\mathbf{z}|\boldsymbol{\theta}) = Z^{-1}(\boldsymbol{\theta})e^{-H(\mathbf{z};\boldsymbol{\theta})}, \tag{35}$$

with $Z(\boldsymbol{\theta}) = \sum_{\mathbf{z}} e^{-H(\mathbf{z};\boldsymbol{\theta})}$ (assuming a discrete phase space), and $H(\mathbf{z};\boldsymbol{\theta})$ an effective Hamiltonian, already multiplied by the inverse temperature β. The likelihood associated to the Boltzmann distribution reads:

$$P(\bar{\mathbf{X}}|\boldsymbol{\theta}) = \prod_{k=1}^{M}\sum_{\mathbf{z}} \delta(\mathbf{X}(\mathbf{z}) - \mathbf{X}^k)\frac{e^{-H(\mathbf{z};\boldsymbol{\theta})}}{Z(\boldsymbol{\theta})} = \prod_{k=1}^{M}\frac{1}{\tilde{Z}(\boldsymbol{\theta})}e^{-\tilde{H}(\mathbf{X}^k;\boldsymbol{\theta})}, \tag{36}$$

1 The uniform prior can properly be defined only if we assume that parameters are bounded. Alternatively, suitable decaying priors might be defined, that would make regularization terms appear in Eq.(33)

where we recall that the k label indexes repeated observations of a set of observables $\mathbf{X}(\mathbf{z})$. The final equivalence in Eq. (36) is for now just formal, and serves us to write $P(\bar{\mathbf{X}}|\boldsymbol{\theta})$ in the Gibbs-Boltzmann style.

Let us introduce the reduced minus log-likelihood:

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{M}\log P(\bar{\mathbf{X}}|\boldsymbol{\theta}) = \log Z(\boldsymbol{\theta}) - \langle \tilde{H}(\mathbf{X};\boldsymbol{\theta})\rangle_{\text{exp}} \tag{37}$$

where $\langle f(\mathbf{X})\rangle_{\text{exp}} = \frac{1}{M}\sum_{k=1}^{M} f(\mathbf{X}_k)$ denotes the experimental average over M independent measurements. Now we further assume that the effective Hamiltonian $H(\mathbf{z};\boldsymbol{\theta})$ depends linearly on the model parameters:

$$H(\mathbf{z};\boldsymbol{\theta}) = \sum_{\mu=0}^{K} \theta_\mu X_\mu(\mathbf{z}), \tag{38}$$

where X_μ are the scalar components of the measured observable. This kind of parametrization is recurrent in statistical inference, as it corresponds to the only family of probability distributions admitting a sufficient statistic with a number of scalar components which does not grow with the sample size (Pitman-Koopman theorem [152, 108]). The Hessian matrix of (37) is easily computed and reads:

$$\mathcal{H}_{\mu\nu} = \frac{\partial^2 \mathcal{L}}{\partial\theta_\mu \partial\theta_\nu} = \langle X_\mu X_\nu\rangle - \langle X_\mu\rangle\langle X_\nu\rangle. \tag{39}$$

Because fluctuations are non-negative, the Hessian matrix element (39) defines a positive semi-definite bilinear form $\sum_{\mu\nu} q_\mu \mathcal{H}_{\mu\nu} q_\nu \geq 0$, for any non-null vector $\mathbf{q}$. If none of the K observables X_μ appearing in the effective Hamiltonian has exactly vanishing connected correlations, all the eigenvalues of the Hessian matrix are positive and the minus-log-likelihood is strictly convex. Thus the concavity of the likelihood is basically always guaranteed when it belongs to the canonical exponential family and the associated Hamiltonian is linear in the model parameters.

These facts remain true when the static equilibrium setting is abandoned in favor of a dynamical one, provided that the probability distribution of the observed trajectory keeps a similar structure to the one we discussed above. An approach that realizes this condition is the maximum caliber approach, which is better described in Section 3.1.4.

3.1.3 The Maximum Entropy method

The maximum entropy method is a powerful inference tool which combines information theory and variational calculus. It was introduced by E. T. Jaynes in two celebrated papers [101, 102], where he rephrased the traditional ensemble derivation of equilibrium Statistical Mechanics in terms of a constrained maximization task. Despite the controversial reception of Jaynes' much discussed claims about the foundations of Statistical Mechanics, the reversal of the usual line of reasoning that he suggested turned out to be particularly well-suited for inverse problems. Over the past 50 years, the maximum entropy principle has been used in many diverse contexts: spectral analysis, image restoration, neural networks, signal processing, and other vast applications in all areas of physics — including biophysics.

In a nutshell, the maximum entropy method offers a prescription to find the 'least structured' model that is consistent with experimental measurements. The inference process must indeed be viewed as a process to extract information from the available data [194]. These data, however, are partial. As a guiding principle, we want the inferred model to be maximally committed to the observed data and maximally non-committed to the data which are not available. In this sense, the model to seek is the most unbiased, or least structured.

The notion of 'least structured' is mathematically formulated as 'maximizing the uncertainty' associated to the statistical model of interest. The key concept of *uncertainty* of a probability distribution was introduced by Shannon in 1948 [171]: in this seminal work it was proved that there exists a unique well-defined measure $H(p_1, \dots, p_n)$ to quantify the uncertainty of any information source, represented by a discrete probability distribution $\{p_i, i = 1, \dots, n\}$. This quantity, better known as Shannon entropy, is:

$$H(p_1 \dots p_n) = -k \sum_i p_i \log p_i, \tag{40}$$

with k an arbitrary constant, usually set equal to 1.

Commitment to the available data is expressed in the form of constraints[2]:

$$0 = g_\mu(p_1, \dots, p_n) = \langle X_\mu(\mathbf{z}) \rangle - \langle X_\mu(\mathbf{z}) \rangle_{\text{exp}}, \quad \mu = 0, \dots, K, \tag{41}$$

so that the problem reduces to the optimization of the Lagrangian function

$$\tilde{S}(p_1, \dots, p_n) = H(p_1, \dots, p_n) - \sum_{\mu=0}^{K} \theta_\mu \left[\langle X_\mu(\mathbf{z}) \rangle - \langle X_\mu(\mathbf{z}) \rangle_{\text{exp}} \right], \tag{42}$$

where $\langle X_\mu(\mathbf{z}) \rangle_{\text{exp}}$ is the experimental average, and $\langle X_\mu(\mathbf{z}) \rangle = \sum_i p_i X_\mu(\mathbf{z}_i)$ is the ensemble average. We denote by $\mathbf{z}$ the state variable of the system, and by $\mathbf{z}_i$ the set of discrete values which $\mathbf{z}$ can take. In order to account for the case of continuous random variables $\mathbf{z}$, a generalization of Eqs. (40) and (42) is needed. A common way to write the Shannon entropy in these cases is through the differential (or continuous) entropy:

$$S[P] = -\int d\mathbf{z} P(\mathbf{z}) \log P(\mathbf{z}), \tag{43}$$

where $P(\mathbf{z})$ is the continuous probability density function (p.d.f.). This quantity is known to present several problems: first of all, it is not guaranteed to be non-negative and bounded, in contrast to the discrete case; secondly, it is not invariant under coordinate transformation. Hence $S[P]$ is not a good definition of absolute uncertainty and it can only be used to calculate *variations* in the information content of a p.d.f.. In order to remove these issues, Eq. (43) is conveniently replaced by the Kullback-Leibler divergence of $P(\mathbf{z})$ from an arbitrary reference probability measure $Q(\mathbf{z})$:

$$S[P||Q] = D_{KL}[P||Q] = \int d\mathbf{z} P(\mathbf{z}) \log \left(P(\mathbf{z}) / Q(\mathbf{z}) \right). \tag{44}$$

2 Because of the required normalization of the probability distribution, one of the constrained observables in Eq. (41) is always the constant function: $X_0(\mathbf{z}) = 1$.

This quantity — also known as relative entropy — is well-defined for both continuous and discrete random variables. In the second case it reads:

$$D_{KL}(p_1, \dots, p_n || q_1, \dots, q_n) = \sum_i p_i \log\left(p_i/q_i\right). \tag{45}$$

The Kullback-Leibler divergence is always non-negative, $S[P||Q] \geq 0$, with equality holding only when $P(\mathbf{z}) = Q(\mathbf{z})$. Moreover, the chain rule ensures that the quantity in Eq. (44) is invariant under any coordinate transformation.

The Kullback-Leibler divergence measures the information gain achieved in passing from a distribution $Q(\mathbf{z})$ — representing our prior knowledge — to $P(\mathbf{z})$. As we seek for the least structured model (i.e. having the minimal information content) that is compatible with the given data, Jaynes' maximum entropy principle must be restated as a minimum relative entropy principle in the continuous case [172]. The minimal agnostic choice is that of a uniform distribution $Q(\mathbf{z})$: in this case, Eq.(44) reduces to the ill-defined quantity in Eq. (43), if we change the sign and add up an infinite constant, and minimization of $S[P||Q]$ is mapped into a maximization of $S[P]$.

The constrained optimization problem can finally be solved by means of the Lagrange multipliers method. We define the new Lagrangian function

$$\tilde{S}[P||Q] = S[P||Q] + \theta_0 \left(\int d\mathbf{z} P(\mathbf{z}) - 1 \right) + \sum_\mu \theta_\mu \left(\int d\mathbf{z} P(\mathbf{z}) X_\mu(\mathbf{z}) - \langle X_\mu(\mathbf{z}) \rangle_{\text{exp}} \right) \tag{46}$$

and solve the variational problem[3] associated to it:

$$\delta \tilde{S}[P||Q] = 0 \iff \int d\mathbf{z} \delta P(\mathbf{z}) \left[\log \frac{P(\mathbf{z})}{Q(\mathbf{z})} + 1 + \theta_0 + \sum_{\mu=1}^{K} \theta_\mu X_\mu(\mathbf{z}) \right] = 0. \tag{47}$$

The maximum entropy p.d.f. is therefore

$$P(\mathbf{z}) = Q(\mathbf{z}) \frac{1}{Z(\boldsymbol{\theta})} e^{-\sum_{\mu=1}^{K} \theta_\mu X_\mu(\mathbf{z})}, \quad \text{with } Z(\boldsymbol{\theta}) = e^{1+\theta_0}. \tag{48}$$

The values of the Lagrange multipliers $\boldsymbol{\theta}$ are determined imposing Eqs. (41).

When $Q(\mathbf{z})$ is uniform, the optimal p.d.f. $P(\mathbf{z})$ in Eq. (48) takes the form of a canonical Boltzmann distribution, with the same Hamiltonian as in Eq. (38). The minimum relative entropy method is therefore equivalent to the Bayesian ML approach, with the likelihood belonging to the canonical exponential family. This equivalence already provides a first clue about the correctness of the maximum entropy method, which was originally introduced with vague and intuitive motivations based on the analogy with equilibrium thermodynamics. Another proof for the correctness of the maximum entropy method in the limit of infinite data has been later provided by Jaynes himself, resorting to combinatorial arguments and the notion of typicality [103].

3.1.4 Learning dynamical models

A ubiquitous application of statistical inference methods regards time series analysis. The goal of time series analysis is to answer to the following question: given a

3 Uniqueness of the solution is guaranteed if the constrained observables $X_1 \dots X_K$ have nonzero fluctuations, as we explained above.

set of sequentially measured data, how can we reliably extract a dynamical model that generates the experimental traces and can be used to make predictions?

Of course, the problem is vast and of interest in many disciplines, and it is not possible to give a fair account of the colossal corpus of scientific work which has been done over the past century, nor of the work which, in spite of this long tradition, is still in progress. Let us just notice that the question above can be tackled in two distinct ways. The first one is seeking for a *deterministic* description of the observed dynamical system from suitable analysis of the experimental data. The second one is building a *stochastic* dynamical model, where fluctuations are not considered as superimposed random variables due to an external noise source (e.g. due to measurement errors) but as a characterizing part of the system's evolution. Much effort has been devoted to the first kind of problems, and especially to the analysis of linear one-dimensional time series [19, 107]. Our interest is however focused on the reconstruction of stochastic processes.

Non-parametric inference of stochastic processes

Within the stochastic class, a first distinction must be made between continuous-time and discrete-time models. Experimental data always come in the form of discrete time series, but it is often convenient to consider them as discrete observations of an underlying continuous-time process. We assume to have measured a state variable $\mathbf{x}$ whose dynamics is governed by a stochastic differential equation (we restrict to the case of continuous processes, without jumps):

$$d\mathbf{x}(t) = \mathbf{F}(\mathbf{x}(t), t)dt + d\mathbf{f}(\mathbf{x}(t), t). \tag{49}$$

$\mathbf{F}(\mathbf{x}(t), t)$ indicates the deterministic force, while $d\mathbf{f}(\mathbf{x}(t), t)$ is the stochastic force, integrated over the infinitesimal time interval dt. The most common situation is that of diffusion processes, where

$$d\mathbf{f}(\mathbf{x}(t), t) = \mathbf{g}(\mathbf{x}(t), t)d\mathbf{W}(t), \tag{50}$$

with $\mathbf{W}(t)$ a Wiener process. Eq. (49) then reduces to a standard Langevin equation with drift vector $\mathbf{D}^{(1)}(\mathbf{x}, t) = \mathbf{F}(\mathbf{x}, t)$ and diffusion matrix $D^{(2)}_{ij}(\mathbf{x}, t) = \sum_k g_{ik}(\mathbf{x}, t)g_{kj}(\mathbf{x}, t)$. Conventionally, we assume that the Langevin equation is integrated à la Itô. The corresponding Fokker Planck (or forward Kolmogorov) equation reads:

$$\frac{\partial}{\partial t}P(\mathbf{x}, t|\mathbf{x}', t') = -\sum_i \frac{\partial}{\partial x_i}\left[D^{(1)}_i(\mathbf{x}, t)P(\mathbf{x}, t|\mathbf{x}', t')\right] + \frac{1}{2}\sum_{ij}\frac{\partial^2}{\partial x_i \partial x_j}\left[D^{(2)}_{ij}(\mathbf{x}, t)P(\mathbf{x}, t|\mathbf{x}', t')\right]. \tag{51}$$

The two quantities $\mathbf{D}^{(1)}(\mathbf{X}(t), t)$ and $\mathbf{D}^{(2)}(\mathbf{X}(t), t)$ are sufficient to completely describe the dynamical behavior of the system, and the goal of inference is to provide an estimate for them.

Even in this case, two types of approaches — namely parametric and non parametric — can be employed. Non parametric approaches apply when the functional form of $\mathbf{F}$ and $\mathbf{g}$ is not specified but must be reconstructed from the data. A straightforward strategy is to adopt the method of cumulants [80]. The method exploits the

definitions of $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$ [159] and reconstructs the drift vector and the diffusion matrix assigned to the bin located at $\mathbf{x}_\alpha$ through the following extrapolations:

$$\mathbf{D}^{(1)}(\mathbf{x}_\alpha, t) = \lim_{\tau\to 0} \frac{1}{\tau}\mathbf{M}^{(1)}(\mathbf{x}_\alpha, t, \tau), \qquad \mathbf{D}^{(2)}(\mathbf{x}_\alpha, t) = \lim_{\tau\to 0} \frac{1}{\tau}\mathbf{M}^{(2)}(\mathbf{x}_\alpha, t, \tau), \tag{52}$$

where $\mathbf{M}^{(1)}$ and $\mathbf{M}^{(2)}$ are the first and second conditional moments, respectively:

$$\mathbf{M}^{(1)}(\mathbf{x}_\alpha, t, \tau) = \frac{1}{N_\alpha} \sum_{\mathbf{x}(t)\in I_\alpha} [\mathbf{x}(t+\tau) - \mathbf{x}(t)] \tag{53}$$

$$\mathbf{M}^{(2)}(\mathbf{x}_\alpha, t, \tau) = \frac{1}{N_\alpha} \sum_{\mathbf{x}(t)\in I_\alpha} \left[\mathbf{x}(t+\tau) - \mathbf{x}(t) - \tau\mathbf{D}^{(1)}(\mathbf{x}(t), t)\right]^2. \tag{54}$$

We denote by N_α the number of points contained in the bin I_α. Higher order moments $\mathbf{M}^{(n)}(\mathbf{x}_\alpha, t, \tau)$ can also be calculated; they are expected to converge in the small τ limit to the Kramers-Moyal (KM) coefficients of the process [159], so, depending on how small they are, they determine how good the assumption of Gaussian white noise is. If $\mathbf{M}^{(3)}$ and $\mathbf{M}^{(4)}$ are significantly different from zero, the p.d.f. of fluctuations has non-Gaussian tails, and the Fokker-Planck equation (51) must be replaced by the full KM expansion.

The method of cumulants for simple diffusion processes is robust and general, but it requires an enormous amount of data. For this reason, it is practically impossible to use in the absence of a stationary process. Many real experiments produce few, or even single time series, from which the generative model must be extracted. Since it is only the *expectation* of the cumulant to converge to the corresponding KM coefficient, N_α must be large in all the considered bins. Yet, the number of bins grows exponentially with the dimension of the system's phase space: as a result, non-parametric methods cannot be applied but to very few-dimensional stationary time series.

Semi-parametric inference approaches

Improved performance with a reduced statistics is exhibited by semi-parametric inference methods, such as the recently introduced Stochastic Force Inference (SFI) algorithm [81]. The key idea of semi-parametric methods is to use general, tunable parametrizations involving, in principle, infinite-dimensional sets of parameters. A typical example of (static) semi-parametric model for a random variable $\mathbf{x}$ is the one we can obtain by expressing the p.d.f. of interest, $P(\mathbf{x})$, as a linear combination of a family of functions, like polynomials or Fourier modes.

Similar ideas can be adopted in the dynamical case: SFI approximates the force field $\mathbf{F}(\mathbf{x})$ of Eq. (49) — independent of time for a steady-state process — as a linear combination of a finite basis of n_b orthonormal functions $\{c_\alpha(\mathbf{x})\}_{\alpha=1\dots n_b}$. The algorithm allows us to reconstruct the drift vector as:

$$F_i(\mathbf{x}) \simeq \sum_{\alpha=1}^{n_b} \hat{F}_{i,\alpha}(\mathbf{x}) c_\alpha(\mathbf{x}), \tag{55}$$

where $\hat{F}_{i,\alpha}$ are the maximum likelihood estimators of the projection coefficients $F_{i,\alpha}(\mathbf{x}) = \int d\mathbf{x} P(\mathbf{x}) F_i(\mathbf{x}) c_\alpha(\mathbf{x})$. In a similar way, the inhomogeneus diffusion ma-

trix can be reconstructed through the estimated projection coefficients $D_{ij,\alpha}(\mathbf{x}) = \int d\mathbf{x} P(\mathbf{x}) D^{(2)}_{ij}(\mathbf{x}) c_\alpha(\mathbf{x})$ (denoted $\hat{D}_{ij,\alpha}$):

$$D^{(2)}_{ij}(\mathbf{x}) \simeq \sum_{\alpha=1}^{n_b} \hat{D}_{ij,\alpha}(\mathbf{x}) c_\alpha(\mathbf{x}). \tag{56}$$

For large enough n_b and a suitable choice of the basis of functions, Eqs. (55)–(56) provide a reliable estimate of the force field and of the inhomogeneous diffusion tensor. An advantage of this approach is that it does not demand for an exact knowledge of the generative model of the observed data, like in the non-parametric case (although it is helpful to make an informed choice of the basis, which may have a strong impact on the performance of the algorithm), yet it allows us to tackle the inference problem as in the parametric setting. The semi-parametric SFI algorithm has been successfully applied to infer the force field and Langevin equations describing the dynamics of motile cells [21] and to develop new methods of analysis of Brownian movies [91].

Parametric estimation of diffusion processes

The last class of methods for continuous-time stochastic processes is that of parametric ones. A recent review of parameteric inference approaches for Langevin equations can be found in [120]; additional reviews are referenced therein.

As already pointed out, the common trait of all parametric techniques is that the statistical model must be fully specified. In the case of stochastic dynamics, this demands us to explicitly derive the finite-time propagator of the process $P_\tau(\mathbf{x}|\mathbf{x}') \equiv P(\mathbf{x}, t+\tau|\mathbf{x}', t)$ — at least approximately, in the small τ limit. The finite-time propagator is what we need in order to build the likelihood function of the discrete process:

$$P(\mathbf{x}_{0:L}|\boldsymbol{\theta}) = \prod_{i=1}^{L} P_{\tau,\boldsymbol{\theta}}(\mathbf{x}_i|\mathbf{x}_{i-1}) P_{\boldsymbol{\theta}}(\mathbf{x}_0), \tag{57}$$

where $\mathbf{x}_{0:L} = (\mathbf{x}_0, \ldots, \mathbf{x}_L)$ is the observed time series and $\boldsymbol{\theta}$ is the set of parameters. $P_{\boldsymbol{\theta}}(\mathbf{x}_0)$ is the probability distribution of the initial condition. Since a very small class of processes are exactly solvable, the propagators appearing in Eq. (57) are typically approximated using a Taylor-Itô expansion. If observation lags τ are too coarse, Markov Chain Monte Carlo (MCMC) approaches can be used to augment the data, by introducing latent time series points, and reach a more accurate estimation of the likelihood [72, 71].

Parametric methods generally show very high precision, even with a moderate amount of data, and a lower computational cost compared to the previously described approaches. This beneficial effect of a model-specific design of the inference procedure comes however at the cost of reduced robustness. The accuracy of the result strongly depends on the accuracy of the hypotheses about the dynamical model as well as on the quality of the data; besides, in this context the presence of measurement noise can be seen as inaccurate modeling.

As a concluding remark, let us notice that, independently of the generative model we choose, parametric methods always demand to formulate the problem in terms

of discrete stochastic processes. Experimental data are indeed obtained from recordings of the considered dynamical process (e.g. the motion of a tracer colloidal particle in a solvent) with a finite number of frames per second. If we are interested in finding a continuous-time description of the process, which is generally easier to interpret and more analytically tractable to make predictions, we need to employ a proper discretization. Once the process is discretized, both discrete-time and continuous-time models are treated on equal grounds, and the same methods apply.

I will conclude this introductory overview about parametric inference methods for dynamic processes by discussing the maximum caliber principle, i.e. the dynamic counterpart of the maximum entropy principle.

Maximum caliber method

The maximum entropy method described in Section 3.1.3 naturally extends to path distributions. The formalism can be developed for general processes, including continuous ones, but it was originally introduced for trajectories composed of discrete time steps [76]. The idea is just to switch from probability density functions of random state variables to probability density functionals for stochastic paths.

The path entropy for a discrete-time process reads:

$$H_T = -\sum_{i_{0:T}} P(i_{0:T}) \log P(i_{0:T}), \tag{58}$$

where $i_{0:T} = (i_0, \dots, i_T)$ is the micro-trajectory, i.e. any sequence of states visited by the system at times $0, \dots, T$. The sum in Eq. (58) denotes that, in addition to the discrete time dynamics, the discrete nature of the state space is also assumed; a generalization to a continuous states space can be obtained as in Eq. (44). Constraints on the dynamics, indexed by μ, are expressed through a set of equations $g_\mu[P(i_{0:T})] = 0$. These constraints are linear in $P(i_{0:T})$ when the maximum caliber method is used as an inference tool. Indeed the constraints must express the equivalence between expected values and experimental averages for the dynamical quantities of interest, like average fluxes, velocities, or rates of conversion. Explicitly, the conditions to impose read:

$$\sum_{i_{0:T}} P(i_{0:T}) X_\mu(i_{0:T}) = \langle X_\mu(i_{0:T}) \rangle_{\text{exp}}, \qquad \mu = 0, \dots, K, \tag{59}$$

where $X_0 = 1$ to guarantee the normalization of the path probability density.

The same maximization principle adopted in Section 3.1.3 can now be applied to the path probability of the dynamical system. The resulting pathways weights are

$$P(i_{0:T}) = \frac{1}{Z_T(\boldsymbol{\theta})} e^{-\sum_{\mu=1}^{K} \theta_\mu X_\mu(i_{0:T})}, \quad \text{with } Z_T(\boldsymbol{\theta}) = \sum_{i_{0:T}} e^{-\sum_{\mu=1}^{K} \theta_\mu X_\mu(i_{0:T})}. \tag{60}$$

The Lagrange multipliers θ_μ are computed after Eq. (59). The dynamical partition function $Z_T(\boldsymbol{\theta})$ plays the same role as its static counterpart: all the statistical properties of the process can be derived from it. At this level, the micro-trajectory is not different from any standard static micro-state living in a $(T \times D)$-dimensional space (if D is the phase space dimension): what carries information about the dynamics of the system is the causality relation — i.e. how information is transferred — among the components of this $(T \times D)$-dimensional vector.

In this framework, the measured observables dictate the nature of the dynamical model. When Eqs. (59) consist of linear constraints on the joint probabilities of k subsequent points, $P(i_h, i_{h+1}, \ldots, i_{h+k}; t_h)$ for $k = 0, 1, \ldots, n$, then the path probability which maximizes the generalized path entropy corresponds to that of an n-th order Markov process, whose transition probability is of the form $P(i_T|i_{T-1}, \ldots, i_{T-n})$ [115, 156]. A particularly relevant case is $n = 1$. Here we have only singlet and pairwise constraints expressed, respectively, by K_0 and K_1 equations:

$$\langle X_{\mu_0} \rangle_{\text{exp}} = \frac{1}{T+1} \sum_{t=0}^{T} \sum_{i_t} X_{\mu_0}(i_t) P(i_t) \qquad \mu_0 = 1, \ldots, K_0, \tag{61}$$

$$\langle Y_{\mu_1} \rangle_{\text{exp}} = \frac{1}{T} \sum_{t=1}^{T} \sum_{i_t} Y_{\mu_1}(i_t, i_{t-1}) P(i_t, i_{t-1}) \qquad \mu_1 = 1, \ldots, K_1, \tag{62}$$

in addition to the normalization condition $\sum_{i_{0:T}} P(i_{0:T}) = 1$. Maximization of the generalized path entropy

$$\begin{aligned} \tilde{H}_T = &- \sum_{i_{0:T}} P(i_{0:T}) \log P(i_{0:T}) - \sum_{\mu_0=1}^{K_0} \theta_{\mu_0} \left[\frac{1}{T+1} \sum_{t=0}^{T} \sum_{i_t} X_{\mu_0}(i_t) P(i_t) - \langle X_{\mu_0} \rangle_{\text{exp}} \right] \\ &- \sum_{\mu_1=1}^{K_1} \theta_{\mu_1} \left[\frac{1}{T} \sum_{t=1}^{T} \sum_{i_t} Y_{\mu_1}(i_t, i_{t-1}) P(i_t, i_{t-1}) - \langle Y_{\mu_1} \rangle_{\text{exp}} \right] - (1+\rho) \left[\sum_{i_{0:T}} P(i_{0:T}) - 1 \right] \end{aligned} \tag{63}$$

yields a dynamic partition function of the form

$$Z_T(\boldsymbol{\theta}) = \mathbf{v}^{\dagger} \cdot \mathbf{G}^T \cdot \mathbf{v}, \tag{64}$$

where each component v_t of the vector $\mathbf{v}$ is a function of the state i_t, and $\mathbf{G}$ is the transfer matrix, having nonzero entries only in the elements $G_{t,t}$ and $G_{t,t\pm1}$. From Eq. (64) it is possible to deduce that the conditional probability

$$P(i_m|i_{m-1}, \ldots i_0) = P(i_0, \ldots i_m)/P(i_0, \ldots i_{m-1}) \tag{65}$$

only depends on the states i_m and i_{m-1}, so it corresponds to the transition rate of a first order Markov process, $P(i_m|i_{m-1})$.

Even when $n > 1$, the maximum caliber principle induces Markovianity: this is a general property akin to the fact that the maximum entropy principle lets the likelihood belong to the canonical exponential family — Eq. (48). Maximum caliber can therefore be used as a principled method to derive from the data dynamical models described by n-th order master equations. Successful applications are reported in the literature and include the analysis of neuron firing patterns, single-molecule dynamics, biochemical cycles, to name but a few [156]. However, it is important to remember that Markovianity is a strong property which is not always satisfied by real processes.

3.2 PARAMETRIC INFERENCE FROM BIRDS DATA: PREVIOUS RESULTS

Active matter systems are typically made up of at least hundreds or thousands of constituents; reconstructing the stationary p.d.f. or the stochastic dynamics of

all their degrees of freedom through non-parametric methods, without any model assumption, is clearly an out-of-reach task. On the contrary, parametric inference has been successfully applied to some available data-sets [11, 136]. In this section I will review previous results obtained from the experimental analysis of flocks of European starlings.

The analyzed experimental data come from two different data collection campaigns in the field. Experiments consist in recordings of several flocking events, which are realized using stereometric photography and computer vision techniques [45, 46]. During the experiment, three synchronized high speed cameras are placed at different angles of view[4]. The 2D image sequence collected by each camera is segmented and combined with the remaining ones, in order to reconstruct the 3D trajectories of all the individuals in the flock. Sophisticated algorithms [43, 2] allow for a detailed and pretty accurate reconstruction of the microscopic state of the system over time. This unprecedented availability of data motivated rephrasing the study of real flocks as an inverse problem and opened the possibility to create a tight connection between ethological observation and theoretical condensed matter models.

3.2.1 Inverse Heisenberg problem

In a first work, Bialek et al. [11] directly reconstructed from the data the stationary probability distribution of birds' directional degrees of freedom by means of a maximum entropy method. Assuming that birds are in a stationary state, and that their relevant degrees of freedom are just the set of normalized 3D velocity vectors $\mathbf{v}_i/|\mathbf{v}_i| = \boldsymbol{\sigma}_i$, the authors look for the minimally structured probability density function $P(\{\boldsymbol{\sigma}_i\})$ that is compatible with the constraint

$$\langle \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j \rangle_{\text{exp}} = C_{ij}, \tag{66}$$

where $\langle \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j \rangle_{\text{exp}}$ is the experimental average and $C_{ij} = \int d\boldsymbol{\sigma}_1, \dots d\boldsymbol{\sigma}_N P(\{\boldsymbol{\sigma}_i\})\, \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j$ is the two-point correlation function associated to the probability density function to determine. The maximum entropy model resulting from this set of constraints is a Heisenberg model with general couplings J_{ij}, whose stationary distribution reads:

$$P(\{\boldsymbol{\sigma}_i\}) = \frac{1}{Z(\{J_{ij}\})} \exp\left[\frac{1}{2}\sum_{ij} J_{ij}\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j\right]. \tag{67}$$

A couple of remarks are in order. Firstly, the symmetry of the correlation matrix implies the symmetry of the couplings J_{ij}, so that the independent Lagrange multipliers to fix from the data are $N(N-1)/2$, with N the number of birds. Eq. (67) thus represents an equilibrium canonical ensemble, where we can recognize the classical Heisenberg Hamiltonian. A second remark concerns the parametrization of the coupling matrix. In principle, each independent J_{ij} parameter can take a different value. In this case, however, the complexity of the inverse problem would escalate rapidly

4 In the first campaign six cameras were used, with interlaced synchronization, in order to increase the acquisition frame rate [45].

for large sized systems and would make it impossible to have enough statistics to reconstruct all the parameters of the model. The correlation of a pair of birds C_{ij} is indeed supposed to change over time, as birds reshuffle their relative positions, so that experimental averages cannot be computed as time averages. On the contrary, it is reasonable to assume that the interaction is local and that C_{ij} does not depend on birds' identity but only on their mutual distance, be it metric or topological. Therefore, a parametrization of J_{ij} as a function of an interaction radius is conveniently used in [11]. We should then interpret Eq.(67) as a conditional probability distribution $P(\{\sigma_i\}|\{\mathbf{r}_i\})$, for a fixed spatial configuration of the flock, not as the marginalized distribution $\int \left(\prod_{i=1}^{N} d\mathbf{r}_i\right) P(\{\sigma_i, \mathbf{r}_i\})$.

The results obtained by Bialek et al. with their inference procedure show that a parametrization of J_{ij} by means of a single coupling strength J and a topological connectivity matrix n_{ij} is robust across different flocking events. The entry of the topological connectivity matrix n_{ij} is equal to one when bird j is among the first n_c neighbors of bird i, and it is zero otherwise (n_c-nearest neighbor model). If, in contrast, a metric interaction mechanism is assumed in the inference procedure, it yields results that depend on the sparseness of the flock, confirming that the notion of distance ruling birds' social behavior is not a Euclidean one, a fact firstly verified in [4].

In order to test whether the obtained maximum entropy parameter-free model offers a good mechanistic description of the whole system, the authors compared its predictions with measurable quantities which were not included among the constrained observables. Specifically, they found a good agreement between the predictions of the maximum entropy model and the experimental observations for the four-point velocity correlations as well as for the full longitudinal and transverse velocity-velocity correlations as functions of the spatial distance. Such agreement indicates that the chosen set of constrained observables is a sufficient statistics for the inverse problem. From a physical point of view, the maximum entropy model tells us that interactions are pairwise, short-ranged and topological (see also [41]).

3.2.2 First order dynamics

The static maximum entropy approach described above completely ignores the sequential relation between the analyzed snapshots, which are only used in [11] to reconstruct from the data an equilibrium-like steady state distribution. However, since experiments give access to full birds' trajectories, it is also possible to build dynamical models based on empirical observations.

This has been firstly done in [136] by applying the systematic approach of maximum caliber to another data-set of European starlings. The authors extracted the minimally structured dynamical model which is consistent with the measured correlation functions $\langle \sigma_i(t) \cdot \sigma_j(t) \rangle_{\text{exp}}$ and $\langle \dot{\sigma}_i(t) \cdot \sigma_j(t) \rangle_{\text{exp}}$. In practice, since recorded trajectories are made up of discrete time points, these correlation functions are estimated through the two-point spatio-temporal correlations of birds' flight directions at equal time — $\langle \sigma_i(t) \cdot \sigma_j(t) \rangle_{\text{exp}}$ — and at a temporal distance of Δt — $\langle \sigma_i(t) \cdot \sigma_j(t + \Delta t) \rangle_{\text{exp}}$ — where Δt is the inverse frame rate. Therefore we have sin-

glet and pairwise constraints for the caliber maximization, yielding a first order Markov process.

The model corresponds to the discretization of a first order SDE of the form:

$$\dot{\sigma}_i = \left(\sum_j J_{ij}\sigma_j + \xi_i \right)_\perp , \tag{68}$$

where ξ_i are independent white noises and $\perp$ indicates the orthogonal projection onto the plane perpendicular to σ_i, for any $i = 1 \dots N$. An exact mapping between the maximum caliber approach and the maximum likelihood method that can be developed for the SDE (68) is exhibited in [35]. The dynamical model in Eq. (68) is a continuous-time variant of the standard Vicsek model, having, in principle, an independent J_{ij} parameter for each bird pair. However, as in the static case, a parametrization is not only mandatory, for statistical reasons, but also realistic. In [136] the authors adopt again the assumption of topological interaction, motivated by the experimental evidence of previous works [4, 11]. The set of inferred parameters then comprises: the range of interaction n_c, i.e. the fixed number of interacting neighbors, the strength of interaction J and the temperature of the heat bath, or noise amplitude, T.

Let us notice that in this dynamical case the constraint on the symmetry of J_{ij} may be lifted, allowing for out-of-equilibrium modeling. Enforcing the symmetrization of the J_{ij} matrix, the authors could compare the results of the dynamic inference procedure to those of the static maximum entropy approach obtained in [11]. This comparison shows that dynamic and static inference procedures give compatible results only in *local equilibrium*, i.e. in an adiabatic regime where the rewiring of the neighbors' network occurs on much longer time scales than the local relaxation dynamics of birds' orientations. Precisely, denoting τ_{relax} the time scale needed by orientation fluctuations to relax on a spatial scale of the order of the effective interaction radius, and τ_{network} the typical time scale of network reshuffling, compatibility of the two approaches is ensured for $\tau_{\text{relax}} \ll \tau_{\text{network}}$. Conversely, if the system is far from the local equilibrium condition, the static approach overestimates the number of interacting neighbors, since unresolved reshuffling creates effective long range interactions. This fact has been checked in [136] through numerical simulations, where the separation between the two time scales is tunable. Real flocks seem to work in a regime of local equilibrium: the two time scales are well separated, at least if τ_{relax} is inferred using the first order model in Eq. (68).

3.3 A NOVEL INFERENCE METHOD FOR SECOND ORDER STOCHASTIC DIFFERENTIAL EQUATIONS

3.3.1 Partial observation of stochastic processes

As discussed in Section 2.2.1 experiments on natural starling flocks revealed that their dynamics is better described by second order SDEs for birds' flight directions, rather than first order ones. In order to take this evidence into account, a step for-

ward is required in the development of dynamical inference approaches, compared to the ones discussed in the previous section. The interest of the methodological problem is broad. Birds are just an example of biological systems where second order dynamical models are needed to account for the phenomenon under study: recent experimental findings on a variety of living systems, including migrating cells [22], bacteria [85], worms [180], to groups of animals on a larger scale [83, 84], indicate that in all these cases the observed behavior cannot be explained with a first order dynamic process, but requires a higher order description.

The emergent dynamics of all the above systems share three fundamental ingredients: an effective inertia, dissipation, and a stochastic contribution. These features are quite general and do not depend on the presence of interactions in the system, nor on whether detailed balance is violated. Therefore, we focus on the simplest case of equilibrium Langevin dynamics, assuming that only a partial observation of the state variable of the system can be performed, and try to develop a general Bayesian inference scheme for such stochastic processes. The general framework is the following one: let us assume that the available experimental data are sequences of points $x_{0:_L} = (x_0, x_1, \dots x_L)$ uniformly separated in time by Δt, and that the underlying dynamics is described by a Langevin equation of the form:

$$\ddot{x} = -\eta \dot{x} + f(x) + \sigma \xi, \tag{69}$$

where $f(x)$ is a general deterministic force, $\sigma^2 = 2T\eta$, and ξ is a standard white noise: $\langle \xi \rangle = 0$, $\langle \xi(t)\xi(t') \rangle = \delta(t - t')$. Without lack of generality, the inertial mass is set to 1. Since the noise is additive, it is unnecessary to distinguish between Itô and Stratonovich integration.

Let us call $\boldsymbol{\theta}$ the irreducible set of parameters that enter into Eq. (69), namely the effective damping coefficient η, the effective temperature T, and the parameters of the force term $f(x)$. Following the Bayesian ML approach, the goal becomes to find a tractable expression for the dynamical likelihood associated to the observation of process (69). The theory of stochastic processes provides us with an explicit but formal expression for the transition probability $P\left(x(t)|x(0), \dot{x}(0)\right)$, involving, in general, integro-differential operators. A closed form solution for the stochastic process may be generally unknown or complicated, especially for many body or off-equilibrium systems, but finely time-resolved data are available. What we look for is then an approximated expression for the probability of discrete trajectories.

A general recipe is the following:

1. As a preliminary step, Eq. (69) can be conveniently rewritten as a set of two first order equations:

$$\dot{x} = v\,, \qquad \dot{v} = -\eta v + f(x) + \sigma \xi. \tag{70}$$

2. Since the dynamics is Markovian when parametrized by the vector variable $\mathbf{q} = (x, v)$, the probability of a discrete trajectory in this space, given the initial condition $\mathbf{q}_0 = (x_0, v_0)$, can be split into a product of propagators:

$$P(\mathbf{q}_1, \dots, \mathbf{q}_L | \mathbf{q}_0) = \prod_{n=1}^{L} P(\mathbf{q}_n | \mathbf{q}_{n-1}). \tag{71}$$

3. Following [68], one can exploit any update rule based on a Taylor-Itô expansion to approximate, within a certain order of accuracy, the propagator over a small time interval Δt:

$$P(\mathbf{q}_n|\mathbf{q}_{n-1}) = P_{(k)}(\mathbf{q}_n|\mathbf{q}_{n-1}) + o(\Delta t^k). \tag{72}$$

Eq. (72) can be replaced into Eq. (71) to get an approximated expression for the probability density of the sequence of points in phase space:

$$P_{(k)}(\mathbf{q}_1, \dots, \mathbf{q}_L|\mathbf{q}_0) = \prod_{n=1}^{L} P_{(k)}(\mathbf{q}_n|\mathbf{q}_{n-1}) + o(\Delta t^k). \tag{73}$$

4. Marginalizing over the velocity-like degrees of freedom one gets a probability distribution depending on the x's only. This projection operation on the subspace of x variables is where the original Markov property of Eq. (71) is generally lost. A crucial remark, beyond the non-Markovian nature of the resulting dynamics, is that this procedure does not simply consist of removing the intermediate variables $v_1, \dots, v_L$, but also of eliminating the initial condition v_0.

When this strategy is adopted, the first thing we need is a discrete integration scheme for Eq. (69) or Eq. (70). The choice of the integration scheme is a delicate point. Although the naive intuition is that any convergent — even if slowly — discretization scheme should work for small Δt, in fact the order of approximation of the temporal discretization affects the mathematical properties of the discrete path integral measure and, consequently, the consistency of ML estimators [68, 90].

Alternatively, one can follow a second strategy, summarized as *'first marginalize, then discretize'*, in contrast to the *'first discretize, then marginalize'* strategy discussed above. The starting point is here the generalized Langevin equation (GLE) corresponding to the desired process, Eq. (69), which can be obtained adopting the Mori-Zwanzig formalism [200]:

$$\dot{x} = v_0 e^{-\eta t} + \int_0^t ds K(t-s) f(x(s)) + \zeta(t), \tag{74}$$

where $K(t) = e^{-\eta t}$ and $\zeta(t) = \int_0^t ds K(t-s)\xi(s)$ is the effective noise. This formalism shows that, when projected from the full phase space into the x subspace, the dynamics acquires memory, described by a friction kernel $K(t)$, and color in the noise. In the limit of infinitely long trajectories, the relation $\langle \zeta(t)\zeta(t') \rangle \propto K(|t-t'|)$ holds asymptotically, and it reduces to the second fluctuation dissipation theorem when $f(x)$ is linear.

Discrete update equations can now be obtained by integrating Eq. (74) on Δt intervals, and by self-consistently removing v_0. We notice that, for arbitrary forces $f(x)$, the corresponding term cannot be exactly integrated and it needs to be approximated at small Δt. The fact that the derivative of the measured coordinate x enters parametrically through v_0 in the GLE stems from the second order nature of the process. Its elimination, which is necessary to retrieve a stochastic difference equation where only the x variable appears, is connected to the problem anticipated in point 4 of the procedure outlined above.

Since the order of the discretization and marginalization operations should be exchangeable, the two strategies must be equivalent. In the following section we show how the simplest inference schemes derived from Euler-like discretizations of Eq. (70) do not satisfy this requirement, whereas higher order discretization schemes, strongly convergent as at least $O(\Delta t^{3/2})$, 'commute' with the marginalization operation and allow us to find consistent results with both strategies.

3.3.2 Naive Maximum Likelihood scheme based on Euler discretization

Derivation of the Likelihood

Discrete integration schemes for SDEs are well known in the literature in connection to numerical simulation methods [153]. The simplest one among them is the Euler-Maruyama scheme. If we apply the Euler-Maruyama discretization to Eq. (69), the resulting discrete update equations are:

$$x_{n+1} - x_n = \Delta t\, v_n\,, \qquad v_{n+1} - v_n = -\eta \Delta t\, v_n - \Delta t\, f(x_n) + \sigma \Delta t^{1/2}\, r_n, \tag{75}$$

with r_n IID random variables of normal distribution $\mathcal{N}(0,1)$, for $n = 0, \ldots, L-1$. They define a two-dimensional Markov process with a deterministic update for the x variables, which lets δ-functions appear in the discrete propagator of the process. In this case one can explicitly marginalize over the velocity degrees of freedom, and eliminate the initial condition v_0. Indeed, to this order of approximation, information on v_0 is fully equivalent to information on x_1. From this marginalization, a fully factorized probability distribution for the discrete sequence is obtained:

$$P_{(1)}(x_2, \ldots, x_L | x_0, x_1) = \prod_{n=1}^{L-1} P_{(1)}(x_{n+1} | x_n, x_{n-1}), \tag{76}$$

where transition probabilities are defined as follows:

$$P_{(1)}(x_{n+1} | x_n, x_{n-1}) = \frac{1}{Z_n} e^{-S_n(x_{n+1}, x_n, x_{n-1})}\,, \tag{77}$$

with

$$S_n = \frac{1}{2\sigma^2 \Delta t^3} \left[x_{n+1} - 2x_n + x_{n-1} + \eta \Delta t (x_n - x_{n-1}) - \Delta t^2 f(x_n) \right]^2\,, \quad Z_n = \sqrt{2\pi\sigma^2 \Delta t^3}. \tag{78}$$

A factorization of $P(x_L, \ldots x_2 | x_1, x_0)$ into a product of transition probabilities of this kind is possible because the random variables appearing in the x difference equation, obtained from Eq. (75) through variable elimination, are independent. This is a crucial but artificial feature occurring only at this level of approximation: more accurate discretization procedures produce an effective noise for the x variables which is correlated in time. We can nonetheless derive the ML inference scheme associated to this Euler discretization.

Connection to Maximum Caliber Principle

Using Eqs. (77)–(78) , an expression for the dynamical likelihood as product of transition probabilities for a second order master equation is recovered. This corresponds to the discrete path probability one would obtain adopting a maximum

caliber approach [156] when certain time-dependent observables are taken as fixed. In the simple example of a one-dimensional harmonic oscillator, they are the equal-time correlations, one-time-step correlations and two-time-step correlations of the process. Indeed, taking $f(x) = -\omega_0^2 x$ in Eq.(69) and rearranging the sum of S_n's in Eq. (76), the reduced minus-log-likelihood can be written as:

$$\frac{\mathcal{L}(\eta, T, \omega_0^2)}{L-1} = \frac{1}{2}\ln(2\pi\sigma^2\Delta t^3) + \frac{1}{2\sigma^2\Delta t^3}\Big[C_s' + (2-\eta\Delta t+\omega_0^2\Delta t^2)^2 C_s + (1-\eta\Delta t)^2 C_s'' \\ +2(1-\eta\Delta t)F_s - 2(2-\eta\Delta t+\omega_0^2\Delta t^2)G_s - 2(1-\eta\Delta t)(2-\eta\Delta t+\omega_0^2\Delta t^2)G_s'\Big], \tag{79}$$

where:

$$C_s = \frac{1}{L-1}\sum_{n=1}^{L-1} x_n x_n; \quad C_s' = \tfrac{1}{L-1}\textstyle\sum_{n=1}^{L-1} x_{n+1}x_{n+1}; \quad C_s'' = \frac{1}{L-1}\sum_{n=1}^{L-1} x_{n-1}x_{n-1}; \tag{80}$$

$$G_s = \frac{1}{L-1}\sum_{n=1}^{L-1} x_n x_{n+1}; \quad G_s' = \tfrac{1}{L-1}\textstyle\sum_{n=1}^{L-1} x_n x_{n-1}; \quad F_s = \frac{1}{L-1}\sum_{n=1}^{L-1} x_{n-1}x_{n+1}. \tag{81}$$

For long enough trajectories, the time averages in Eqs. (80)–(81) can be thought of as equivalent to ensemble averages defining the autocorrelation function of the process, evaluated at equal time points — $C_s, C_s', C_s'' \sim C(0) = \langle x(t_0)x(t_0)\rangle_{t_0}$ —, at a time distance of Δt — $G_s, G_s' \sim C(\Delta t) = \langle x(t_0+\Delta t)x(t_0)\rangle_{t_0}$ — and at a time distance of $2\Delta t$ — $C_s, C_s', C_s'' \sim C(2\Delta t) = \langle x(t_0+2\Delta t)x(t_0)\rangle_{t_0}$. Fixing these time-dependent observables in the maximum-caliber procedure yields the discrete path probability of a linear second order Markov process [115], also known as AR(2) process [19].

Euler-ML estimators

Minimization of the quantity in Eq. (79) with respect to η, T and ω_0^2 gives us the inference formulas for the parameters of the harmonic oscillator. Let us notice that the Euler scheme adopted in Eq. (79) is not the unique choice we have, at this order of approximation: several variants are available. We focus on three particular examples: the standard explicit Euler-Maruyama scheme (EM-fwd), its implicit variant (EM-bkd), and the symmetric BBK scheme [23]. The three of them may be obtained from the second order SDE Eq. (69) by approximating first and second time derivatives adopting a forward, backward or symmetric prescription respectively:

$$\text{[EM-fwd]} \qquad x_{n+1} - (2-\eta\Delta t)x_n + (1-\eta\Delta t+\omega_0^2\Delta t^2)x_{n-1} = \sigma\Delta t^{3/2} r_{n-1} \tag{82}$$

$$\text{[EM-bkd]} \qquad (1+\eta\Delta t)x_{n+1} - (2+\eta\Delta t-\omega_0^2\Delta t^2)x_n + x_{n-1} = \sigma\Delta t^{3/2} r_{n+1} \tag{83}$$

$$\text{[BBK]} \qquad \left(1+\tfrac{\eta\Delta t}{2}\right)x_{n+1} - (2-\omega_0^2\Delta t^2)x_n + \left(1-\tfrac{\eta\Delta t}{2}\right)x_{n-1} = \sigma\Delta t^{3/2} r_n. \tag{84}$$

From each of them, we extract a factorized expression for the dynamical likelihood, of the form of Eq. (76). The reduced minus-log likelihood, defined as

$$\frac{\mathcal{L}}{L-1} := \frac{-\ln P(x_2,\ldots,x_L|x_0,x_1)}{L-1}, \tag{85}$$

corresponds in this factorized case to the temporal average of the quantity $S_n + \ln Z_n$. This quantity results into slightly different expressions in the three cases above.

Minimizing it with respect to the parameters of the model yields the following optimal values:

$$\eta^*_{\text{fwd}} = \frac{1}{\Delta t} \frac{G_s + G'_s - 2C_s + \frac{G'_s}{C''_s}(2G'_s - C''_s - F_s)}{-C_s + \frac{G'^2_s}{C''_s}} \; ; \tag{86}$$

$$\omega^{2*}_{0\,\text{fwd}} = \frac{1}{\Delta t^2} \frac{(2 - \eta\Delta t)G'_s - (1 - \eta\Delta t)C''_s - F_s}{C''_s} \; ; \tag{87}$$

$$\begin{aligned} T^*_{\text{fwd}} = &\frac{1}{2\eta\Delta t^3}[C'_s + (2 - \eta\Delta t^2)^2 C_s + (1 - \eta\Delta t + \omega_0^2\Delta t^2)^2 C''_s - 2(2 - \eta\Delta t)G_s + \\ &2(1 - \eta\Delta t + \omega_0^2\Delta t^2)F_s - 2(2 - \eta\Delta t)(1 - \eta\Delta t + \omega_0^2\Delta t^2)G'_s] \; . \end{aligned} \tag{88}$$

For the sake of brevity, we report here only the parameter estimators obtained with the Euler-forward discretization; the remaining formulas for $\left(\eta^*_{\text{bkd}}, T^*_{\text{bkd}}, \omega^{2*}_{0\text{bkd}}\right)$ and $\left(\eta^*_{\text{BBK}}, T^*_{\text{BBK}}, \omega^{2*}_{0\,\text{BBK}}\right)$ are reported in Appendix A. The same procedure we followed for the stochastic harmonic oscillator can be adopted to derive the Euler-ML estimators of any other second order process, be it additive or non-additive. A more general formalism is illustrated in Appendix A.

The 2/3 bias

At this point, having an explicit inference method for a solvable test process (the harmonic oscillator), it can be both numerically and analytically tested. We simulate discrete trajectories of the stochastic harmonic oscillator in several damping conditions using an exact integrator [86], with a numerical time step $\tau^{sim} = 0.005$. We apply the inference formulas to discrete data sets sampled from synthetic trajectories at time intervals $\Delta t \geq \tau^{sim}$. This choice mimics real experiments, where the time resolution is fixed by the acquisition apparatus, while the true microscopic time-scale of the dynamics is unknown. Filtering the synthetic trajectories in time is a good blind inspection tool to check the robustness of the inferred continuous stochastic model, without prior knowledge about the time scales of the process.

Results in Fig. 3 show that a systematic error in the estimation of the damping coefficient emerges, which can be cast into a constant rescaling factor close to $2/3$ for the inferred value η^*, as compared to the true value η^{sim}. It is worth remarking that this rescaling is *independent of* Δt, as clearly visible in Fig. 3a, so increasing the resolution of the acquisition instruments is of no help in improving the estimation of the damping coefficient. On the contrary, the estimation of the remaining parameters is in agreement with the parameter values used in the simulations, as shown in Figs. 3c – 3d.

Our numerical results for the stochastic harmonic oscillator agrees with the results of Refs. [147, 117], whose authors pointed out, in a non-Bayesian framework, the failure of the same naïve embedding strategy for second order SDEs. For a more rigorous discussion in the Bayesian (ML) setting, see [90]. We stress that the Euler discretization is the simplest and most commonly used extrapolation of the derivative of an observed variable from its finite increment:

$$\dot{x}(t) \simeq \frac{x(t + \Delta t) - x(t)}{\Delta t}. \tag{89}$$

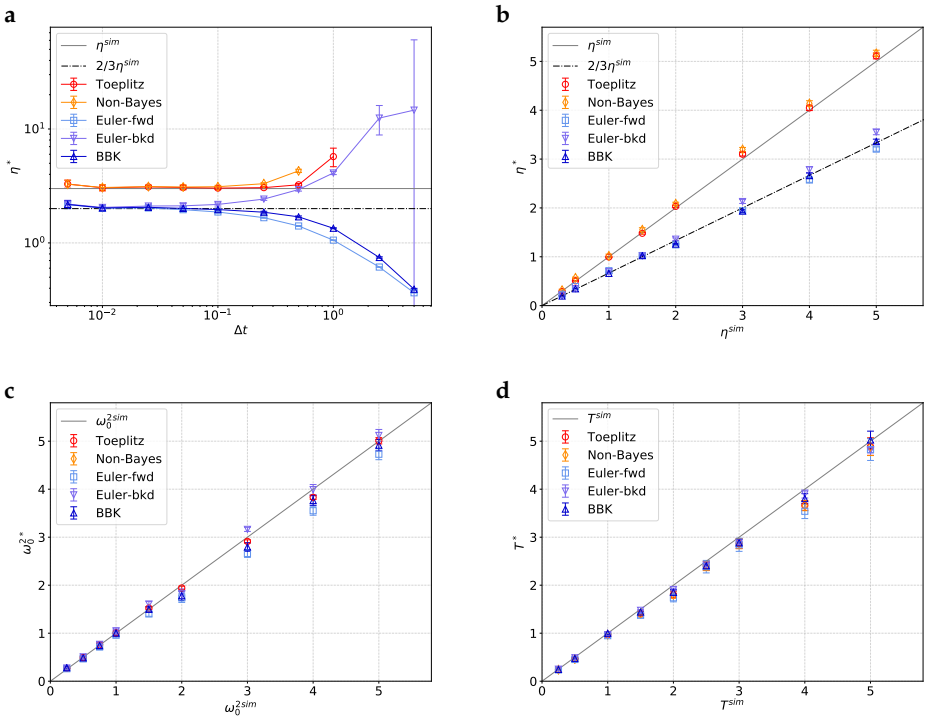

Figure 3: Inference results for the stochastic harmonic oscillator. Sample trajectories are obtained from exact numerical integration of Eq. (69) with parameters η^{sim}, ω_0^{sim} and T^{sim}. The simulation time step τ^{sim} is always $\tau^{sim} = 0.005$, and it corresponds to the minimum displayed value of Δt in **a**. Points at higher values of Δt are obtained applying the inference procedure to sub-trajectories extracted from the original one. Each of the points displayed in **b** – **d** is obtained from the average of the inference results from different Δt values, in the range where the small Δt approximation is valid. We compare the accuracy of all the schemes derived in App. A, both from a first order Taylor-Itô expansion (Euler-fwd, Euler-bkd, BBK) and from a second order expansion (Toeplitz, Non-Bayes). **a** : Inferred damping coefficient η^*. Averages over 10 sample trajectories of 5000 points (for any Δt) are reported with their 1.96 SD (0.95 CI for Gaussian distributions). Simulation parameters are: $T = 1$, $\omega_0 = 1$, $\eta = 3$. **b** : Inferred damping coefficient η^* vs true simulation parameter η^{sim}: results from higher order methods follow the line of slope 1, whereas numerical results from naïve methods fall on the line of slope $2/3$. The remaining parameters are fixed: $T = 1$, $\omega_0 = 1$. **c** : Inferred stiffness of the harmonic oscillator ω_0^{2*} vs true simulation parameter ω_0^{2sim}. All the schemes give correct results in this case in the whole explored range of values. Simulation parameters: $\eta = 3$, $T = 1$. **d** : Inferred temperature T^* vs the true value of the simulation parameter T^{sim}: again, results from all schemes fall on the line of slope 1 in the whole explored range of values. Remaining simulation parameters: $\eta = 1.5$, $J = 1$.

This approximated estimation of the velocity works if one observes the system in the overdamped regime, i.e. when $\eta\Delta t \gg 1$ and $\omega_0/\eta < \infty$, and the effective dynamics can be described by a first order equation: in this case, Euler-based inference schemes provide excellent results. However, when a non-Markovian signal is observed, coming for instance from the partial observation of a higher dimensional Markov process, these schemes are bound to fail.

A simple argument can help us to understand what is missing, and why the parameter η is the one affected by the approximation. Assuming that experimental averages perfectly reproduce ensemble averages, we can replace into the formula for the estimator η^* the known analytical expression for the self-correlation of the harmonic oscillator in the stationary regime:

$$C(t) = \frac{T}{\omega_0^2} e^{-\gamma t} \left[\cos\left(\sqrt{\omega_0^2 - \gamma^2}\, t \right) + \gamma \frac{\sin(\sqrt{\omega_0^2 - \gamma^2}\, t)}{\sqrt{\omega_0^2 - \gamma^2}} \right], \tag{90}$$

and perform a Taylor expansion around $t = 0$ to compute $C(\Delta t)$ and $C(2\Delta t)$. The underlying assumption of the whole procedure is indeed that the time lag Δt between subsequent points is small, compared to the typical time scales of the dynamics. Using this expansion in Eq. (86) we obtain an expression for η^* depending only on the derivatives of $C(t)$ at $t = 0$:

$$\eta^* \simeq \frac{1}{\Delta t} \frac{2\dot{C}(0) - \frac{2}{3}\dddot{C}(0)\Delta t^2 - \frac{\ddot{C}(0)\dot{C}(0)}{C(0)}\Delta t^3}{2\dot{C}(0) + \ddot{C}(0)\Delta t + \frac{1}{C(0)}\left[\dot{C}(0) + \frac{1}{2}\ddot{C}(0)\Delta t\right]^2 \Delta t}. \tag{91}$$

Knowing explicitly $C(t)$ for the harmonic oscillator, one can compute the desired derivatives:

$$C(0) = \frac{T}{\omega_0^2}\,; \quad \dot{C}(0) = 0\,; \quad \ddot{C}(0) = -T\,; \quad \dddot{C}(0) = \eta T\,. \tag{92}$$

Proper combinations of these quantities allow us to extrapolate all the parameters of the model[5]. By replacing Eqs. (92) into Eq. (91), we obtain:

$$\eta^* = -\frac{2}{3}\frac{\dddot{C}(0)}{\ddot{C}(0)}\left[1 + O(\Delta t)\right] = \frac{2}{3}\eta + O(\Delta t). \tag{93}$$

We find then, at the leading order, a rescaling factor of $2/3$, as observed in numerical tests. No rescaling factors appear for the other inferred parameters: performing the same replacement and expansion of the analytical correlation functions in the inference formulas of T and ω_0, we see that temperature and pulsation are correctly retrieved from proper combinations of $C(0)$ and $\ddot{C}(0)$.

This result gives us a clue to understand the origin of the Δt-independent rescaling factor for η. We recall that the order of strong and weak convergence of the Euler-Maruyama scheme for an additive process like (69) is 1. The elimination of the velocity variable in Eq. (75) makes terms of order $O(\Delta t^{3/2})$ appear, even if the starting accuracy of the expansion is $O(\Delta t)$. This means that Eq. (79) has been inconsistently derived retaining only some of the $O(\Delta t^{3/2})$ contributions; in turn there are missing $O(\Delta t^3)$ contributions to the quadratic fluctuations of x. This fact explains why Eq. (93) is incorrect and shows the need of higher order discretization schemes for stochastic second order dynamics.

5 The importance of the first derivative as a quantity to discriminate between first and second order dynamics in oscillator-like models has already been stressed in [44, 34], with explicit reference to complex interacting systems. With inference, we can go beyond the binary answer provided by $\dot{C}(0)/C(0)$, proportional – through a time scale factor – to 1 or to 0 for first or second order dynamics respectively, and give a quantitative estimation of the damping regime in which a system operates, employing all the derivatives at $t = 0$ up to the third one.

3.3.3 An approximate analytical method

Higher order discretizations

The minimum requirement for an inference method that exploits only local dynamical information is to reproduce fluctuations correctly at least up to $O(\Delta t^3)$, for second order processes. The lowest order of convergence of the discretization scheme is therefore $O(\Delta t^{3/2})$.

If we follow the recipe described in 3.3.1 with an $O(\Delta t^{3/2})$ scheme, we reduce to a sequence of intertwined Gaussian integrals for the marginalization of $v_1 \dots v_L$, which may be cumbersome to compute for arbitrary length of the trajectory. Therefore, it is convenient to work with update equations in x space. They can be obtained either from a temporal discretization of the GLE (74) or from a self-consistent elimination of the velocity variables in the discrete-time equations resulting from a second order Taylor-Itô expansion of Eq. (70). In the first case, it is possible to integrate the GLE (74) in the interval between two subsequent observation points, t_n and t_{n+1} and manipulate it to remove the dependency on the initial condition v_0. We obtain:

$$x_{n+1} - x_n - e^{-\eta\Delta t}(x_n - x_{n-1}) = \frac{1 - e^{-\eta\Delta t}}{\eta} \int_{t_{n-1}}^{t_{n+1}} \Psi(t - t_n) f(x(t)) dt + \zeta_n \tag{94}$$

where $\zeta_n = \int_{t_{n-1}}^{t_{n+1}} \Psi(t - t_n)\xi(t)dt$ and

$$\Psi(t) = \frac{e^{\eta t} - e^{-\eta\Delta t}}{1 - e^{-\eta\Delta t}} \left[\theta(t + \Delta t) - \theta(t)\right] + \frac{1 - e^{\eta(t-\Delta t)})}{1 - e^{-\eta\Delta t}} \left[\theta(t) - \theta(\Delta t - t)\right] \tag{95}$$

where $\theta(x)$ denotes the Heaviside function.

So far, these equations are exact. Some approximation is needed to evaluate the integral of the force. Various methods have been investigated in the literature; among the simplest ones is the Langevin Impulse method [173], which approximates the integral with the function at the midpoint, leading to

$$x_{n+1} = x_n + e^{-\eta\Delta t}(x_n - x_{n-1}) + \frac{1 - e^{-\eta\Delta t}}{\eta} \Delta t f(x_n) + \zeta_n. \tag{96}$$

Alternatively, taking the first order expansion of the force around the midpoint, $f(x(t)) \simeq f(x_n) + \frac{t - t_n}{\Delta t}\left[f(x_n) - f(x_{n-1})\right]$, one recovers the stochastic Verlet algorithm [93], which is one order more accurate than Eq. (96).

Independently of the chosen method, we can schematically rewrite the discretized equation as:

$$x_{n+1} + F(x_n, x_{n-1}; \boldsymbol{\mu}) = \zeta_n, \tag{97}$$

with $\boldsymbol{\mu}$ a set of effective parameters for the drift. In addition to them, we have an extra parameter, the temperature T, associated to ζ_n. The noise variables ζ_n in Eq. (97) are still zero-mean Gaussian variables, but no longer independent. This is the crucial difference with the Euler-Maruyama scheme, which takes into account only the diagonal entries of the covariance matrix $C_{nm} = \langle \zeta_n \zeta_m \rangle$.

Toeplitz approximation of the noise covariance matrix

To order $O(\Delta t^3)$, for any consistent derivation of Eq. (97), one can deduce that the covariance matrix of random increments $\mathbf{C}$ has a 'nearest-neighbour' structure of the kind:

$$C_{nm} = \langle \zeta_n \zeta_m \rangle = a\,\delta_{n,m} + b\,\delta_{n,m\pm 1} \tag{98}$$

where

$$a \simeq \frac{2}{3} 2T\eta\Delta t^3\,; \quad b \simeq \frac{1}{6} 2T\eta\Delta t^3. \tag{99}$$

Hence $\mathbf{C}$ has the form of a symmetric tridiagonal Toeplitz matrix of order $L-1$. These mathematical features carry a physical meaning: first of all, the presence of non-vanishing off-diagonal elements is the signature of a colored noise. Secondly, the fact that the matrix is banded means that the correlation of the noise variables is finite-ranged, i.e. that the associated memory kernel, in a continuous-time description, decays fast [135]. Finally, the Toeplitz structure is synonymous with shift invariance.

A more careful derivation of the update equations in x space would require shift invariance not to hold and the first entry of the covariance matrix C_{11} to be different from the other elements of the main diagonal. Eq. (94) is in fact not valid for the first integration step, where the initial conditions intervene. In this respect the structure of the data also poses the problem of the elimination of the initial condition v_0 in favor of x_0 and x_1. Even if not able to perform it explicitly without stationarity assumptions, we can argue[6] that it has the effect of modifying the covariance matrix in the following way:

$$C = \begin{pmatrix} \tilde{a} & b & \dots & 0 \\ b & a & . & \vdots \\ \vdots & . & \ddots & b \\ 0 & \dots & b & a \end{pmatrix}, \tag{100}$$

where the shift invariance expressed by the Toeplitz structure of Eq. (98) is then broken at the beginning of the time series. Despite that, the error we make by replacing $\tilde{a}$ with a in the quasi-Toeplitz matrix (100) is negligible in the limit of long trajectories, as better discussed in the next section about the method's limitations. Intuitively, since the breaking of the shift invariance occurs only at the first step, the longer the trajectory, the more similar this is to a truly shift invariant situation.

Apart from the difficulty in determining correctly $\tilde{a}$, the advantage of replacing the true covariance matrix Eq. (100) with a Toeplitz matrix is that the inverse of the Toeplitz matrix is explicitly known, as well as the eigenvalues [104, 134]:

$${C^{-1}}_{nm} = \frac{2}{L} \sum_{k=1}^{L-1} \frac{\sin\left(\frac{nk\pi}{L}\right) \sin\left(\frac{mk\pi}{L}\right)}{a + 2b\cos\left(\frac{k\pi}{L}\right)}; \quad \lambda_k = a + 2b\cos\left(\frac{k\pi}{L}\right). \tag{101}$$

6 The argument comes from the fact that the elimination of the velocity variables $v_1 \dots v_n$ can be independently performed from that of v_0. This is explicitly visible from the derivation in the previous section.

These quantities are those which enter into our likelihood expression:

$$P_{(2)}(x_L,\dots,x_2|x_1,x_0) = \frac{1}{Z}\exp \\ -\frac{1}{2}\sum_{n,m=1}^{L-1}(x_{n+1}+F(x_n,x_{n-1};\boldsymbol{\mu}))C^{-1}{}_{nm}(x_{m+1}+F(x_m,x_{m-1};\boldsymbol{\mu})), \tag{102}$$

being $Z = \left[(2\pi)^{L-1}\det\mathbf{C}\right]^{1/2} = \left[\prod_{k=1}^{L-1}2\pi\lambda_k\right]^{1/2}$. Eq. (102) can be derived by noticing that Eq. (97) identifies a transformation $\zeta_{1:L-1}\mapsto x_{2:L}$ whose Jacobian determinant is one. The reduced minus log-likelihood associated to Eq. (102) reads:

$$\mathcal{L} = \frac{L-1}{2}\ln\left(2\pi\frac{2}{3}T\eta\Delta t^3\right)+\frac{1}{2}\sum_{k=1}^{L-1}\ln\left(2+\cos\left(\frac{k\pi}{L}\right)\right) \\ +\frac{3/2}{LT\eta\Delta t^3}\sum_{n,m=1}^{L-1}\left[(x_{n+1}+F(x_n,x_{n-1};\boldsymbol{\mu}))\tilde{A}_{nm}(x_{m+1}+F(x_m,x_{m-1};\boldsymbol{\mu}))\right], \tag{103}$$

where $\tilde{A}_{nm}$ is the rescaled inverse covariance matrix:

$$\tilde{A}_{nm} = \sum_{k=1}^{L-1}\frac{\sin\left(\frac{nk\pi}{L}\right)\sin\left(\frac{mk\pi}{L}\right)}{2+\cos\left(\frac{k\pi}{L}\right)}. \tag{104}$$

Let us highlight that the inverse of the covariance matrix does not preserve a banded structure. This means that, even if noise correlations are local in time, two-time functions of every pair of points of the trajectory enter into the minus-log-likelihood. Hence Eq. (102) cannot be factorized. Factorization corresponds to a block structure for $\mathbf{C}^{-1}$, which implies a block structure for $\mathbf{C}$. This is incompatible with the tridiagonal Toeplitz or quasi-Toeplitz nature of the covariance matrix, where off-diagonal elements are of the same order as the diagonal ones.

Nonetheless, having built an explicit discrete path integral measure, a maximum likelihood approach is practicable, and it reduces to minimizing the quantity $\mathcal{L} = -\ln P(x_L,\dots,x_2|x_1,x_0)$ with respect to the parameters of the model. Thanks to the regularities of Eq. (103), the minimization of $\mathcal{L}$ can be performed analytically in the case of the harmonic oscillator and of other simple single-particle processes. The optimization procedure can be performed semi-analytically also for some many-particle systems. In these cases an additional parameter is typically the interaction range of the particles' interaction. In general, once an expression for $\mathcal{L}$ is given, a large number of optimization algorithms are available to minimize it with respect to all the extra parameters that do not allow for a fully analytical approach. Explicit formulas for the Toeplitz-ML estimators in the case of some simple processes that we studied numerically can be found in Appendix A.

3.3.4 Application

The inference method we proposed is quite general and can in principle be applied to a variety of processes of the following form (including their generalizations in more than one dimension):

$$\dot{x} = v\,, \qquad \dot{v} = -\eta v + f(x) + \sigma(x)\xi, \tag{105}$$

where $f(x)$ and $\sigma(x)$ are arbitrary functions. The only strict requirement is that just a simple linear dependence on the RHS of Eq. (105) is allowed. The linearity in v of the force terms is exploited in the discretization scheme of the LI integrator [173], and it is pivotal for us to have easily invertible tridiagonal covariance matrices.

We checked the performance of the scheme on several processes of interest:

1. The damped harmonic oscillator in a heat bath, which is exactly solvable and can be used to prove the inconsistency of the naive Euler-ML estimators;
2. Brownian motion in a double-well potential;
3. A one-dimensional mutliplicative process with a suitable positive definite covariance matrix for the random increments;
4. The Inertial Spin Model on a fixed network, in its ferromagnetic phase.

Complete formulas for all of these processes are reported in Appendix A. Here we show and comment on the numerical results obtained from their application.

Linear additive processes

The first process we considered is the stochastic damped harmonic oscillator ($f(x) = -\omega_0^2 x$ in Eq. (69)). We used it as a sample process to test the validity of our method and to compare it to ML methods derived by adopting discretization schemes of order $1/2$. We simulated the oscillator dynamics in its phase space (x, v) using the exact Gillespie algorithm [86], but just observed the positional coordinates to mimic the real experimental scenario. To analyze the performance of the inference method at different values of Δt and observe the impact of time discretization, we further subsampled the simulated trajectories to progressively increase the time separation between subsequent points.

The comparison between higher order and naïve lower order inference schemes for the example of the harmonic oscillator confirms the analytical predictions discussed in Section 3.3.2 (Fig. 3). In any damping regime, the higher order inference method outperforms the naïve scheme in two ways: perturbatively, since the convergence of the parameter estimators is extended to a larger Δt window due to the higher order Taylor-Itô expansion (see Fig. 3a), and non-perturbatively in Δt, since no rescaling factor for the η parameter is required (Fig. 3b). The different behaviour of the various Euler-like schemes at large Δt, where the series expansion is non-asymptotic, is probably related to the details of the discretization rules and their stability properties.

In addition to the Toeplitz-ML method, we show in Fig.3 the results obtained from the application of an alternative parametric but non-Bayesian inference method. Parameter estimators are indeed not unique: we can define many different consistent estimators for the same parameter, with different convergence rates. Maximum likelihood estimators have the advantage of being derived from an optimization principle: the problem is hence well-defined for any kind of process. When the model is simple enough, it is also possible to find relations between measurable quantities and model parameters that can be used to infer the latter. For instance, let us take the discrete-time equation (96) and multiply both sides by x_{n-1}, x_n and

x_{n+1}. Self-consistently averaging over the noise distribution, we obtain a set of three independent equations, from which all the parameters of the dynamical model can be extracted (explicit formulas are derived in Appendix A). Remarkably, the $2/3$ rescaling is absent, since the method is based on a discretization of the SDE which is convergent as $\Delta t^{3/2}$.

Non-linear additive processes

Fig. 4 shows numerical results based on the Toeplitz inference scheme for the the following process, describing the motion of a Brownian particle of unit mass in an anharmonic double-well potential:

$$\dot{x} = v\,, \qquad \dot{v} = -\eta v - kx - \lambda x^3 + \sigma\xi\,, \tag{106}$$

We explored various values of the parameters λ^{sim} and k^{sim} (Fig. 4). In all the explored regimes the inference scheme provides excellent results, showing, in particular, that no bias is introduced by the possible imbalance between linear and nonlinear force terms (values close to the origin are correctly estimated in Fig. 4.d and Fig. 4.f), even if, for a fixed Δt, an increase in the relative error or more noisy estimations cannot be prevented in these conditions (Fig. 4.g). Moreover, no bias is introduced by the fact that, when k assumes a negative value, the particle may be confined in a single minimum of the double-well potential for all the length of the sampled trajectory (see Fig. 4.e).

Inference formulas can be easily generalized to any nonlinear additive process with a polynomial force term $f(x) = \sum_\alpha \mu_\alpha x^\alpha$ – see Appendix A.

Multiplicative processes

We now consider possible generalizations to the case of non-additive noise. An adaptation of our non-Markovian Bayesian inference scheme can be developed for the following class of multiplicative processes:

$$\ddot{x} = -\eta\dot{x} + f(x) + \sigma(x)\xi, \tag{107}$$

with $\xi(t)$ a standard white noise and initial conditions $x(0) = x_0$, $\dot{x}(0) = v_0$. This model has two features: linear dissipation, and a velocity-independent diffusion coefficient proportional to $\sigma^2(x)$. Under these conditions, the memory kernel of the GLE associated to Eq. (107) is explicitly known and, following the same procedure that led to the discretization of the additive process, we obtain an approximated discrete time update rule of the form:

$$x_{n+1} - x_n - e^{-\eta\Delta t}(x_n - x_{n-1}) - \frac{1 - e^{-\eta\Delta t}}{\eta}\Delta t f(x_n) = \zeta_n, \tag{108}$$

where the stochastic term is defined as

$$\zeta_n = \frac{1 - e^{-\eta\Delta t}}{\eta}\int_{t_{n-1}}^{t_{n+1}} dt' \Psi(t' - t_n)\sigma(x(t'))\xi(t'). \tag{109}$$

The function $\Psi(t)$ is defined in the same way as in Eq. (95). The approximation of the force term in Eq. (108) corresponds to that of the Langevin impulse integrator

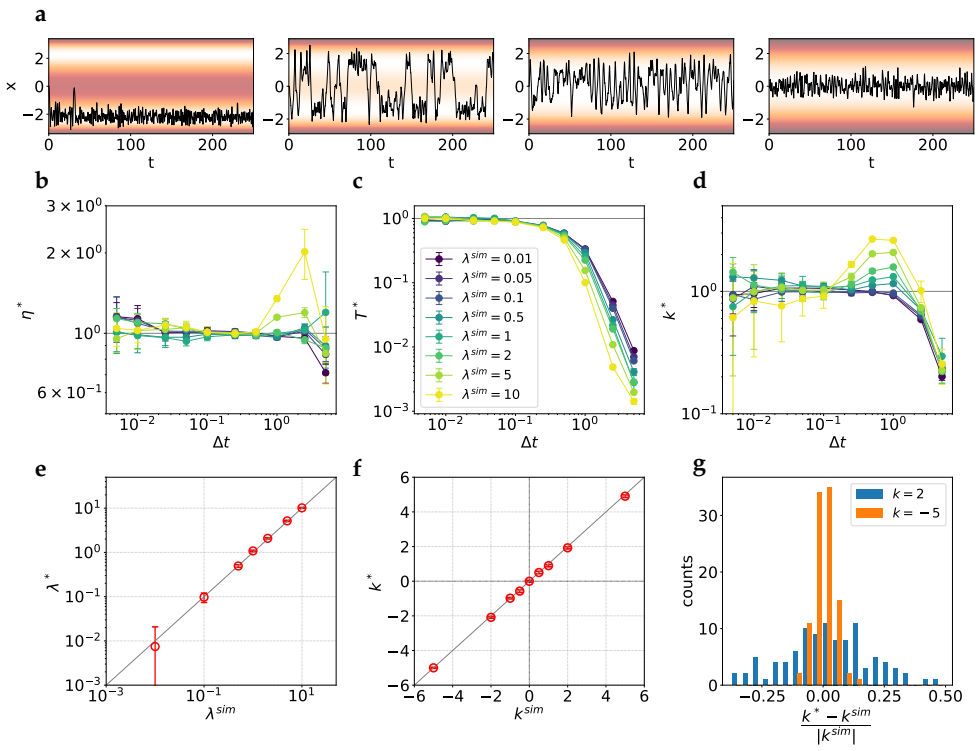

Figure 4: Bayesian inference of the dynamical parameters of a Brownian motion in a force field $f(x) = V'(x)$, with $V(x) = \frac{1}{2}kx^2 + \frac{1}{4}\lambda x^4$. Only the Toeplitz method is applied. **a** : Excerpts of sample trajectories in various landscapes. The strength of the confining potential is qualitatively indicated by the colormap, with light areas corresponding to the minimum of the potential. The following parameters of the simulation are kept fixed: $T = 1$, $\eta = 1$, $\lambda = 1$. By varying the parameter k we realize, from left to right: a strong confinement in a double well potential, with long exit times, at $k = -5$; a switching dynamics with relatively short switching times, at $k = -2$; a marginal situation at $k = 0$; confined Brownian motion in the vicinity of the origin at positive values of k ($k = 5$). **b–d** : Inferred model parameters against subsampling parameter Δt. The true value is equal to 1 in all cases and is marked by the straight grey line. **e** : Inferred vs true value of the nonlinear coefficient λ. **f** : Inferred vs true value of the parameter of the linear force k, assuming both positive and negative values. **g** : Histogram of counts for the relative distance of the inferred parameter k^* to the simulation parameter k^{sim}. With fixed $\lambda^{sim} = 1$ and $k^{sim} = \{2, -5\}$, the weight of anharmonicity varies, but the variance of all the estimated parameters seems to be unaffected. As a result, relative errors decrease for larger $|k|$. 100 trajectories are sampled for each k value shown in the histogram, and $\Delta t = 0.025$ in all cases. In Figs. **b–f** 10 sample trajectories of length 5000 points are considered for each Δt. Errorbars are 1.96 SD.

[173]. As in the additive case, alternative numerical integration schemes for GLEs, such as the stochastic Verlet algorithm [93], can also be used.

From now on we will implicitly refer to the Itô integration prescription; however, thanks to the fact that $\sigma(x)$ only depends on the configurational degree of freedom,

x, the mean square convergence of ζ_n is not affected by a switch to the Stratonovich convention. Up to $O(\Delta t^3)$, the stochastic increments satisfy

$$C_{nm} = \langle \zeta_n \zeta_m \rangle \simeq \frac{2}{3}\Delta t^3 \sigma^2(x_n)\delta_{n,m} + \frac{1}{6}\Delta t^3 \sigma(x_n)\sigma(x_m)\delta_{n,m\pm 1}\,. \tag{110}$$

This choice of off-diagonal terms ensures the positiveness of the matrix, if $\sigma^2(x) > 0$[7]. The covariance matrix also preserves a tridiagonal symmetric structure, but the Toeplitz property is lost since, in the presence of multiplicative noise, shift invariance cannot hold. Nevertheless, we can still build an efficient maximum likelihood inference routine, exploiting the short-time Gaussian approximation of the random increments of the process. Let us rewrite the minus log-likelihood associated to Eq. (108) as

$$\mathcal{L} = \frac{1}{2}\sum_{k=1}^{L-1} \ln \lambda_k(\mathbf{x};\boldsymbol{\nu}) + \sum_{n,m=1}^{L-1} \left[x_{n+1} + F(x_n, x_{n-1};\boldsymbol{\mu})\right] {C^{-1}}_{nm}(\mathbf{x};\boldsymbol{\nu}) \left[x_{m+1} + F(x_m, x_{m-1};\boldsymbol{\mu})\right]. \tag{111}$$

We denote by $\boldsymbol{\mu}$ the subset of parameters including η and the parameters of the conservative potential, and by $\boldsymbol{\nu}$ the subset of parameters appearing in the x-dependent diffusion coefficient $\sigma(x;\boldsymbol{\nu})$. For the $\boldsymbol{\mu}$ parameters, analytical formulas for their max-likelihood estimators can be found as functions of $\boldsymbol{\nu}$, if $f(x)$ is polynomial. On the contrary, inferring the latter set of parameters generally requires numerical optimization (unless $\sigma(\mathbf{x};\boldsymbol{\nu})$ is univariate and has a purely multiplicative dependence on its single ν parameter).

To illustrate the method, we applied it to the multiplicative process in Eq. (107), with $f(x) = -kx$ and $\sigma(x) = \sqrt{a + bx^2}$, where a and b are non-negative parameters. In this case the max-likelihood procedure can be reduced to a one-dimensional numerical optimization. Complete inference formulas are reported in App. A and the results are shown in Fig. 5. These confirm that the method provides a reliable inference tool also in the case of a non-equilibrium multiplicative process, independently of the relative strength of the additive and multiplicative contributions to the noise term, and that the procedure does not require stationarity nor equilibrium assumptions to work.

Inertial Spin Model

We finally consider a passive variant of the Inertial Spin Model (ISM) in the deeply ordered phase. The ISM equations of motion in $d = 3$ are:

$$\dot{\mathbf{v}}_i = -\frac{1}{\chi}\mathbf{v}_i \times \mathbf{s}_i\,, \qquad \dot{\mathbf{s}}_i = -\frac{\eta}{\chi}\mathbf{s}_i + \frac{1}{v_0^2}\sum_j J_{ij}\left(\mathbf{v}_i \times \mathbf{v}_j\right) + \boldsymbol{\xi}_{i\perp}, \tag{112}$$

for $i = 1 \dots N$. Each $\boldsymbol{\xi}_i$ is an independent isotropic white noise:

$$\langle \xi_i^\alpha \rangle = 0\,, \qquad \langle \xi_i^\alpha(t)\xi_i^\beta(s)\rangle = 2T\eta\delta_{\alpha,\beta}\delta(t-s). \tag{113}$$

The symbol $\perp$ denotes the orthogonal projection on the plane perpendicular to the direction of motion of the i-th bird. Compared to Eqs. (18)–(20), we are ignoring

7 There exists a similarity transformation that transforms the matrix in Eq. (110) into a strictly diagonally dominant matrix with positive entries. Since the spectrum is real and it is unchanged under similarity transformations, this ensures the non-negativity of all the eigenvalues of the original covariance matrix.

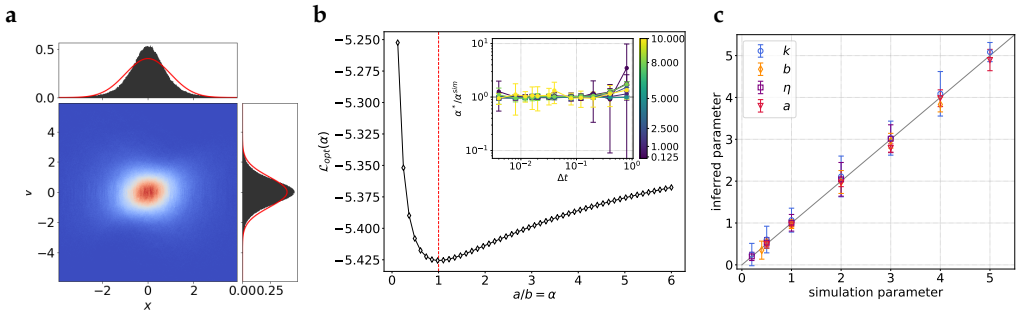

Figure 5: Inference method applied to a multiplicative process. The process is described by Eq. (107) with $f(x) = -kx$ and $\sigma(x) = \sqrt{a + bx^2}$. **a**: The steady state p.d.f. is reconstructed through the fraction of time spent by the system in each region of the phase space for a sample trajectory of length $4 \cdot 10^4$, with $k = 1$, $\eta = 1$, $a = 1$, $b = 1$ and initial condition $(x_0 = 0, v_0 = 0)$. There is a clear difference with the Gaussian distribution having the same second moment (red line), showing the effect of the multiplicative noise. **b**: Analytically optimized negative log-likelihood as a function of the effective parameter $\alpha = a/b$, computed on a sample sub-trajectory of 5000 points, $\Delta t = 0.016$, with the same parameters as in Fig. a. In the inset optimal values of α as a function of Δt are reported. Errorbars correspond to 1.96 SD on 10 sample trajectories of 5000 points for each Δt. The color code refers to the value of α^{sim}, measuring the relative contribution of additive and multiplicative part of the noise term. **c**: Performance of the method in inferring the whole set of parameters of the model.

the update of particles' positions, which therefore behave as passive ones. This approximation is justified in the highly polarized phase of low density systems, since reshuffling of the interaction network occurs on very long timescales in these conditions [8]. Since birds' relative positions do not evolve with time, their constant speed v_0 does not play any role and we set it to 1. Model (112) therefore describes the inertial dynamics of a three-dimensional Heisenberg model on a random graph specified by the J_{ij} matrix.

Notice, however, that the process is not an equilibrium one, in general, as no restriction is imposed on the couplings J_{ij}. Motivated by the findings of [4], we focus on the topological variant of the inertial spin model, where the coupling constant is parametrized as $J_{ij} = J\, n_{ij}$, with $n_{ij} = 1$ if bird j is among the first n_c nearest neighbours of bird i, and $n_{ij} = 0$ otherwise. For a general spatial configuration of the particles, the J_{ij} matrix may be non-symmetric and the action-reaction principle may be not valid. The spatial configuration we choose in our simulations is a random uniform point pattern in a cubic box of linear size L. Periodic boundary conditions are employed.

8 Typical reshuffling times in real flocks of starlings have been quantified in [136], showing a clear separation of scales with respect to the relaxational dynamics of the local order parameter in a first order stochastic description. How reshuffling lowers in the polarized phase is also discussed in Chapter 4 for a different flocking model.

We implemented a numerical integrator in $d = 3$ that combines the leapfrog method with Boris's trick to ensure speed conservation [13]. The set of update equations of our integrator read:

$$\mathbf{v}_i^{n+1} = \mathbf{v}_i^n + (\mathbf{v}_i^n + \mathbf{v}_i^n \times \mathbf{t}^n) \times \mathbf{u}^n, \tag{114}$$

$$\mathbf{s}_i^{n+1/2} = \left(1 + \frac{\eta \Delta t}{2\chi}\right)^{-1} \left[\left(1 - \frac{\eta \Delta t}{2\chi}\right) \mathbf{s}_i^{n-1/2} + \mathbf{v}_i^n \times \left(\frac{J\Delta t}{\chi} \sum_j n_{ij} \mathbf{v}_j^n + \mathbf{g}_i^n\right)\right]. \tag{115}$$

with $\mathbf{t}^n = -\frac{1}{2\chi}\mathbf{s}^{n+1/2}\Delta t$ and $\mathbf{u}^n = 2\mathbf{t}^n/(1 + |\mathbf{t}^n|^2)$. The vector $\mathbf{g}_i^n$ is a three-dimensional isotropic Gaussian variable of zero mean and of variance $\langle \mathbf{g}_i^n \cdot \mathbf{g}_j^m \rangle = \delta_{ij}\delta_{mn} 3 \cdot 2T\eta\, \Delta t$.

Systems of $N = 1000$ boids are simulated to obtain the results shown in this chapter, with a topological interaction range $n_c = 6$ (except for the data in Fig. 6b), alignment strength $J/\chi = 5$ and effective temperature T/χ in the range $[0.2, 1.2]$. When not explicitly indicated, we took $T/\chi = 0.4$, approximately corresponding to a polarization of $\Phi = 0.97$ (for $n_c = 6$). We tried to ensure that the system was sampled in a stationary regime by firstly performing a relaxation run. The stationarity of the system has been evaluated by just looking at the stationarity of the polarization time series. The polarization is defined, in analogy to the global magnetization of a Heisenberg model, as

$$\Phi = \frac{1}{N v_0} \left| \sum_{i=1}^{N} \mathbf{v}_i \right|. \tag{116}$$

Working in the deeply ordered phase (Φ close to 1) allows us to simplify the inference scheme by means of the spin-wave approximation (SWA), introduced in Section 2.2.3. The effect of the approximation is to linearize the force terms, so that the equations of motion (112) take the form of a set of second order SDEs for N coupled harmonic oscillators:

$$\chi \ddot{\boldsymbol{\pi}}_i = -\eta \dot{\boldsymbol{\pi}}_i - J \sum_{j=1}^{N} \Lambda_{ij} \boldsymbol{\pi}_j + \tilde{\boldsymbol{\xi}}_{i\perp}. \tag{117}$$

Here the $\boldsymbol{\pi}_i$'s are the birds' velocity fluctuations, defined as in Section 2.2.3, which lie on the orthogonal plane to the direction of collective motion $\mathbf{n}$. $\Lambda_{ij} = n_c \delta_{ij} - n_{ij}$ is the discrete Laplacian of the birds' network, and $\tilde{\boldsymbol{\xi}}_{i\perp}$ is a two-dimensional isotropic white noise that lives on the same plane as $\boldsymbol{\pi}_i$. The same inference formulas which can be derived for a system of coupled harmonic oscillators can then be applied to Eq. (117) to extract the ISM parameters: $\eta/\chi, J\Lambda_{ij}/\chi, T/\chi$. We report them in Appendix A.

Outside the regime of applicability of the SWA, our inference algorithm cannot be generally applied to extract the parameters of the ISM. Recalling Eq. (21), it is evident that the process does not belong to the class (105), due to the presence of a nonlinear dependence on $\dot{\mathbf{v}}_i$ in the centripetal force term and of a $\mathbf{v}$-dependent drag coefficient $\eta(\mathbf{v}) = \eta/\chi \left(\delta^{\alpha\beta} - v_i^\alpha v_i^\beta / v_0^2\right)$. The only exception is the $d = 2$ case, where we can define $\mathbf{v}_i = v_0 e^{i\theta_i}$ and study the dynamics of the angular variables θ_i, whose equations of motion reduce to those of an inertial Kuramoto model [111].

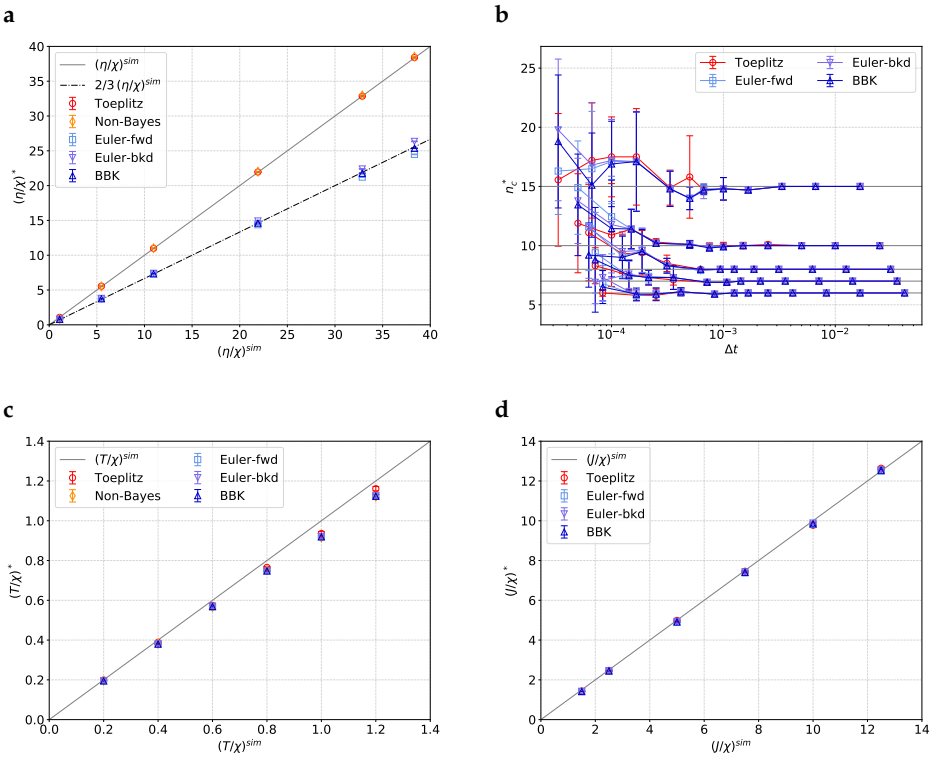

Figure 6: Inference results for the Inertial Spin Model. **a** : Inferred values for the effective damping coefficient η/χ. We notice the emergence of a $2/3$ rescaling factor for naïve methods derived from first order Taylor-Itô expansions. **b** : Inferred topological interaction range from numerical minimization (only obtained in the Toeplitz scheme and in the three Euler variants). Horizontal lines correspond to simulation values. **c** : Inferred values for the parameter T/χ, as derived from Eq. (117). We notice a slight divergence from the slope-1 line, which is especially evident at large temperatures. This is due to the spin-wave approximation (SWA), whose first correction only impacts the temperature parameter and can be explicitly evaluated, as explained in the text. **d** : Inferred values of the interaction strength, $(J/\chi)^*$ vs the parameter value used in simulations, $(J/\chi)^{sim}$. All methods retrieve the correct results. We remark that only the parameters in the left panels, η/χ and T/χ, can be estimated by the non-Bayesian method. In all the simulations we took flocks of $N = 1000$ birds. Points in **a**, **c** and **d** are obtained as in the case of the harmonic oscillator (see Fig. 3). For the $O(\Delta t^{1/2})$ methods we consider different integration schemes: standard Euler (Euler-fwd), inverse (Euler-bkd) and BKK.

We applied different inference strategies to the synthetic trajectories: results are in qualitative agreement to those of the one-dimensional harmonic oscillator. In particular, the expected rescaling factor of $2/3$ for the damping coefficient is retrieved using any Euler-like scheme, as shown in Fig 6a. In contrast, both Bayesian and non-Bayesian inference schemes derived from a higher order Taylor-Itô expansion do not require any rescaling — at least for sufficiently long trajectories.

However, in contrast to the simple one-dimensional case, we could not estimate all of the ISM parameters using the non-Bayesian method. One additional difficulty we must face in the case of N-body dynamics is indeed the estimation of the interaction range. In the ML setting — independently of the order of convergence of the

discretization — numerical approaches can always be employed, even though an explicit analytical minimization of $\mathcal{L}$ with respect to n_c is not operable. The problem is not even costly, in our case, since it simply consists of a one-dimensional optimization, once the n_{ij} matrix is parametrized. In the topological case the problem is further simplified by the fact that n_c is an integer parameter.

We always manage to find the correct value of the parameter used in the simulations, if the statistics is good enough (see Fig. 6b). Wrong estimations of n_c are just due to a rugged reconstruction of the likelihood from the data, which creates fake minima in $\mathcal{L}$. As the number of birds N or the number of trajectory points L is increased, the improved statistics smooths the rugged reconstructed likelihood and the real minimum becomes easier to detect. To this end, another parameter playing a relevant role is the time lapse Δt: when the separation between subsequent data-points is very small compared to the time scales of the system, increments are also very small. Smaller increments correspond to smaller terms in $\mathcal{L}$, whose minimization is then subject to bigger relative errors: this effect is at the origin of what we observe on the left-hand side of Fig. 6b. Once the optimal value of n_c is recovered, it is used to compute the spatially structured correlation functions which enter into the formulas of the remaining parameters (see Appendix A).

Non-Bayesian methods are not based on any likelihood definition: as a result, they do not allow us to infer n_c through numerical optimization. Despite that, an approximated estimation of the effective temperature T/χ and of the damping coefficient η/χ is still possible, as shown in Figs. 6a and 6c. On the contrary, the parameters associated to the interaction potential, n_c and J, are not evaluated. In a non-Bayesian setting, alternative strategies employing semi-parametric methods can be adopted to infer the force terms in the stochastic dynamics of interacting many-body systems [81, 20].

Applied to large interacting systems, our Toeplitz-ML method performs well even for relatively short trajectories. Taking, for instance, trajectories of length[9] $L = 200$ for systems of $N = 1000$ particles already enables us to achieve good accuracy (see Fig. 6). The advantage of moving from the single oscillator to the many-body interacting case is that a restricted number of "local" quantities turn out to dominate and they self-average in sufficiently large systems. As a result, the statistical issue can be at least partially mitigated by averaging over the sample size, rather than relying only on temporal averages as we are compelled to do in the case of the harmonic oscillator.

A final remark concerns the estimated effective temperature T/χ. It is visible from Fig. 6c that all the points are slightly displaced below the line of slope 1, indicating that a systematic bias is present in the inference procedure. The origin of this little bias, which is more or less the same for all the adopted methods, must be identified in the use of the spin-wave approximation. As better highlighted in Fig. 7, as the polarization of the system increases, the difference between the true and inferred values of T/χ becomes more and more negligible, confirming our hypothesis on the origin of the bias. It is possible to include a leading order

9 We recall that what matters in the Toeplitz-ML method is just the number of subsequent data-points, not the length of the trajectory in time units.

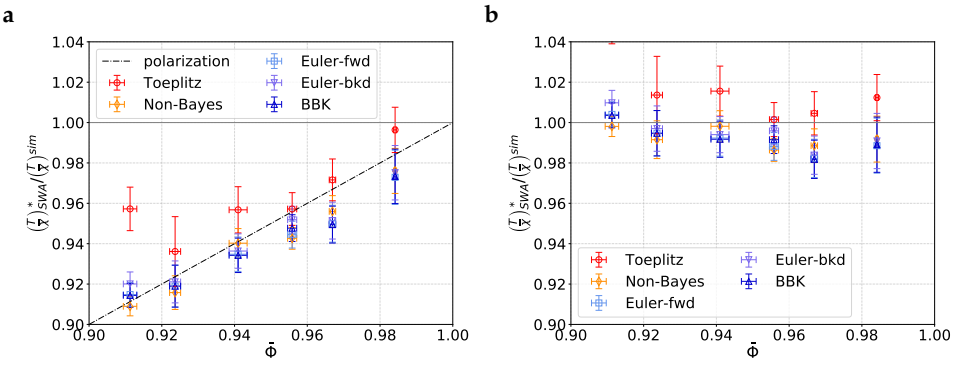

Figure 7: First correction to the SWA. In **a** the raw inferred values of T/χ, obtained using the inference formulas derived from Eq. (117), are reported. In **b** we included the first correction by rescaling the output with the time-averaged polarization, for each sample trajectory. $\bar{\Phi}$ is the average (among different simulated flocks) of the time-averaged polarizations of each flock, at any given temperature. Errorbars for $\bar{\Phi}$ correspond to standard errors, whereas vertical bars represent, as in the other figures, 1.96 SD.

correction by rescaling the inferred parameter by $\frac{1}{N}\sum_{i=1}^{N} v_i^L$, which is by definition equivalent to the polarization of the flock. The result of this a posteriori rescaling is shown in the right-hand panel in Fig. 7.

3.3.5 Limitations

Asymptotic estimator consistency

An estimator θ^* is said to be consistent when it converges in probability to the true value of the parameter θ it aims at estimating, as the sample size goes to infinity:

$$\lim_{M\to\infty} \Pr\left(|\theta^* - \theta| > \epsilon\right) = 0. \tag{118}$$

Providing a mathematical proof of the consistency of our method goes beyond our scope: we limit ourselves to a numerical check of the consistency of estimators (in Section 3.3.4) for various test processes of interest, and for big enough M.

In our case $M = (L-1) \times n_s$, where n_s is the number of sample trajectories and $L-1$ is the number of points in each sequence. To lighten the notation, all the formulas above are given in the case $n_s = 1$, but the generalization is straightforward, even releasing the assumption that all the trajectories have equal length. There are different ways to take the $M \to \infty$ limit: for instance, we can choose to let n_s go to infinity while keeping L fixed, or let L diverge, while keeping n_s fixed. These two situations do not produce the same result, in terms of estimators' consistency. Only in the $L \to \infty$ limit, estimators seem to satisfy the consistency condition (118) — at least, within our numerical accuracy.

This problem leads back to the quasi-Toeplitz approximation of the covariance matrix, which served us as a trick to remove the unobserved initial condition v_0. To explain this, let us take a step back. In a maximum likelihood setting, the first task is to calculate the probability of observing a given sequence of data-points,

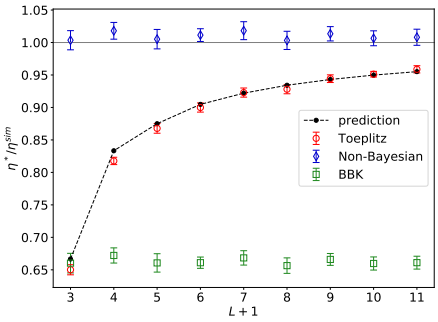

Figure 8: The shift-invariant approximation introduces a finite-size distortion: black points represent the analytical prediction of the rescaling factor $\varrho(L+1) = \eta^*(L+1)/\eta^{sim}$, with $\varrho(L=3) = 2/3$ and $\varrho(L+1) \to 1$ as $L \to \infty$. Numerical results are in agreement with this prediction. No dependence on the trajectory length is found for the non-Bayesian method, nor for Euler-like methods (BBK used here).

knowing the parameters of the model $\boldsymbol{\theta}$. In first order stochastic processes, when all the degrees of freedom allowing for a Markovian description of the dynamics are experimentally accessible and measurements are free of error, there is no ambiguity on how this likelihood should be computed. The probability of a Markov discrete trajectory $(y_0, y_1, \dots y_L)$ is by definition decomposed as

$$P(y_0, y_1, \dots, y_L|\boldsymbol{\theta}) = \prod_{n=1}^{L} P(y_n|y_{n-1};\boldsymbol{\theta}) P(y_0|\boldsymbol{\theta}) \tag{119}$$

If the process is fully observed, y_0 is known and the distribution of the initial condition can be assumed to be sharply peaked in the measured value: $P(Y_0|\boldsymbol{\theta}) = \delta(Y_0 - y_0)$. The measured initial condition acts as a fixed parameter which does not need to be inferred.

On the contrary, for a continuous-time second order stochastic process the initial condition is given by the pair $x(0) = x_0$, $\dot{x}(0) = v_0$ and the corresponding discrete-time propagator is $P(x_L, \dots, x_1|x_0, v_0;\boldsymbol{\theta})$. Unlike x_0, the initial condition on the velocity is not empirically known, so the propagator and the likelihood do not have the same form as before. Let us briefly note that this issue is strictly connected to the embedding problem in stochastic processes, as we might be willing to extrapolate v_0 from the measurements of x_0 and x_1.

In the presence of noise, replacing the conditioning on x_0 and v_0 with a conditioning on x_0 and x_1 is, strictly speaking, incorrect. A well-grounded way to bypass the problem is to use the steady state distribution of v_0 and marginalize over the hidden degree of freedom. Nonetheless, for the development of the suggested inference method, we decided to deal with the initial condition problem in a different way, in order to avoid introducing any prior on v_0 and resorting to purely numerical methods. Firstly, the choice of the LI discretization scheme we made confined the effect of the initial condition only to the first time step, independently of the total number of data-points and of the relation between the decay time of the memory kernel and Δt. Concurrently, we introduced a Toeplitz approximation for the noise covariance matrix, which means neglecting the breaking of shift invariance.

This approximation works well for long trajectories (with many data-points), whereas it fails for very short ones. This fact tells us that when only a fragmented observation of the system is achievable, through repeated short-term recordings, the method is in general not reliable. The Toeplitz method is exact only in the infinite trajectory limit, so the smaller the number of subsequent points, the less accurate the inference scheme becomes. However, the convergence is quite fast, as it is visible in Fig. 8, and the adopted strategy has several advantages: it is simpler than exact marginalization, and applies even when a steady state distribution is not available (e.g. in the multiplicative case, for certain parameter values).

We checked this prediction in numerical simulations of the stochastic harmonic oscillator, keeping constant the total number of points used in the inference procedure, $(L+1)n_S$, and adapting the number of samples n_S as the length $L+1$ of the sample trajectories is varied. A significant deviation of the inferred value from the simulated one is visible in Fig. 8 for small values of L. For small L it is also possible to approximately estimate the distortion introduced by the finite size of the trajectory under the Toeplitz assumption. Following the same idea that led to the prediction of the $2/3$ function for the η parameter of the harmonic oscillator, we expand the two time correlation functions appearing in the Toeplitz inference formulas for small L, obtaining:

$$\eta^* \simeq -\frac{1}{\Delta t} \ln \left(1 + \varrho(L+1) \frac{\dddot{C}(0)}{\ddot{C}(0)} \Delta t \right) [1 + O(\Delta t)] \,. \tag{120}$$

From Eq. (120) we deduce that the Δt-independent rescaling factor of the damping coefficient can be identified with $\varrho(L+1)$ in Eq. (120). The first few values of these rescaling factors are: $\varrho(3) = 2/3$, $\varrho(4) = 5/6$, $\varrho(5) = 7/8$, $\varrho(6) = 19/21$, in good agreement with numerical results. The exact value is only retrieved in the $L \to \infty$ limit, yet time lapse recordings in common experiments with motile objects are typically composed by a much larger number of frames than those shown in Fig. 8. Although we showed that the wrong marginalization of the initial condition can play a role, in practice this effect can hopefully be neglected in many situations.

In summary, what matters in this case is not only the total number of points for statistical reasons — which is the only thing to worry about for standard definitions of estimators' consistency — but also their succession in time. Notice that the total length $(L+1)\Delta t$ of the trajectory in units of the physical time scales of the process is irrelevant; the relevant element is just the number of points $L+1$ of which the trajectory is made up [10].

Robustness to measurement noise

So far, we have not included observation errors in the developed inference scheme, but we assumed that stochastic trajectories are sampled with infinite accuracy. However, data are typically affected by accuracy limitations and other sources of experimental errors. In the current section we show the effects of an additional source of noise on the estimation of the model parameters.

10 This is not surprising if one carefully looks at the expression of the inverse of the tridiagonal Toeplitz matrix Eq. (101), which closely resembles Fourier series expansions. Increasing the number of points

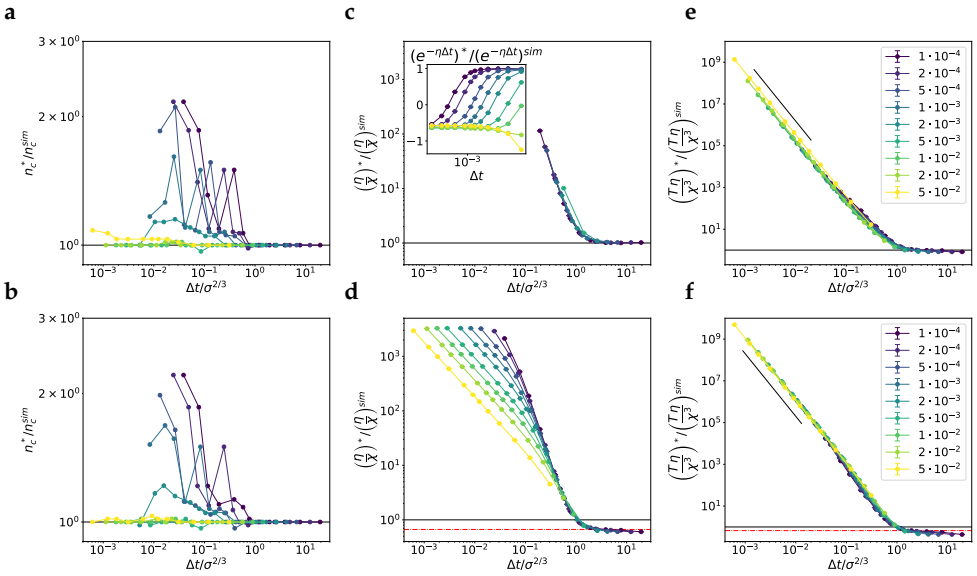

Figure 9: Effect of measurement error on some selected parameter estimators for the ISM. In the top row results from the Toeplitz inference scheme are reported; in the bottom row results from the BBK inference scheme are reported. The rescaling of the inverse sampling rate in the abscissa makes the curves in **c–f** depart at the same point ($\Delta t/\sigma^{2/3} \sim 1$) from the expected value in absence of experimental errors (1 for the Toeplitz method, $2/3$, marked by the red dot-dashed line, for the BBK method). The collapse of the curves shown in **e–f** proves that the control parameter is the ratio between stochastic and experimental noise: $T\eta\Delta t^3/\sigma^2$. The black lines, having a slope -3, are a guide for the eye. We notice that for large noise-to-signal ratio the estimate of η with the Toeplitz method may be problematic since estimators of positive definite quantities built with noisy data can become negative, as visible in the inset of **c**. Errorbars on Figs. **a–b** are not shown, for sake of clarity, whereas in Figs. **c–f** the 1.96 SD errorbar is smaller than the marker size.

The simplest (still realistic, in many practical cases) way to model experimental errors is through a superposition of the discretely sampled trajectory with a sequence of i.i.d. Gaussian random variables $\mathcal{N}(0, \sigma^2)$. As pointed out by several authors, even when σ^2 is very small, measurement noise can impact dynamical inference [147, 117, 20]. A large modification of the high-frequency region of the power spectrum of reconstructed velocities is introduced in the presence of noise [147], which in turns results in a diverging bias in parameter estimation as $\Delta t \to 0$ [117]. This bias and its trend with Δt appear also in our inference method (see Fig. 9).

The ML inference scheme fully extracts information from the increments of the measured degree of freedom, Δx, whose average absolute value grows as Δt, and needs to be compared with the amplitude of measurement errors σ, which we assume to be independent of the sampling rate ($\sigma \sim \Delta t^0$). At very high sampling rates experimental errors will dominate over the effective dynamics, resulting into an artificial trend $\sim \Delta t^{-1}$ for the parameter η, and $\sim \Delta t^{-2}$ for the effective temperature and pulsation of the harmonic oscillator.

corresponds to including an increasing number of harmonics; finite size corrections to parameters estimators can be seen as a counterpart of the Gibbs phenomenon.

Since noise cannot be ignored, let us formally include it to consider a hidden (non) Markov model. Suppose we measure discrete data-points $(\hat{x}_0, \hat{x}_1, \dots, \hat{x}_L)$ corresponding to the noisy reproduction of trajectory points $(x_0, x_1, \dots, x_L)$. Following a maximum likelihood argument, the parameters $\boldsymbol{\theta}$ of the dynamical hidden model are estimated as:

$$\boldsymbol{\theta}^*_H = \arg\max_{\boldsymbol{\theta}} P(\boldsymbol{\theta}|\hat{x}_{0:L}) = \arg\max_{\boldsymbol{\theta}} P(\hat{x}_{0:L}|\boldsymbol{\theta}), \tag{121}$$

where

$$P(\hat{x}_{0:L}|\boldsymbol{\theta}) = \int dx_0 \dots dx_L P(x_{0:L}|\boldsymbol{\theta}) \cdot \prod_{n=0}^{L} P(\hat{x}_n|x_n). \tag{122}$$

We assume

$$P(\hat{x}_n|x_n) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(\hat{x}_n - x_n)^2}{2\sigma^2}; \tag{123}$$

$P(x_{0:L}|\boldsymbol{\theta})$ is determined by the hypothesized dynamical model. At variance with the noiseless measurement case, we cannot replace the path probability with the conditional distribution $P(x_2, \dots, x_L|x_0, x_1; \boldsymbol{\theta})$, as x_0 and x_1 are not known. Using the Toeplitz approximation to remove the conditioning on v_0 in favor of a conditioning on x_1 is now of no use to find an analytical solution of the problem. As long as we deal with linear models, as in the interacting and non-interacting cases considered above, $P(\hat{x}_{0:L}|\boldsymbol{\theta})$ reduces to Gaussian integrals and the marginalization over the hidden variables can be performed explicitly [106]. A full treatment at any noise-to-signal ratio is then possible, but not easily generalizable beyond the harmonic case.

For this reason here we limit ourselves to showing the predicted effect of experimental uncorrelated noise on numerical simulations. Explicit rewriting of the likelihood in Eq. (122) allows us to identify the combination of parameters that control the transition from the small to large noise regime. If $T\eta\Delta t^3/\sigma^2 \ll 1$, noise dominates and, to lowest order,

$$P(\hat{x}_{0:L}|\boldsymbol{\theta}) \simeq \prod_{n=0}^{L} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\hat{x}_n^2}. \tag{124}$$

If $T\eta\Delta t^3/\sigma^2 \gg 1$, the effect of noise will be small, and the likelihood will converge to the one we found in absence of experimental errors.

We conclude, in agreement with Ref. [20, 117], that whenever the experimental apparatus and the observed process are such that the chain of conditions $\sigma^2 \ll T\eta\Delta t^3 \ll 1$ holds, the developed inference strategy still provides a reliable methodology to infer the parameters of the dynamics. When that condition is not fulfilled, controlled denoising procedures or inference strategies based on hidden modelling must be employed.

3.4 GENERAL RESULTS ON THE DISCRETIZATION OF GAUSSIAN PROCESSES

3.4.1 A Renormalization Group approach to link discrete and continuous-time descriptions of Gaussian processes

The Euler-Maruyama scheme: different performance for different use

Choosing a *good* discretization scheme for a continuous-time stochastic process is key to find consistent estimators in parametric inference procedures. The discussion and the results presented in Section 3.3 provide a strong argument for this statement. In particular, they show how the discretization issue cannot be overlooked when the inference task is challenged by partial observation of the state of the system. Partial observation of noise-driven systems breaks indeed the Markovian nature of the process, when observed and unobserved degrees of freedom are coupled, bringing in temporal noise correlations and memory effects [200, 135, 97].

The example of the damped Langevin equation we focused on, in which only the positional coordinates (but not momenta) are directly measurable, is a paradigmatic one. We have shown in section 3.3 that in this case maximum likelihood methods based on a naive use of the Euler-Maruyama discretization do not provide consistent parameter estimators, even though the same rule can be successfully used to numerically integrate the SDE. The inconsistency of Euler-based inference approaches had been already noticed in the literature [147, 116, 117] and rigorously proven [90, 162, 66, 54, 154]. However, it is also well-known that the Euler integration scheme is strongly convergent in the limit of vanishing increments and it can be successfully used for numerical integration [153]. A natural question is then: why is the behavior of the same scheme so different for the two tasks?

The condition to impose for a successful application of this approximate scheme in numerical integration is that the simulation time step, τ_{sim}, is sufficiently small compared to the observation scale, τ_{obs}. While this separation of scales is possible in numerical simulations, state space inference formally requires to discretize the process over time scales τ_{obs}. The lack of such separation of scales explains why the Euler discretization performs so badly in parametric inference tasks, compared to numerical integration.

A possible strategy to recover the separation of scales, and hence improve the performance of inference schemes, is to introduce and marginalize additional intermediate states between pairs of observed points. Many Markov Chain Monte Carlo methods exploiting this idea have been proposed and applied in several contexts, for inferring both fully and partially observed stochastic processes [71, 72]. This augmentation operation amounts to temporal coarse-graining and is reminiscent of the Real Space Renormalizaton Group (RG), with our time playing the role of space in usual RG, and the observed time series playing the role of a microstate for a standard condensed matter system. In the following, we derive this analogy formally for the class of stationary Gaussian processes, for which an exact RG map can be explicitly derived. We use this analogy to identify *good* discretizations of stochastic

processes — i.e. working for both simulation and inference tasks — with the fixed points of such RG map. RG fixed points satisfy indeed the property of invariance under coarse graining, a feature that consistent discretizations of continuous-time models must have. This invariance property amounts to requiring that, when the discretization interval τ is small enough, integrating the process with a given integrator over a single step 2τ or through a combination of two steps on intervals τ should provide the same result.

Analogy with the Renormalization Group

In order to better explain this formal analogy, let us consider again the damped harmonic oscillator in a heat bath:

$$dx = vdt, \qquad dv = -\eta v dt - \kappa x dt + \sigma dW, \tag{125}$$

with $W(t)$ a Wiener process. We assume that the coordinate x is observed at a finite sampling rate τ^{-1}, producing an infinite time series denoted as $\{X_n, n \in \mathbb{N}\}$. Our goal is to find a suitable model to describe the discrete stochastic dynamics which results from this partial observation of the continuous process. For stationary linear processes, such as Eq. (125), an exact solution can be computed and used in inference and simulation problems [106, 86]. However, for more general processes exact solutions are typically not available, and one must resort to Taylor-Itô expansions of the integrated SDE. We start by applying the lowest order nontrivial expansion to Eq. (125). Eliminating v, we obtain:

$$X_n = \psi X_{n-1} + \theta X_{n-2} + \mu \epsilon_n, \tag{126}$$

where $\psi = 2 - \eta\tau - \kappa\tau^2$, $\theta = (-1 + \eta\tau)$, $\mu = \sigma\tau^{3/2}$ and $\epsilon_n \sim \mathcal{N}(0, 1)$.

This process is also known as autoregressive model of order two, AR(2) [19], and is fully characterized by the conditional probability $P(X_n|X_{n-1}, X_{n-2})$. Its Markov structure allows for an easy derivation of the probability of the time series:

$$P(\{X_n, n \in \mathbb{N}\}) = \prod_{n \geq 2} P(X_n|X_{n-1}, X_{n-2}) P(X_0, X_1). \tag{127}$$

Under stationary assumptions[11], we can neglect boundary terms and interpret $P(\{X_n\})$ as the Boltzmann weight of a configuration of continuous variables on a one-dimensional lattice with first- and second-nearest-neighbor interactions (corresponding to the zig-zag ladder topology – see Fig. 10). At this stage, the main idea is to view the approach to continuum as a progressive increase in the number of steps contained within a fixed time window. A possible way to reach the continuum limit is by iterating the two operations that make up the Renormalization Group: (i) coarse graining and (ii) joint rescaling of the time unit and of the parameters of the model.

Explicit RG transformations for ARMA(p, q) time series models

We adopt the strategy of decimation to coarse grain [105, 144, 17], as sketched in Fig. 10. The goal is to get rid of half of the sites (e.g. odd ones) in the sequence

11 Here the initial condition is moved arbitrarily far in the past to ensure stationarity. In contrast to Section 3.3.5, we do not have the problem of broken shift invariance.

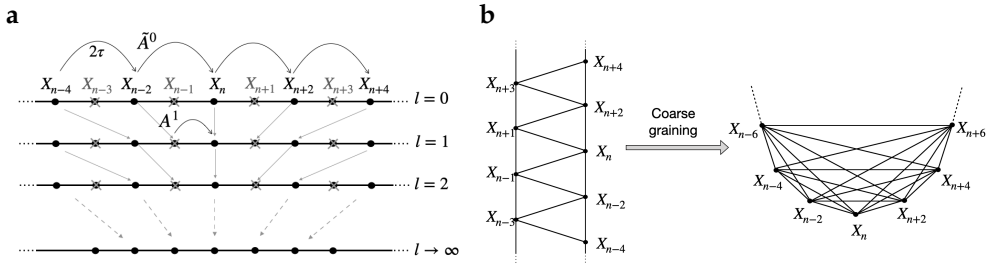

Figure 10: **a**: Sketch of the RG procedure for ARMA processes: the discrete times series is firstly decimated and then rescaled, so that the time step is always kept equal to τ. **b**: In the case of second order processes, the Euler starting point — AR(2) time series — corresponds to a spin chain model with nearest-neighbor and next-nearest-neighbor couplings, which can also be represented as a zig-zag ladder. The ladder structure is not invariant under real space RG transformations. The decimation procedure sketched in (a) produces, since the first iteration, infinite range couplings. However, these couplings are not independent, and the 4 parameters of the ARMA(2,1) model $\phi_1 = \psi$, $\phi_2 = \theta$, μ and $\nu_1 = \nu$ are sufficient to characterize them.

generated by Eq. (126), deriving effective update equations for a sub-series of states having only indexes of the same parity. In order to implement this transformation, we take a suitable linear combination of neighboring update equations of the form of Eq. (126),

$$\mathrm{Eqn}(X_n) + \psi \mathrm{Eqn}(X_{n-1}) - \theta \mathrm{Eqn}(X_{n-2}), \tag{128}$$

which results into an update equation:

$$X_n = \tilde{\psi} X_{n-2} + \tilde{\theta} X_{n-4} + \tilde{r}_n, \tag{129}$$

where $\tilde{\psi} = \psi^2 + 2\theta$ and $\tilde{\theta} = -\psi^2$.

Unlike in the original process, the random increments $\tilde{r}_n = \mu[\epsilon_n + \psi\epsilon_{n-1} - \theta\epsilon_{n-2}]$ are now correlated across nearest neighbors, $\mathbb{E}[\tilde{r}_n \tilde{r}_{n\pm2}] \neq 0$. However, we can exploit the fact that linear combinations of Gaussian variables are Gaussian to decompose them in the following way:

$$\tilde{r}_n = \tilde{\mu}\tilde{\epsilon}_n + \tilde{\nu}\tilde{\epsilon}_{n-2}, \tag{130}$$

with $\tilde{\epsilon}_i \sim \mathcal{N}(0,1)$ new I.I.D. variables, and $\tilde{\mu}$, $\tilde{\nu}$ satisfying:

$$\mathbb{E}\left[\tilde{r}_n^2\right] = \tilde{\mu}^2 + \tilde{\nu}^2 = (1 + \psi^2 + \theta^2)\mu^2; \tag{131}$$

$$\mathbb{E}[\tilde{r}_n \tilde{r}_{n\pm2}] = \tilde{\mu}\tilde{\nu} = -\theta\mu^2. \tag{132}$$

It can be easily verified that $\mathbb{E}\left[\tilde{r}_n^l \tilde{r}_{n\pm2k}^l\right] = 0$ for $k > 1$. Then Eq. (129) becomes an ARMA(2,1) model [19]:

$$X_n = \tilde{\psi} X_{n-2} + \tilde{\theta} X_{n-4} + \tilde{\nu}\tilde{\epsilon}_{n-2} + \tilde{\mu}\tilde{\epsilon}_n. \tag{133}$$

Autoregressive moving-average processes of order (p, q), denoted ARMA(p, q), are time series generated by update equations of the form

$$X_n = \sum_{i=1}^{p} \phi_i X_{n-i} + \sum_{i=1}^{q} \nu_i \epsilon_{n-i} + \mu\epsilon_n, \tag{134}$$

with $\epsilon_n \sim \mathcal{N}(0,1)$ I.I.D., $\nu_i, \mu \in \mathbb{R}$. The autoregressive (AR) part of the equation corresponds to the contribution from the previous p states of the system; the moving average (MA) part, of order q, corresponds to the second sum in the RHS of Eq. (134) and is responsible for the time correlation of random increments.

It is evident that, after coarse-graining, the process in Eq. (129) is no longer Markovian, as its probability distribution cannot be factorized into the product of single-step transition probabilities. Recalling the analogy with the spin chain, RG creates, since the first iteration, infinite-range effective couplings, starting with just first- and second-nearest-neighbor bonds. This is indeed the effect of simple decimation on the zig-zag ladder topology (Fig. 10.b).

However, these emerging couplings are not independent, and the 4 parameters in Eq. (133) are sufficient to characterize them. Let us call $\mathcal{A} = (\psi, \theta, \mu, \nu)$ the set of parameters of the ARMA(2,1) model, and $\tilde{\mathcal{A}} = (\tilde{\psi}, \tilde{\theta}, \tilde{\mu}, \tilde{\nu})$ the parameters of the coarse grained model, prior to rescaling. In order to get a closed-form RG transformation, we need to ensure that further iterations of the decimation procedure do not keep introducing novel higher order terms. Luckily, the ARMA(2,1) structure is stable, as an example of a more general result that we discuss in the following ('Memory selection' paragraph). Hence, applying RG to the class of ARMA(2,1) models yields a well-defined map from $\mathcal{A}$ to $\tilde{\mathcal{A}}$.

The second operation that completes the RG iteration is rescaling the time step, $2\tau \to \tau$, and reabsorbing this change of units through a redefinition of the parameters. The parameters of our models are dimensionless, yet their dependency on τ is what determines how to connect any discrete-time process to its continuous-time counterpart. We express ψ, θ as asymptotic power series of τ, $\psi(\tau) = \sum_k \psi_k \tau^k$ and $\theta(\tau) = \sum_k \theta_k \tau^k$, and work, up to the desired order, with recursive relations for the coefficients of the series expansion, ψ_k and θ_k. These coefficients are dimensional and get rescaled with the time unit. The same idea can be applied to ν and μ, except for expanding them in powers of $\tau^{1/2}$. It is therefore convenient to reparametrize the noise amplitudes using:

$$\alpha = \mathbb{E}[r_n r_n] = \mu^2 + \nu^2 \quad \text{and} \quad \beta = \mathbb{E}[r_n r_{n\pm1}] = \mu\nu, \tag{135}$$

as their asymptotic series expansions involve integer powers of τ: $\alpha(\tau) = \sum_k \alpha_k \tau^k$, $\beta(\tau) = \sum_k \beta_k \tau^k$.

The physical dimension of each coefficient $A_k \in \{\psi_k, \theta_k, \alpha_k, \beta_k\}$ is now set by the order of the corresponding term in the series expansion, and each of them gets rescaled, after coarse graining, as $A_k^{l+1} = 2^{-k} \tilde{A}_k^l$, where l is the index of the RG iteration, and $\tilde{A}_k^l$ is the result of the decimation step described before. With this expansion we define the RG map as a set of recursive equations in the infinite-dimensional space of the Taylor coefficients parametrizing the 4 functions $\psi^l(\tau)$, $\theta^l(\tau)$, $\alpha^l(\tau)$, $\beta^l(\tau)$:

$$\psi_k^{l+1} = 2^{-k} \left[2\theta_k^l + \sum_{i=0}^{k} \psi_i^l \psi_{k-i}^l \right], \tag{136}$$

$$\theta_k^{l+1} = -2^{-k} \sum_{i=0}^{k} \theta_i^l \theta_{k-i}^l, \tag{137}$$

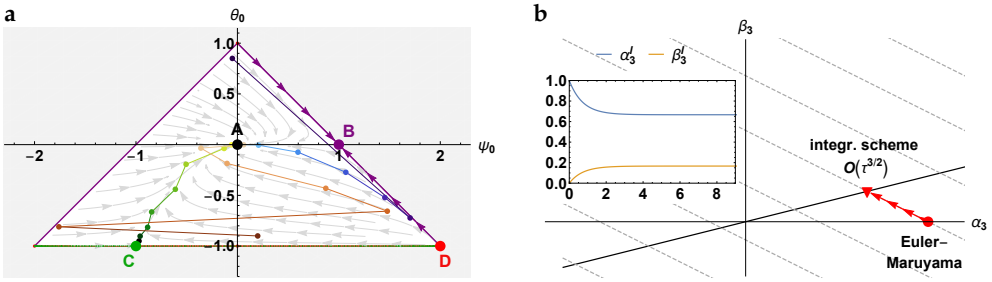

Figure 11: **a**: Fixed points of the RG map, projected on the plane of leading autoregressive coefficients ψ_0, θ_0. The interior of the triangle is the basin of attraction of fixed point A. The equal sides are the basin of attraction of point B. The basins of points C and D are contained in the basis of the triangle (plus vertex on the top for D). The Euler AR(2) process has coordinates (2,-1) in this plane. Shaded trajectories represent the solution of recurrence relations from three sample initial conditions. Arrows show the direction of (discrete) moves and should not be interpreted as continuous flow lines. All the points in the exterior of the triangle escape towards diverging fixed points. **b**: RG flow on the plane of third order MA coefficients for $s = 0$. The black solid line is the manifold of fixed points $\alpha_3^* = 4\beta_3^*$. The dashed parallel lines represent different orbits associated to the solution (149) of the linear system. A Euler-Maruyama discretization of a 2nd-order SDE corresponds to an initial condition $(\sigma^2, 0)$ on the positive α_3 axis, which flows towards the ARMA(2,1) fixed point of coordinates $(2/3, 1/6)\sigma^2$ with an l-dependent decay shown in the inset. When $s \neq 0$ the picture above is just slightly modified, as the RG map becomes an affine one.

$$\alpha_k^{l+1} = 2^{-k}\left\{(1+\psi_0^2+\theta_0^2)\alpha_k^l + 2\beta_k^l\psi_0(1-\theta_0) + \sum_{i=0}^{k-1}\left[\alpha_i^l\sum_{j=0}^{k-i}\left(\psi_j\psi_{k-i-j}+\theta_j\theta_{k-i-j}\right) + 2\beta_i\psi_{k-i}(1-\theta_0) - 2\beta_i\sum_{j=0}^{k-i-1}\psi_j\theta_{k-i-j}\right]\right\}, \tag{138}$$

$$\beta_k^{l+1} = 2^{-k}\left\{\beta_k^l\psi_0(1-\theta_0) - \alpha_k^l\theta_0 - \sum_{i=0}^{k-1}\alpha_i^l\theta_{k-i} + \sum_{i=0}^{k-1}\beta_i^l\left[\psi_{k-i}(1-\theta_0) - \sum_{j=0}^{k-i-1}\psi_j\theta_{k-i-j}\right]\right\}. \tag{139}$$

Notice that, since the recursion equations at order k only involve lower orders, they can be solved recursively over k, and can also be truncated to an arbitrary order while retaining a closed form.

Fixed points

Solving the fixed point equations associated to Eqs. (136)–(139), we find 4 different types of linear processes, parametrized by the Taylor coefficients $\{\psi_k^*, \theta_k^*, \alpha_k^*, \beta_k^*\}$. Results are reported in Table 1 up to order $k = 3$. The four classes of fixed points correspond to four manifolds in the space of all possible ARMA(2,1) models, whose projection onto the plane of leading order AR coefficients, ψ_0 and θ_0, is shown in Fig. 11.

Table 1: Fixed point solutions of the RG recurrence relations up to third order in τ. We find 4 manifolds of fixed points, corresponding to 4 types of processes, parametrized by the arbitrary constants u, s, z and b. In addition to the reported ones, there are diverging fixed points.

			A. MA(0)	B. AR(1)	C. ARMA(2, 1)	D. ARMA(2, 1)
AR coefficients	$k=0$	ψ_0^*	0	1	-1	2
		θ_0^*	0	0	-1	-1
	$k=1$	ψ_1^*	0	u	u	u
		θ_1^*	0	0	$2u$	$-u$
	$k=2$	ψ_2^*	0	$u^2/2$	$-u^2/2$	z
		θ_2^*	0	0	$-2u^2$	$-u^2/2$
	$k=3$	ψ_3^*	0	$u^3/6$	$u^3/6$	$u(6z-u^2)/12$
		θ_3^*	0	0	$(2u)^3/6$	$-u^3/6$
MA coefficients	$k=0$	α_0^*	s	0	0	0
		β_0^*	0	0	0	0
	$k=1$	α_1^*	0	s	$4s$	$-2s$
		β_1^*	0	0	s	s
	$k=2$	α_2^*	0	us	$-8us$	$-2us$
		β_2^*	0	0	$-2us$	us
	$k=3$	α_3^*	0	$2u^2s/3$	$32u^2s/3$	$4b-(2z+3u^2)s$
		β_3^*	0	0	$13u^2s/6$	b

The first class of fixed points A corresponds to sequences of independent random variables ($\psi^* = \theta^* = \nu^* = 0$). Fixed points B are AR(1) processes: they can be interpreted as discretizations of fully observed Markov processes, described by linear first-order SDEs in continuous time:

$$dx = uxdt + \sqrt{s}dW. \tag{140}$$

By induction, it is possible to show that, if $\theta_0^* = \theta_1^* = 0$, then $\theta_k^* = 0\ \forall k$, and that a solution of the recurrence relation (136) is $\psi_k^* = \frac{1}{k!}\left(\psi_1^*\right)^k$. Renaming $\psi_1^* = u$, the sum of the series for the first AR parameter reads $\psi^*(\tau) = e^{u\tau}$, which corresponds to the exact discretization of a first order linear process, whose deterministic part reads $\dot{x} = ux$. Concerning the MA coefficients, it is possible to show that, selecting the fixed point deterministic coefficients of B, $\beta_k^* = 0\ \forall k$ and $\alpha^*(\tau) = \frac{\alpha_1^*}{2u}\left(e^{2u\tau} - 1\right)$.

In addition, there are two other fixed-point manifolds corresponding to ARMA(2,1) models, denoted by C and D. Model C is not a continuous process, but evolves through finite jumps. This fixed point satisfies $\psi_k^* = -\frac{(-u)^k}{k!}$ and $\theta_k^* = -\frac{(-2u)^k}{k!}$, implying $\psi(\tau) = -e^{-u\tau}$ and $\theta(\tau) = -e^{-2u\tau}$. The process has a three-branched phase diagram, and evolves in time jumping from one branch to the next one — approaching the origin or moving away from it depending on the sign of $\psi_1^* = u$, as shown in Fig. 12. This structure is invariant under the RG transformation we defined, but it would not be so under a different one. For instance, if the coarse graining procedure was implemented by trimming two points out of three, it would result into a single-branched phase diagram in Fig. 12.

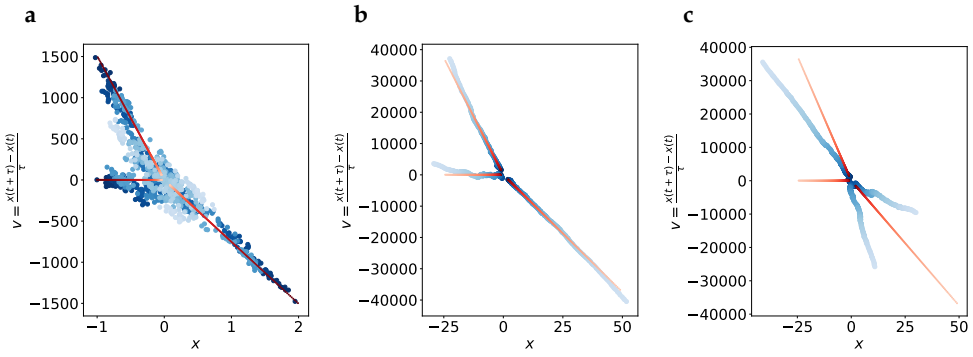

Figure 12: Sample trajectories in the phase diagram of the fixed point process C. Red points are obtained iterating the deterministic process $X_n = \psi X_{n-1} + \theta X_{n-2}$, with $\psi(\tau) = -e^{-\psi_1 \tau}$ and $\theta(\tau) = -e^{-2\psi_1 \tau}$. Blue points correspond to the stochastic process of fixed point C (cfr. Table 1. Color code from dark to light indicates advancing time. The deterministic model exhibits a three-branched phase diagram: at each iteration the system jumps from one branch to the next one, moving towards the origin or far from it depending on the modulus of the complex eigenvalues. **a**: When $\psi_1 > 0$ the origin is an asymptotically stable fixed point ($\psi_1 = 2$, $s = 0.5$, $\tau = 0.002$). **b-c**: The origin becomes unstable for negative ψ_1. We plot two samples with the same parameter values $\psi_1 = -2$, $s = 2$, $\tau = 0.002$. The presence of noise in the unstable case may give rise to very different phase diagrams from the deterministic one, if a large stochastic deviation occurs at small times.

We discuss models of the D class in the next paragraph.

D class: Second order continuous-time models

Fixed point D represents the discretization of a partially observed two-dimensional Markov process of the form:

$$dx = vdt + \sigma_x dW_x, \tag{141}$$

$$dv = -\eta v dt - \kappa x dt + \sigma dW_v, \tag{142}$$

of which (125) is a particular case. The (ψ, θ) coefficients of D reconstruct a second derivative at leading order, and the effect of linear drift at first order (with $u = -\eta$ in Table 1). The variable $s = -\sigma_x^2$ in Table 1 encodes noise added to the x variable, and it is zero in inertial models like Eq. (125).

The parametrization of the partially observed second order process is not unique; in general we can write the set of SDEs in vector form as:

$$d\mathbf{y} = \mathbf{A}\mathbf{y}dt + \mathbf{B}d\mathbf{W}, \tag{143}$$

where

$$\mathbf{y} = (x, v)^\top, \quad \mathbf{W} = (W_x, W_v)^\top, \quad \mathbf{A} = \begin{pmatrix} -\lambda & 1 \\ -\kappa & -\eta \end{pmatrix}, \quad \mathbf{B}\mathbf{B}^\top = \begin{pmatrix} \sigma_{xx}^2 & \sigma_{xv}^2 \\ \sigma_{vx}^2 & \sigma_{vv}^2 \end{pmatrix}. \tag{144}$$

Setting $A_{12} = 1$ does not imply a loss of generality: compared to the arbitrary choice of $A_{12} = a \in \mathbb{R} \setminus \{0\}$, it just corresponds to a rescaling of the time unit, which does not alter the process (the only caveat is that $a < 0$ would revert the time direction).

The case $A_{12} = 0$ is not of interest for us, as it would decouple the dynamics of the unobserved degrees of freedom from that of the observed ones. Assuming that time evolves in the positive direction, we require that $\mathbf{A}$ is negative semidefinite to ensure stability.

Partial observation of the process in Eq. (143) results into a Gaussian process described, at the continuous level, by a Generalized Langevin Equation (GLE), and, at the discrete level, by an ARMA(2,1) model. Thanks to linearity, Eq. (143) is exactly integrable, and its discrete-time description is readily obtained:

$$X_n - \frac{\left(e^{2\mathbf{A}\tau}\right)_{12}}{\left(e^{\mathbf{A}\tau}\right)_{12}} X_{n-1} - \left[\left(e^{2\mathbf{A}\tau}\right)_{11} - \frac{\left(e^{2\mathbf{A}\tau}\right)_{12}}{\left(e^{\mathbf{A}\tau}\right)_{12}}\left(e^{\mathbf{A}\tau}\right)_{11}\right] X_{n-2} = r_n, \tag{145}$$

$$\begin{aligned} r_n = & \int_{t_n-2\tau}^{t_n} \left(e^{\mathbf{A}(2\tau-s)}\right)_{11} [B_{11}dW_x(s) + B_{12}dW_v(s)] \\ & - \frac{\left(e^{2\mathbf{A}\tau}\right)_{12}}{\left(e^{\mathbf{A}\tau}\right)_{12}} \int_{t_n-2\tau}^{t_n-\tau} \left(e^{\mathbf{A}(\tau-s)}\right)_{11} [B_{11}dW_x(s) + B_{12}dW_v(s)] \\ & + \int_{t_n-2\tau}^{t_n} \left(e^{\mathbf{A}(2\tau-s)}\right)_{12} [B_{21}dW_x(s) + B_{22}dW_v(s)] \\ & - \frac{\left(e^{2\mathbf{A}\tau}\right)_{12}}{\left(e^{\mathbf{A}\tau}\right)_{12}} \int_{t_n-2\tau}^{t_n-\tau} \left(e^{\mathbf{A}(\tau-s)}\right)_{12} [B_{21}dW_x(s) + B_{22}dW_v(s)] . \end{aligned} \tag{146}$$

It is now possible to find a noninvertible[12] map between $(\mathbf{A}, \mathbf{B}\mathbf{B}^\top)$ and the Taylor coefficients of $(\psi, \theta, \alpha, \beta)$:

$$u = -(\lambda + \eta), \quad z = -\kappa + \frac{1}{2}(\eta^2 + \lambda^2), \tag{147}$$

$$s = -\sigma_{xx}^2, \quad b = \frac{1}{6}\left[\sigma_{vv}^2 + 2\eta\sigma_{xv}^2 + \sigma_{xx}^2(2\kappa - 3\eta^2 - 6\eta\lambda - 4\lambda^2)\right]. \tag{148}$$

A subclass of models contained in fixed point D is represented by inertial processes, obtained by further setting $\sigma_{xx}^2 = 0$. This condition implies that there are no stochastic contributions to the observed process of order $O(\tau^{1/2})$ — the order of convergence of the Euler scheme. As one can read from Table 1 by setting $s = 0$, leading contributions to noise covariances are in this case determined by third-order coefficients. If we start with the Euler discretization (126) as an initial condition for the RG recurrence relations ($\alpha_k^0 = \sigma^2\delta_{k,3}, \beta_k^0 = 0$), the associated RG flow, on the plane of noise parameters, reads:

$$\alpha_3^l = \sigma^2\left(\frac{2}{3} + \frac{1}{3}4^{-l}\right); \quad \beta_3^l = \sigma^2\left(\frac{1}{6} - \frac{1}{6}4^{-l}\right). \tag{149}$$

While the initial model (Euler) is strongly convergent as $\tau^{1/2}$, the resulting fixed point is at least convergent as $\tau^{3/2}$. Indeed, the asymptotic values (α_3^*, β_3^*) are those we would obtain if we applied a discretization scheme like that of the Langevin impulse integrator [173] to Eq. (125) in the first place.

12 Because of the partial nature of the observation, multiple models are mapped to the same ARMA process. Hence there is no bijection between the continuous-time Markovian description of the dynamical system and its experimental non-Markovian observation. Extracting the parameters of an underlying continuous-time Markov model from time series analysis is therefore an arbitrary task, which postulates the choice of suitable hidden variables. If we set $\lambda = 0$, $\sigma_{xv}^2 = 0$, we retrieve Eqs.(141),(142).

Memory selection

The RG procedure we detailed for ARMA(2,1) models can be generalized to arbitrary ARMA(p, q) processes, defined by:

$$X_n = \sum_{i=1}^{p} \phi_i X_{n-i} + \sum_{i=1}^{q} \nu_i \epsilon_{n-i} + \mu \epsilon_n, \tag{150}$$

with $\epsilon_n \sim \mathcal{N}(0,1)$. Decimation of the time series can be done by combining p neighboring equations (150), in a similar way as in Eq. (128):

$$\mathrm{Eqn}(X_n) + \sum_{i=1}^{p} (-1)^{i+1} \phi_i \mathrm{Eqn}(X_{n-i}). \tag{151}$$

This decimation prescription yields new update equations with an ARMA(p,$\tilde{q}$) structure, with

$$\tilde{q} = \left\lfloor \frac{p+q}{2} \right\rfloor, \tag{152}$$

where $\lfloor x \rfloor$ denotes the rounding down operation (details in Appendix B). The condition of invariance under RG imposes that any fixed point satisfies $\tilde{q} = q$, implying $q = p$ or $q = p - 1$.

This result leads to two important observations. Firstly, it shows that purely autoregressive models — AR(p)$\equiv$ ARMA($p, 0$) — of order $p \geq 2$, cannot be stable points, and thus cannot be exact discretizations of stochastic differential equations of second or higher order. Secondly, the sharp selection of q reveals that longer memory than p steps in the past is irrelevant in the RG sense.

Our result is related to the non-existence of exact delay vector embeddings for noise-driven systems. Delay vector embeddings are equivalent to estimating the derivatives of the observed stochastic process through differences of subsequent measurements. Precisely, they consist of stacking a finite number of subsequent points to define a new dynamical variable $\mathcal{X} = (X_{n-p+1}, \ldots, X_n)$ — known as delay vector [183, 28] — and assuming that it follows a Markov dynamics. This dynamics is described by AR(p) models.

Although this derivative reconstruction method is standard in deterministic contexts, it cannot be directly extended to stochastic processes, as partial observation sets strong limitations to phase space reconstruction for stochastic dynamical systems [178, 179, 142]. Controlled embedding procedures need to be employed, such as in [58, 14]. Another possible suggestion, which can be drawn from our analysis, is to abandon the Markov setting in favor of descriptions with correlated noise. The RG construction shows indeed that time correlations in the noise terms are spontaneously generated to match the original dimension of the partially observed dynamical system. This fact suggests to introduce an additional noise delay for any new coordinate that is appended to the delay vector. Inference of the resulting ARMA(p, q) models and, in general, of nonlinear processes with memory, can be done using several available methods, including the one we introduced in Section 3.3. How to extend the RG formalism to nonlinear processes remains, on the contrary, an open question.

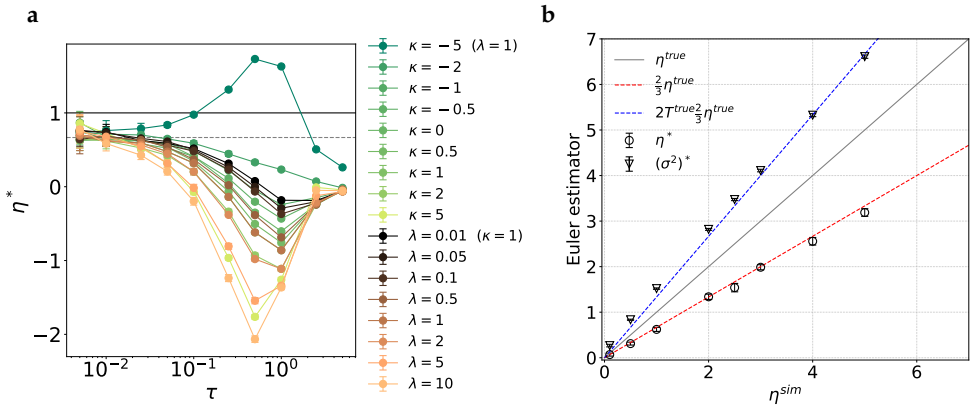

Figure 13: **a** Inconsistent ML estimators (derived from Euler discretization) for the damping coefficient η of a Brownian process in a symmetric quartic potential: $V(x) = \frac{\kappa}{2}x^2 + \frac{\lambda}{4}x^4$, for different values of the simulation parameters κ and λ. **b** The estimator η^* is rescaled of a factor $2/3$ compared to the true value of the parameter used in the simulation (in the $\tau \to 0$ limit). The same rescaling affects the diffusion coefficient $\sigma^2 = 2T\eta$; hence the equilibrium temperature T is correctly estimated from their ratio. Estimators κ^* and λ^* obtained with the same Euler-based procedure are inconsistent: disentangling the contributions of the different terms in the polynomial force requires dynamical information of higher order in τ (not accessible through the Euler scheme).

3.4.2 A 'shortcut' to Markov embedding for partially observed equilibrium processes

The argument presented so far is not based on any stationarity nor equilibrium assumption for the continuous process we observe: it only relies on the requirement that the discretized model must have an invariant structure under coarse graining. Such a requirement ensures that a transition over a time interval 2τ can be equivalently described, up to a given order of approximation, either by a discretization on a single time step of amplitude 2τ or by the composition – using the same scheme – of two τ-steps. When this request is not satisfied, the discretization is inadequate, producing biased estimators when applied to inference problems [147, 89]. Nonetheless, we have seen in Section 3.3.2 that the bias comes in a regular form. Given a damped Langevin equation like

$$dx = vdt, \qquad dv = -\eta v dt - \kappa x dt + \sqrt{2T\eta}dW, \tag{153}$$

the bias consists in a τ-independent rescaling factor equal to $2/3$ for the linear damping coefficient η, whereas the remaining parameters are unbiasedly estimated. In a similar way, by applying Euler-based inference schemes to simulated time series for the Brownian particle in a double well (formulas in Appendix A), we obtain a damping coefficient η rescaled of a factor $2/3$, while the parameter estimator for the temperature of the heat bath is consistent. Results are shown in Fig. 13.

In this section we try to rationalize this finding, in order to understand how universal this rescaling is, and whether it can be exploited to build an *effective* discrete Markov model for generalized Langevin equations, which neglects noise

correlations. We try to provide an argument for this fact, working out the reference problem of the integrated Ornstein-Uhlenbeck (OU) process.

The integrated OU process, described by the Eqs. (153) with $\kappa = 0$, is the simplest example of 2nd-order SDE for which the Euler-related inconsistency appears. Suppose we can only observe (with infinite accuracy) the inertial degree of freedom, at a sampling rate τ^{-1}. Let us recall the notation for the time series of empirical observations $\{X_n\}_{n\in\mathbb{N}}$, and introduce that of reconstructed velocities $\{\bar{V}_n\}_{n\in\mathbb{N}}$, where:

$$\bar{V}_n = \frac{(X_{n+1} - X_n)}{\tau}. \tag{154}$$

Let us also introduce the series of *real* velocities $\{V_n\}_{n\in\mathbb{N}}$, corresponding to the one we would obtain if we were able to measure directly the velocity degree of freedom.

Because the evolution of the velocity degree of freedom is described by an independent first-order SDE in the OU process, the time series $\{V_n\}_{n\in\mathbb{N}}$ is described by an AR(1) model. On the contrary, the evolution of the reconstructed velocities $\bar{V}_n$ inherits a nonzero MA order from the non-Markovian dynamics of the x variables, ending in an ARMA(1, 1) process. Nonetheless, we look for an *effective* AR(2) process for $\{X_n\}_{n\in\mathbb{N}}$, i.e. an effective AR(1) process for the $\{\bar{V}_n\}$ series:

$$\bar{V}_n - (1 - \alpha)\bar{V}_{n-1} = \sigma\epsilon_n, \tag{155}$$

where $\epsilon_n \sim \mathcal{N}(0, 1)$ I.I.D. and α, σ are parameters to fix. The goal is to find a memory-less discrete model for $\{\bar{V}_n\}_{n\in\mathbb{N}}$ that reproduces correctly the sufficient statistics employed by parametric inference approaches based on the exploitation of *local* dynamical information[13].

In the case of the integrated OU process, a sufficient statistics corresponds to the set $\mathcal{S}_1 = \{\mathbb{E}[\bar{V}_n^2], \mathbb{E}[\bar{V}_n\bar{V}_{n+1}]\}$, i.e. the self-correlations of the reconstructed velocities at equal time and at a distance of one time step. A statistics is a collection of functions of the observed data which can be used to extract information from them [77]. It is *sufficient* when it contains all the information contained in the sample, and the addition of any other function does not bring more information [108].

We impose on these observables the two following *consistency* conditions:

i. $\bar{V}_n$ follows a Maxwell Boltzmann distribution with temperature T (equipartition):

$$\mathbb{E}\left[\bar{V}_n^2\right] = T \qquad (k_B = 1). \tag{156}$$

ii. The relation between the reconstructed acceleration $\bar{A}_n = (\bar{V}_{n+1} - \bar{V}_n)/\tau$ and the reconstructed velocity $\bar{V}_n$ is the one we can exactly compute for the integrated Ornstein-Uhlenbeck process:

$$\mathbb{E}[\bar{A}_n|\bar{V}_n] \underset{\tau\to 0}{\simeq} -\frac{2}{3}\eta\bar{V}_n. \tag{157}$$

13 The common characteristic of these approaches (both Bayesian and non-Bayesian) is being derived from a Taylor-Itô expansion in the small τ limit; hence they only exploit short time information to learn the laws of motion. This information is typically carried by the first few elements of the autocovariance of the time series.

A detailed derivation of Eq. (157) is in [147]. From this relation we immediately deduce the condition to impose on the observables of interest:

$$\mathbb{E}[\bar{V}_{n+1}\bar{V}_n] = \left(1 - \frac{2}{3}\eta\tau\right)\mathbb{E}\left[\bar{V}_n^2\right]. \tag{158}$$

The self-correlation function $\mathbb{E}[\bar{V}_n\bar{V}_{n+k}]$ of an AR(1) process of the form of Eq. (155) is explicitly known [19]:

$$\mathbb{E}[\bar{V}_n\bar{V}_{n+k}] = \frac{(1-\alpha)^{|k|}\sigma^2}{1-(1-\alpha)^2}. \tag{159}$$

Taking its value at $k = 1$ and using Eq. (156) in Eq. (158) yields the expected result $\alpha = (2/3)\eta\tau$. Computing the self-correlation of the reconstructed velocities at $k = 0$ and imposing equipartition, we set the value of σ:

$$\sigma^2 = T\left[1-(1-\alpha)^2\right] \underset{\tau\to 0}{\simeq} 2T\alpha. \tag{160}$$

This is the celebrated Einstein relation [109].

In conclusion, it is possible to describe the sequence of measurements of an integrated OU process as an effective AR(2) series with a rescaled damping coefficient $\eta' = (2/3)\eta$. Due to its link to the Einstein relation, we expect such effective rescaling to provide a shortcut to Markovian delay vector embeddings for any stochastic process at equilibrium [127]. Numerical and analytical evidence confirm that the resulting purely autoregressive model can be used for maximum likelihood inference in the simple case of linear Langevin equations, as discussed in Section 3.3.2. In the presence of nonlinear conservative forces, numerical results suggest that the same rescaling procedure can be effectively used to infer η and T, while the remaining parameters of the model do not match the Euler-ML estimators.

As a final remark, we underscore the starting observation about the different behaviour of the Euler scheme in inference and integration tasks. While this seemingly universal bias can be exploited to simplify parametric inference procedures based on 'local in time' sufficient statistics, the same effective rescaled white scheme cannot be used for simulations: it would obviously converge to a continuous process with the wrong parameters.

3.5 FINAL REMARKS

3.5.1 Summary

For the work presented in this Chapter, we employed standard methods from stochastic calculus. The goal was to address the problem of parametric inference for non-Markovian dynamical processes. Our original contribution consists of:

1. The development of a new inference algorithm for second order stochastic differential equations based on the maximum likelihood principle. This work was motivated by the study of starling flocks, whose dynamics seems to be ruled by effective inertial models for the velocity degrees of freedom. The

method we developed is however more general, and can be applied to a whole class of processes in a fast and efficient way.

2. The discussion of a formal analogy between data augmentation methods used in statistical inference and the renormalization group construction. Despite the analogy is obvious, it helps explaining why the naïve Euler-Maruyama discretization of a second order stochastic process can work for numerical integration but not for inference, yielding inconsistent maximum likelihood estimators. The Euler-Maruyama discretization of a second-order SDE is indeed an AR(2) process, which is not an RG fixed point. For simple Gaussian processes, modeled by ARMA(p, q) processes at the discrete level, we showed that a specific relation between p and q must hold if we want the model to be the exact, RG-invariant discretization of a continuous-time process. The result, $q = \lfloor (p+q)/2 \rfloor$, is a manifestation of the lack of finite-dimensional delay vector embeddings for partially observed SDEs, and suggests an alternative rule to embed these processes in a non-Markov way: for any new point added to the delay vector, the MA order of the discrete Gaussian process shall be increased by one, by adding a delay also in the noise variables.

3.5.2 Prospects

The natural prospect of this work is the application to real data. Applying the Toeplitz-based inference method for the ISM to the available experimental data on starling flocks is however not straightforward, as this step is hampered by several problems.

The first one is that there are missing points in the reconstructed data. Because of tracking problems of various nature, including the fact that birds may spend a fraction of time out of the field of view of the cameras, many of the birds' trajectories are cut and do not cover the whole time span of the observation. An inference method like the one we proposed, whose accuracy is only asymptotic in the length of the trajectory, might then not look as an optimal choice. However, we saw that the convergence to the true value is pretty fast. Moreover, thanks to the analysis of the finite-length bias in Section 3.3, we could deduce the rescaling factors coming from the Toeplitz shift-invariant approximation for the biased parameter estimator η^*, which can be used to apply an a posteriori correction.

An alternative strategy, which we have tried to pursue, could be to adopt a Euler-based scheme on triplets of subsequent points. Since the inertial spin model is linear in the SWA, the equations of motion for the velocity fluctuations π correspond to those of N coupled harmonic oscillators. We know, for linear processes, that the effect of modeling the system through an AR(2) process is just a $2/3$ rescaling of the estimated η parameter (Section 3.4.2). So, if we implement again an a posteriori correction, the problem can also be reframed through a maximum caliber approach that produces a second order Markov description.

A second annoying problem, which we better discuss in the next paragraph, comes from the birds' wing flapping. Before describing our attempts of data analysis, let us refer to Appendix D for a description of the dataset.

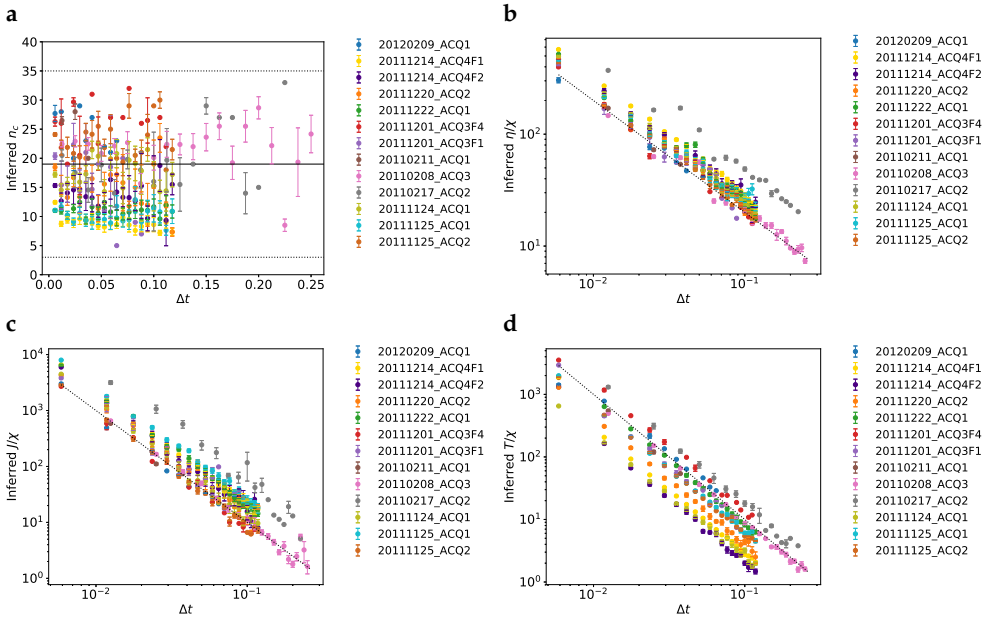

Figure 14: Parameter estimators of the ISM with k-nearest neighbor interaction ($k = n_c$) obtained applying our ML-BBK inference scheme on triplets of subsequent frames, where the velocity time series $\{\bar{\mathbf{v}}_i\}$ is computed as in Eq. (161). Each point is an average over the triplets which compose the system's trajectory; errorobars are the standard deviations of the inference results on the various triplets. **a**: The number of interacting neighbors is found through a numerical minimization of the minus-log-likelihood ($n_c \in \mathbb{N}$). The sampled interval is indicated between the dashed lines; the solid line is the average. The picture is compatible with a measurement noise-dominated estimation of the parameter: since n_c is dimensionless, there is no trend with Δt, but the distribution of the results is not far uniform sampling — see also Fig. 16. **b**: The inferred parameter η/χ scales as $\sim \Delta t^{-1}$ (dashed line). **c–d**: The inferred parameters J/χ and T/χ scale as $\sim \Delta t^{-2}$ (dashed line). Time units are 1s for Δt, 1s^{-1} for η/χ, and 1s^{-2} for J/χ and T/χ.

Preliminary data analysis

We tried to apply the BBK-ML scheme (based on Euler discretization — see App. A) to the available data, but this revealed the presence of strong measurement noise, compared to the temporal resolution of the experimental setup. The fact is evident from the comparison between Fig. 9 and Figs. 14–15. In the latter we see that the trend of the inferred parameters with Δt is purely ruled by the physical dimension of the parameters, in the same way as when uncorrelated measurement errors dominate in numerical simulations. While the inferred $(J/\chi)^*(\Delta t)$ and $(T/\chi)^*(\Delta t)$ scale as Δt^{-2}, $(\eta/\chi)^*(\Delta t)$ scales as Δt^{-1} and $n_c^*(\Delta t)$ is approximately constant. In these plots we vary Δt by selecting one every n points in the time series of reconstructed velocities. Given $\Delta t = n\tau$, with τ the minimal time interval corresponding to the resolution of the apparatus, we firstly derive the reconstructed velocity time series $\left\{\bar{\mathbf{v}}_i^t\right\}$, where:

$$\bar{\mathbf{v}}_i^t = \frac{\mathbf{x}_i^{t+\tau} - \mathbf{x}_i^t}{\tau} \tag{161}$$

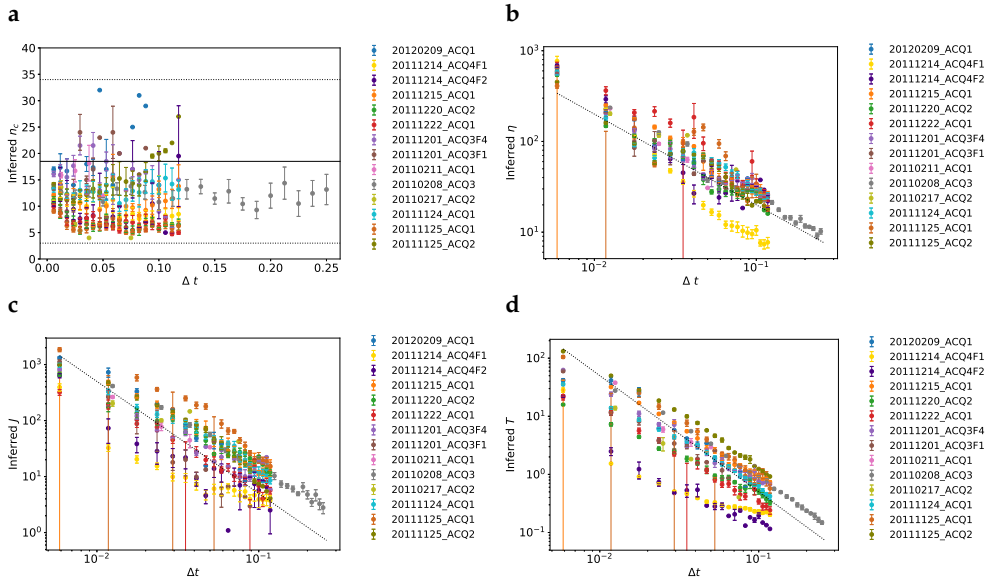

Figure 15: ISM parameter estimators obtained applying our ML-BBK inference scheme applied to triplets of subsequent frames, when the velocity time series is $\{\tilde{\mathbf{v}}_i\}$, with $\tilde{\mathbf{v}}_i$ defined as in Eq. (162). Points and their errorbars correspond to averages and standard deviations as in the previous case. Again, the trends are compatible with the dimensional scaling for all the parameters, i.e., setting $\chi = 1$: (**a**) $n_c^* \sim \Delta t^0$, (**b**) $\eta^* \sim \Delta t^{-1}$, (**c**) $J^* \sim \Delta t^{-2}$, (**d**) $T^* \sim \Delta t^{-2}$. Notice that, in this case, the inferred value of n_c^* is systematically below the uniform average of the sampling interval. Here again the time unit for Δt is 1s, while for the inferred parameters is s^{-1} or s^{-2}, according to the physical dimension of the parameter.

and then select one point out of n in order to produce an *observed* time series at a temporal resolution Δt.

We checked that even cooking the observed time series in other ways, e.g. averaging the velocities over m-point intervals, the behavior of the parameter estimators is the same, and points at a scenario where measurement noise dominates over the intrinsic dynamic noise. Precisely, we reconstructed a smoothed velocity time series $\{\tilde{\mathbf{v}}_i^t\}$, where:

$$\tilde{\mathbf{v}}_i^t = \frac{\mathbf{x}_i^{t+m} - \mathbf{x}_i^t}{m\tau}. \tag{162}$$

At the least coarse-grained level, subsequent points are separated by τ. In order to increase Δt, we select again one point out of n, having $\Delta t = n\tau$. In particular, we carefully analyzed the cases $m = 17$ (for data samples recorded with cameras shooting at 170 fps) and $m = 8$ (for data samples recorded at 80 fps). The spectra of the reconstructed trajectories exhibit indeed a peak at a frequency of 10 Hz, which corresponds to the wing flapping frequency of starlings (cfr. Fig. 18). Even employing Eq. (162) to define the velocity, the inferred parameters show a trend which is compatible with the dimensional scaling, see Fig. 15.

However, this trend is more blurred than in Fig. 14, and a significant improvement is visible in n_c^*. We notice indeed that the inferred values in the first plot of Fig. 15 are typically below the midpoint of the interval $[n_{c\min}, n_{c\max}]$, over which the log-likelihood is minimized, in contrast to Fig. 14. In Fig. 16 we compare the

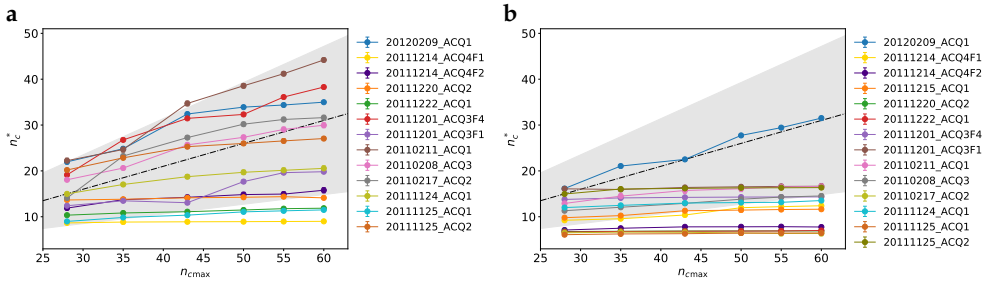

Figure 16: Inferred number of interacting neighbors n_c^* for varying $n_{c\max}$. The dot-dashed line is the midpoint of the interval $[n_{c\min}, n_{c\max}]$; the shaded area is the SD of the uniform distribution. **a** When $\mathbf{v}_i$ is computed as in Eq. (161), n_c^* grows with $n_{c\max}$, suggesting that the minimization results are not far from a uniform sampling of the interval $[n_{c\min}, n_{c\max}]$. **b** Using Eq. (162) to compute $\mathbf{v}_i$ and feeding our inference algorithm with these trajectories, n_c^* seems to be independent of $n_{c\max}$.

results of these log-likelihood minimizations with the expectation coming from a uniform sampling of the considered interval, for different values of $n_{c\max}$ and several flocking events. When the velocity time series is *not* averaged over a period, but computed according to Eq. (161), n_c^* follows an increasing trend compatible with the hypothesis of uniform sampling of the interval. On the contrary, when velocities are computed using Eq. (162), the trend is almost flat, suggesting robustness of the inferred value.

We cannot say, however, that inference works at this stage, due to the above-mentioned Δt dependency of the remaining parameters, shown in Fig. 15. The fact that the supposed dynamical model does not properly match the raw data is also suggested by the low success rate of the inference algorithm (Fig. 17), which is programmed to give a 'failure' output whenever the optimal n_c value corresponds to one of the extremes of the considered interval.

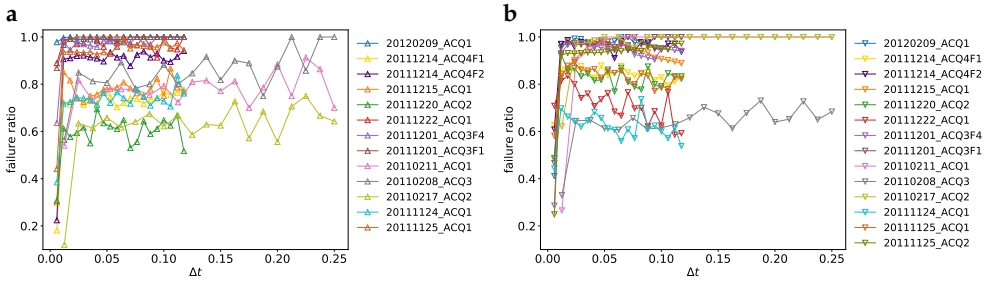

Figure 17: Failure ratio of the BBK inference algorithm on real data. The algorithm is applied to subsequent triplet of frames, for varying Δt. The failure ratio is computed as the number of triplets over which the optimal n_c^* is equal to $n_{c\min} = 3$ or $n_{c\max} = 34$, divided by the total number of frame triplets (which is reduced as Δt increases). The temporal coarse graining is implemented selecting one point out of n in the reconstructed velocity time series $\{\bar{\mathbf{v}}_i^t\}$, defined as in Eq. (161). Here $\Delta t = n\tau$, for $n = 1, \ldots, 20$. The resolution of the cameras is $\tau^{-1} = 170$ fps for all the data samples, except for the three where $\tau^{-1} = 80$ fps. **a**: the $\mathbf{v}_i$'s are defined as in Eq. (161); **b**: the $\mathbf{v}_i$'s are defined as in Eq. (162). The unit for Δt is 1s.

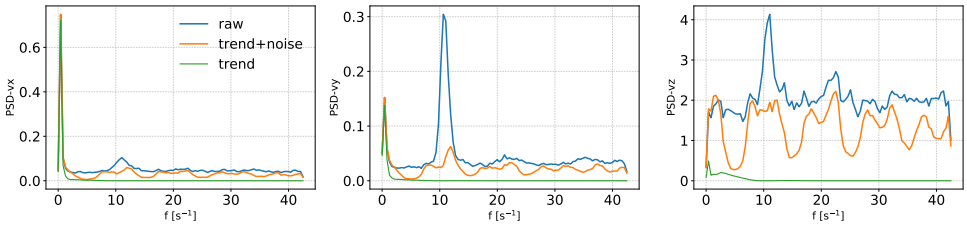

Figure 18: Average spectral density of the velocity components of the 189 birds of the data set 20111201_ACQ3F4 which have been fully recorded for 442 frames. Raw data (with $\mathbf{v}_i$ computed as in Eq. (161)) are displayed in blue. The peak at 10Hz frequency comes form the birds' wing flapping. Green and orange lines correspond to the spectra obtained by computing $\mathbf{v}_i$ from the trend or trend+residual signal resulting from the STL decomposition of each bird's spatial trajectory. The assumed period for the STL decomposition is $T_P = 0.1$s — luckily commensurable with the sampling rate τ: $T_P = 17\tau$. Spectral densities have been computed using the Welch's method.

In all the plots shown here we use the following parametrization for the connectivity matrix, already adopted in [136]:

$$n_{ij} = e^{-k_{ij}/n_c}, \tag{163}$$

where k_{ij} is the rank of the bird j among the neighbors of bird i. No significant difference in the performance of the inference approach nor in the results was observed using the standard binary parametrization: $n_{ij} = \Theta(n_c - k_{ij} + \epsilon)$.

Future strategies

What emerges from our discussion is that, at the current status, data are not processed reliably enough to apply the proposed Bayesian or maximum caliber inference schemes. These methods are very efficient and precise but extremely non robust to contamination of the data. We therefore need a reliable strategy to go from the raw data to smoothed trajectories, where at least the flapping signal is removed.

A standard tool in time series analysis is the STL algorithm (Seasonal-Trend decomposition using LOESS). Through a controlled iterative procedure, the STL algorithm decomposes a time series into three components [55]: a trend, a periodic (or *seasonal*) component, and a residual component, as schematically represented in Fig. 19. One possible strategy is to perform dynamical maximum likelihood inference on the trend component of our multidimensional signal, assuming that, in addition to the periodic flapping signal, also measurement errors are removed with the residual component. However, we expect that the residual will contain also part of the intrinsic dynamic noise, which is not possible to disentangle from other noise sources in a model-independent way.

Further systematic analysis on synthetic data is required to assess what is the effect of the STL decomposition when combined with a maximum likelihood dynamic inference scheme which exploits only local (in time) observables. This would help us determining whether the impact of this or other filters results into a con-

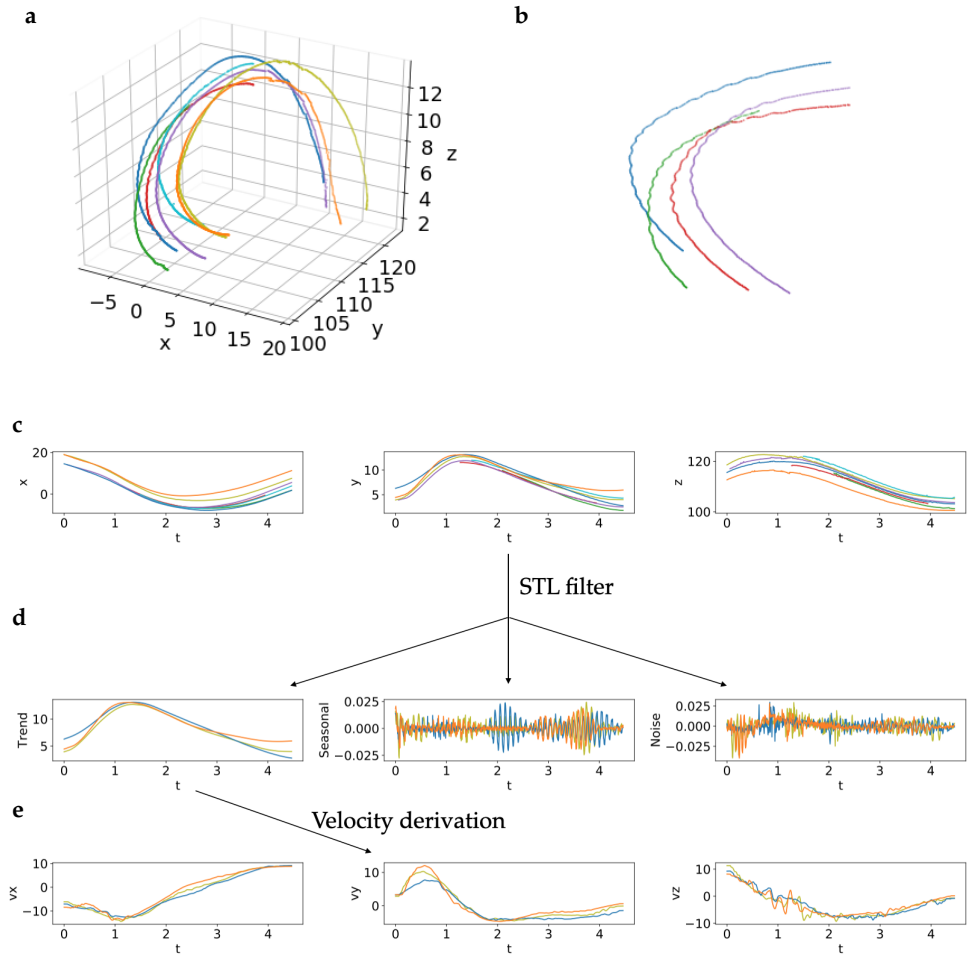

Figure 19: Schematic representation of how the STL filter is employed on a subset of bird trajectories from the data-set 20111125_ACQ2. **a–b**: Three-dimensional reconstruction of few bird trajectories, with a zoom on the right that evidences the flapping. **c**: The recorded (x and y components) and reconstructed signals (z component) are decomposed using the STL filter into three parts: trend, seasonal and noise. **d**: We show here the three components associated to the y signal. **e**: From the trend part of the signal of each component, the velocity vectors are calculated as in Eq. (161). Alternatively, they can be computed, using the same formulas, from trend+noise. The corresponding spectral densities are reported (for a different flock, where the periodic signal is more evident) in Fig. 18. In all the plots, the time unit is 1s, the distance unit is 1m, and the velocity unit is 1m/s.

trolled and predictable bias, and ultimately whether it is reasonable to infer at least some of the dynamical parameters with this strategy. Work along these lines is underway.

4 SIGNATURES OF TIME REVERSAL SYMMETRY BREAKING IN MICROSCOPIC FLOCKING MODELS

4.1 INTRODUCTION

4.1.1 Global order is a genuine non-equilibrium feature

Irreversibility is a distinguishing feature of active systems. Self-propulsion indicates indeed the ability to dissipatively consume resources to sustain a state of motion. In an extended active system made of particles with such property, detailed balance is therefore broken at the constituents' level. In contrast to many passive systems, the non-equilibrium state of an active system is not the result of an extremely slow relaxation process towards equilibrium, nor of external gradients imposed at its boundaries; we can rather conceive it as a homogeneous internal driving generated by activity, which creates irreversible currents in the system's phase space.

Irreversibility plays a crucial role in active matter, as it allows for the emergence of collective phenomena that would be impossible at equilibrium. Two major examples for scalar and polar matter are the phenomena of motility-induced phase separation and of flocking. In both cases, the role of activity can be mechanistically seen as creating effective interactions that would not be present otherwise. In the case of MIPS, it consists of an effective attraction due to the persistence of motion; in flocks, it consists of an effective long-range ferromagnetic alignment between flying XY spins produced by their motility, which violates the hypotheses of the Mermin-Wagner theorem [133].

The case of flocks is of particular interest to us. Understanding rigorously which are the crucial ingredients that allow for spontaneous ordering in Vicsek-like models is not a completely closed question. The Mermin-Wagner theorem only holds for ferromagnetic or antiferromagnetic spin systems on a regular lattice or recursive graphs [29], with short-ranged — or sufficiently fast-decaying — interactions[1]. Different arguments can be made to invalidate both the hypotheses in Vicsek-like models. Firstly, particles' motility potentially induces effective long-range inter-

1 Technical requirements for the original proof by Mermin and Wagner are the possibility to work in a discrete Fourier space and that the isotropic coupling $J(R)$ decays fast enough to guarantee that $\sum_{\mathbf{R}} R^2|J(R)|< \infty$, with $\mathbf{R}$ the vectors of a Bravais lattice. Subsequent or contemporaneous versions of the theorem also exist, both for classical and quantum spin systems. Mermin and Wagner focused on quantum systems, only highlighting in the concluding part of [133] that the theorem must extend to the classical case. An alternative proof for classical $SO(n)$-symmetric ferromagnets was later provided by McBryan and Spencer [132], who used the microscopic lattice just to take a continuous-space limit. The hypothesis that particles lie on the sites of a regular lattice can then be relaxed, but we expect the theorem not to hold in the case of a (metric) Vicsek-like model, due to giant number fluctuations, which create a strongly irregular neighbors network.

actions; secondly, the structure of the system resembles that of a liquid or a gas, rather than that of a regular crystal lattice, especially in the metric case. At this point, though, the arguments are not very rigorous. In a recent paper, Tasaki elucidated that motility alone is not able to give rise to true long range order in $d = 2$, unless it is accompanied by detailed balance violation [184]. In the presence of detailed balance, a Hohenberg-Mermin-Wagner-type theorem can be derived, even when particles are self-propelled. This result corroborates the widely shared belief that the crucial ingredient for collective motion is precisely the irreversibility of the microscopic process.

Significant progress has been made over the last decade in the attempt of characterizing the non-equilibrium steady-state of self-propelled systems, with special emphasis on the two prototypical models of ABPs and AOUPs, both with and without inter-particle interactions [63, 47, 130, 27, 56]. An approach which has often been used is to look for approximations of the steady state of the active system through equilibrium-like descriptions, for which extended fluctuation-dissipation relations (FDR) or equations of state may hold [175, 79, 88, 75, 128]. However, whether such descriptions are good or not depends on the features we want to reproduce and on the regime in which the system operates. Finding a good way to quantify the departure from equilibrium in active systems has been a hot topic in the past years, and, as it is common in non-equilibrium systems, the answer seems to be really dependent on the model under investigation and on the spatio-temporal scale at which this is observed.

Except for single particle models, most of the efforts have focused so far on the case of non-aligning scalar matter [140]. Among the preferred systems are AOUPs or ABPs with repulsive interactions, at the microscopic level [119, 78, 47, 61], or scalar active field theories which have been introduced to describe MIPS, at a coarse-grained level [137, 24]. There is a good reason for this fact: equilibrium-like state functions hold to describe the bulk of phase-separated systems, far from the interfaces [75]. Moreover the RG analysis of scalar ϕ^4 theories like Active Model B or their variants [31], typically used to describe systems of active particles undergoing a motility-induced phase separation, shows that activity is irrelevant in the RG sense and that these systems belong to the Ising universality class [25, 145], as confirmed by numerical simulations [125]. Quantifying how far the system is from an equilibrium condition, which seems to be recovered at the macroscopic scale, is then an interesting question. On the contrary, activity is relevant (in the RG sense) in polar active matter, whose field theories generally have their own dynamic universality classes, which are different from equilibrium ones [187, 38, 50]. In addition, it has been shown that, when repulsive forces are replaced with non-conservative torques that reorient the polar particles, pressure is no more a state function [175]. Although irreversibility has a key role at all scales in the polar case, relatively little has been done to characterize at the microscopic level the non-equilibrium steady state of self-propelled particles with ferromagnetic interactions. Few results, mainly employing fluctuating hydrodynamic models, have been obtained so far [15].

In this Chapter we want to adopt the formalism of stochastic thermodynamics to characterize microscopic irreversibility in a Vicsek-like model with purely aligning

active particles. We carried out a numerical analysis, which allowed us to measure the entropy production rate and identify signatures of time reversal symmetry breaking. Inertia is neglected in this context [2]. I will finally discuss our attempts of reproducing the observed features in the framework of kinetic theory.

4.1.2 Review of general formalism

In this section I will present some of the main concepts of stochastic thermodynamics, mainly following Seifert's review [168], and some results about the symmetries of NESS distributions. These results are then applied to simple active Brownian particle models for flying spins, in order to show that time reversal symmetry breaking induces an explicit asymmetry in the steady state distribution of the system's microstates.

Path integral formalism

Let us stick to the paradigm of Langevin dynamics: we consider a continuous Markov process described by the SDE:

$$\dot{x}_\alpha(t) = A_\alpha(\mathbf{x}, t) + B_{\alpha\beta}(\mathbf{x}, t)\xi_\beta(t), \tag{164}$$

with $\boldsymbol{\xi}$ an isotropic white noise. The trajectory of the degrees of freedom is denoted by $\mathbf{x}(t)$. As the process is Markovian, the weight of the trajectory, given an initial condition $\mathbf{x}(0) = \mathbf{x}_0$, is

$$p[\mathbf{x}(t)|\mathbf{x}_0] = \mathcal{N} \exp -\mathcal{A}[\mathbf{x}(t), \lambda(t)], \tag{165}$$

with $\mathcal{A}$ the Onsager-Machlup action, $\mathcal{N}$ a (diverging) normalization factor, and $\lambda(t)$ a set of possibly time-dependent control parameters, hidden in $\mathbf{A}$ and $\mathbf{B}$ in Eq. (164). To lighten the notation, we will take a single scalar control parameter. The explicit expression of the Onsager-Machlup action depends on the adopted Itô or Stratonovich convention. In the following we will focus on the case of additive processes where the $\mathbf{B}$ matrix is independent of $\mathbf{x}$ and of time. For such processes, the weight of the trajectory between the time points $t = 0$ and $t = \tau$ is specified by

$$\mathcal{A}[\mathbf{x}(t), \lambda(t)] = \frac{1}{2}\int_0^\tau dt \left\{ \frac{1}{2}[\dot{\mathbf{x}} - \mathbf{A}(\mathbf{x}, t)]^\top \mathbf{D}^{-1}[\dot{\mathbf{x}} - \mathbf{A}(\mathbf{x}, t)] + \nabla \cdot \mathbf{A}(\mathbf{x}, t) \right\}, \tag{166}$$

where $\mathbf{D} = \frac{1}{2}\mathbf{B}^\top \mathbf{B}$ is assumed to be an invertible matrix[3] and the latter term $\nabla \cdot \mathbf{A}$ comes from the discretization of the Jacobian under the Stratonovich convention. Discretizing with the Itô convention, the last term on the RHS of Eq. (166) disappears[4].

2 Although inertia can be introduced in the dynamical equations through terms that do not break the TR-symmetry, can change the steady state of the system, thus modifying the degree of irreversibility of the process. Since inertial effects may mingle with pre-existing non-equilibrium effects in an unknown way, we prefer to focus here on the primary causes of irreversibility, and defer the analysis of inertial effects to a later stage.

3 For underdamped dynamics, $\mathbf{D}$ is non-invertible, but the path integral formalism can still be adopted, as pioneered by Onsager and Machlup [121].

4 The difference between the two conventions however affects only the path integral weight, and disappears when this is used to compute any physical quantity expressed through an ensemble average — the two discretizations must indeed be equivalent in the case of additive processes.

First law of thermodynamics

The path integral formalism is useful to compute ensemble averages of trajectory-dependent observables. In general, for an observable $\Omega[\mathbf{x}(t)]$,

$$\langle \Omega[\mathbf{x}(t)] \rangle = \int d\mathbf{x}_0 \int \mathcal{D}[\mathbf{x}(t)]\, \Omega[\mathbf{x}(t)] p[\mathbf{x}(t)|\mathbf{x}_0] p_0(\mathbf{x}_0). \tag{167}$$

Among the path-dependent observables one can define, there are the stochastic analogues of standard thermodynamic quantities [169], i.e.

- Work applied to the system (e.g. a colloidal particle in a heat bath). It is convenient to decompose the force $\mathbf{A}(\mathbf{x}, t)$ into a conservative and a non-conservative part[5] (occasionally including external forces) as:

  $$\mathbf{A}(\mathbf{x}, t) = \nabla V(\mathbf{x}; \lambda(t)) + \mathbf{f}(\mathbf{x}, t). \tag{168}$$

  The infinitesimal work is identified with:

  $$đw = (\partial V/\partial \lambda) d\lambda + \mathbf{f} \circ d\mathbf{x}, \tag{169}$$

  so that the work along the stochastic trajectory reads:

  $$w[\mathbf{x}(t)] = \int_0^\tau dt [(\partial V/\partial \lambda)\dot{\lambda} + \mathbf{f} \circ \dot{\mathbf{x}}]. \tag{170}$$

  The $\circ$ symbol indicates that the integrals must be interpreted in the Stratonovich sense [169].

- Heat dissipated into the medium. The infinitesimal increment is defined as

  $$đq = \mathbf{A}(\mathbf{x}, t) \circ d\mathbf{x}, \tag{171}$$

  so that the corresponding path-dependent observable becomes

  $$q[\mathbf{x}(t)] = \int_0^\tau dt [\mathbf{A}(\mathbf{x}, t) \circ \dot{\mathbf{x}}]. \tag{172}$$

These two quantities satisfy the first law:

$$w[\mathbf{x}(t)] = q[\mathbf{x}(t)] + \Delta V(\mathbf{x}_\tau, \mathbf{x}_0), \tag{173}$$

where $\Delta V(\mathbf{x}_\tau, \mathbf{x}_0) = V(\mathbf{x}_\tau; \lambda_\tau) - V(\mathbf{x}_0; \lambda_0)$ is the internal energy variation, with obvious notation for the values assumed by the coordinates and the control parameter of the system at the final and starting times. It is worth noting that alternative definitions of work have been provided, and a debate on the interpretation of stochastic work, heat, and the first law took place soon after their introduction [193, 98, 148, 149].

5 In this decomposition we implicitly set the mobility tensor of the colloidal particle to $\mathbf{1}$. We would otherwise have $\mathbf{A}(\mathbf{x}, t) = \boldsymbol{\mu}\,[\nabla V(\mathbf{x}, \lambda(t)) + \mathbf{f}(\mathbf{x}, t)]$. We recall that, at equilibrium, mobility and diffusion tensors, $\boldsymbol{\mu}$ and $\mathbf{D}$, are linked by the Einstein relation $\mathbf{D} = \boldsymbol{\mu} T$.

Stochastic entropy and irreversibility

A notion of entropy at the level of a single trajectory can also be defined [185, 123]. Thermodynamics teaches us that heat dissipation must be associated to an increase in the entropy of the medium. Since an expression of heat as a functional of the stochastic trajectory is provided, whenever the system admits the definition of a single temperature[6] $T = D_{\alpha\beta}/2$, a first quantity is defined:

$$\Delta s^m[\mathbf{x}(t)] = \frac{q[\mathbf{x}(t)]}{T}. \tag{174}$$

At the same time, a Gibbs entropy

$$\mathcal{S}_G(t) = -\int d\mathbf{x}\, p(\mathbf{x},t) \log p(\mathbf{x},t) \tag{175}$$

is associated to the probability density $p(\mathbf{x},t)$, which represents the solution of the Fokker-Planck equation (FPE) associated to Eq. (164):

$$\begin{aligned} \partial_t p(\mathbf{x},t) &= \mathcal{L}_{FP}\, p(\mathbf{x},t) \\ &= -\partial_\alpha \left[A_\alpha(\mathbf{x},t) p(\mathbf{x},t)\right] + \partial_\alpha \partial_\beta D_{\alpha\beta} p(\mathbf{x},t) \\ &\equiv -\nabla \cdot \mathbf{j}(\mathbf{x},t). \end{aligned} \tag{176}$$

The quantity in Eq. (175) can be interpreted as the average of $-\log p(\mathbf{x},t) \equiv s(t)$. The rate of variation of $s(t)$ along the dynamics generated by $\mathcal{L}_{FP}$ takes the following form, when the noise is additive:

$$\dot{s}(t) = -\partial_t \log p(\mathbf{x}(t),t) + \mathbf{D}^{-1}\left[\frac{\mathbf{j}(\mathbf{x}(t),t)}{p(\mathbf{x}(t),t)} - \mathbf{A}(\mathbf{x}(t),t)\right] \circ \dot{\mathbf{x}}. \tag{177}$$

Once we recognize the (minus) rate of heat dissipation into the medium in the LHS of Eq. (177), it follows that the rate of increase of the total entropy, defined as the sum of the entropy increase in the medium and the change of entropy of the system, reads [167]:

$$\dot{s}^{tot}(t) = \dot{s}(t) + \dot{s}^m(t) = -\partial_t \log p(\mathbf{x}(t),t) + \mathbf{D}^{-1}\frac{\mathbf{j}(\mathbf{x}(t),t)}{p(\mathbf{x}(t),t)} \circ \dot{\mathbf{x}}. \tag{178}$$

This quantity has the nice property that, taking the ensemble average, it is always non-negative and it is equal to zero only at equilibrium. Moreover, the integrated quantity $\Delta s^{tot} = \int_0^\tau dt\, \dot{s}^{tot}(t)$, where $p(\mathbf{x},t)$ is replaced by the stationary solution of the time-independent FPE, satisfies integral and detailed fluctuation theorems [167, 114].

Another definition of fluctuating entropy production, which shares the same features of s^{tot} and is valid for general stochastic processes, was formerly introduced in, e.g., [114, 59]. This definition is as follows:

$$\sigma[\mathbf{x}(t)] = \log \frac{p[\mathbf{x}(t),\lambda(t)]}{p[\mathbf{x}^\dagger(t),\lambda^\dagger(t)]} = \log \frac{p_0(x_0,\lambda_0)}{p_0(x_\tau,\lambda_\tau)} + \log \frac{p[\mathbf{x}(t),\lambda(t)|x_0,\lambda_0]}{p[\mathbf{x}^\dagger(t),\lambda^\dagger(t)|x_\tau,\lambda_\tau]}. \tag{179}$$

Notice that, in contrast to s^{tot}, σ is a functional of the trajectory $\mathbf{x}(t)$. The trajectory can be either a function of a continuous parameter t or a time series, in a unified formalism for both discrete and continuous-time processes. In Eq. (179) p_0 is the

6 In one-dimensional additive processes the definition (174) does not pose any problem. Notice also that, even if there is no unique thermostat temperature T, the heat definition in Eq. (172) is well-posed.

p.d.f. of the initial condition, τ the fixed length of the trajectory, and the † symbol indicates the time reversal operation. The time reversal transformation maps any functional $g_\alpha[\mathbf{x}(t), \lambda(t)]$ into $g_\alpha^\dagger[\mathbf{x}^\dagger(t), \lambda^\dagger(t)]$. Observables having a definite parity under time reversal behave as follows:

$$g_\alpha^\dagger[\mathbf{x}^\dagger(t), \lambda^\dagger(t)] = \epsilon_\alpha g_\alpha[\mathbf{x}(\tau - t), \lambda(\tau - t)] \quad \text{with} \quad \epsilon_\alpha = \pm 1. \tag{180}$$

They are, in other words, eigenfunctions of the time-reversal operator associated to the ± 1 eigenvalues. Assuming that all system coordinates and control parameters have a definite parity, they are transformed as:

$$x_\alpha(t) \mapsto x_\alpha^\dagger(t) = \varepsilon_\alpha x_\alpha(\tau - t), \quad \lambda(t) \mapsto \lambda^\dagger(t) = \varepsilon_\lambda \lambda(\tau - t), \tag{181}$$

whereas the reversed dynamics is described by the same model as the direct one. In the case of the Langevin equation in Eq. (164), this means that the functional form of $\mathbf{A}(\mathbf{x}, t)$ and $\mathbf{B}(\mathbf{x}, t)$ is unchanged, hence the path probability functional is computed following the same rules.

Time reversal belongs in fact to a more general class of 'conjugate' transformations that can be used to derive general fluctuation theorems [177, 166], where the generator of the dynamics can be modified as well. Our interest is limited here to the standard time reversal transformation, as this is the operation used to define the equilibrium condition, and our goal is to quantify irreversibility in active matter models. The path ensemble average of Eq. (179) is the ideal quantifier for it, as it corresponds to the Kullback-Leibler divergence between the path probabilities of the forward and time-reversed dynamics:

$$\mathcal{S} = \langle \sigma[\mathbf{x}(t)] \rangle = \int \mathcal{D}[\mathbf{x}(t)]\, p[\mathbf{x}(t)] \log \frac{p[\mathbf{x}(t)]}{p[\mathbf{x}^\dagger(t)]} = D_{KL}\left[p[\mathbf{x}(t)] || p[\mathbf{x}^\dagger(t)] \right]. \tag{182}$$

Eq. (182) satisfies the property $D_{KL}\left[p[\mathbf{x}(t)] || p[\mathbf{x}^\dagger(t)] \right] \geq 0$, with the equality holding only when detailed balance is realized, i.e. $p[\mathbf{x}(t)] = p[\mathbf{x}^\dagger(t)]$. The parallelism between the fact that $\mathcal{S} \geq 0$, for any $p[\mathbf{x}(t)]$, and the second law of thermodynamics has often been stressed.

A final quantity of interest is the entropy production *rate* (EPR). For a trajectory length τ, the entropy production in Eq. (182) is a function of this parameter, so it is natural to introduce a dissipation rate that is independent of it:

$$\dot{\mathcal{S}} = \lim_{\tau \to \infty} \frac{\mathcal{S}(\tau)}{\tau} = \lim_{\tau \to \infty} -\frac{1}{\tau} \langle \mathcal{A}[\mathbf{x}(t), \lambda(t)] - \mathcal{A}[\mathbf{x}^\dagger(t), \lambda^\dagger(t)] \rangle. \tag{183}$$

To derive the latter identity, Eq.(179) is decomposed into bulk and boundary terms, and the following assumption on the initial and final states of the system is exploited:

$$\lim_{\tau \to \infty} \frac{1}{\tau} \left[\log p_0(\mathbf{x}_0, \lambda_0) - \log p_0(\mathbf{x}_\tau, \lambda_\tau) \right] = 0. \tag{184}$$

A final remark concerns the fact that the entropy production rate defined in Eq. (183) is equivalent to the rate of heat dissipation into the medium in units of $k_B T$ ($k_B = 1$), when the thermostat temperature T is well-defined [166].

Symmetry of non-equilibrium steady states

Let us now focus on the characterization of purely non-equilibrium steady states (NESS). In a NESS the system is not driven out of equilibrium by any external protocol $\lambda(t)$ but by internal irreversible currents. We recall the definition of the probability current

$$\mathbf{j}(\mathbf{x}) = \mathbf{A}(\mathbf{x})\psi(\mathbf{x}) - \mathbf{D}(\mathbf{x})\nabla\psi(\mathbf{x}). \tag{185}$$

coming from rewriting the FPE as a continuity equation (176). We rename ψ the solution of the stationary FPE: $\mathcal{L}_{FP}(\mathbf{x})\psi(\mathbf{x}) = 0$. By *reversible* and *irreversible* we refer to the usual decomposition of current and drift terms [159]:

$$A^{rev}_\alpha(\mathbf{x}) = \frac{1}{2}\left[A_\alpha(\mathbf{x}) - \epsilon_\alpha A_\alpha(\boldsymbol{\epsilon}\mathbf{x})\right], \qquad A^{irr}_\alpha(\mathbf{x}) = \frac{1}{2}\left[A_\alpha(\mathbf{x}) + \epsilon_\alpha A_\alpha(\boldsymbol{\epsilon}\mathbf{x})\right], \tag{186}$$

with $\boldsymbol{\epsilon}$ denoting the time reversal operator. For now, we assume that $\boldsymbol{\epsilon}$ acts (linearly) on coordinate variables $\mathbf{x}$ having a definite parity ($\epsilon_\alpha = \pm 1$). The probability current is correspondingly decomposed as $\mathbf{j} = \mathbf{j}^{rev} + \mathbf{j}^{irr}$:

$$j^{rev}_\alpha(\mathbf{x}) = A^{rev}_\alpha(\mathbf{x})\psi(\mathbf{x}), \qquad j^{irr}_\alpha(\mathbf{x}) = A^{irr}_\alpha(\mathbf{x})\psi(\mathbf{x}) - D_{\alpha\beta}(\mathbf{x})\partial_\beta\psi(\mathbf{x}). \tag{187}$$

It can be proven that a necessary and sufficient condition for detailed balance to hold, i.e. :

$$p(\mathbf{x}_\tau, \tau|\mathbf{x}_0, 0)\psi(\mathbf{x}_0) = p(\boldsymbol{\epsilon}\mathbf{x}_0, \tau|\boldsymbol{\epsilon}\mathbf{x}_\tau, 0)\psi(\boldsymbol{\epsilon}\mathbf{x}_\tau) \quad \forall\tau, \mathbf{x}_0, \mathbf{x}_\tau, \tag{188}$$

is that the two following conditions are verified (in addition to stationarity, $\nabla\cdot\mathbf{j} = 0$):

$$\mathbf{j}^{irr}(\mathbf{x}) = 0 \quad \text{and} \quad D_{\alpha\beta}(\mathbf{x}) = \epsilon_\alpha\epsilon_\beta D_{\alpha\beta}(\boldsymbol{\epsilon}\mathbf{x}). \tag{189}$$

As a consequence, a non-equilibrium steady state in the absence of external driving can be realized only in two- or higher-dimensional systems, since the first equation in (189) can always be solved by quadrature in one dimension (unless the state space is an interval with periodic boundary conditions)[7].

If detailed balance is broken, $\mathbf{j}^{irr}(\mathbf{x}) \neq 0$. This condition is sufficient to guarantee the positivity of the entropy production in a NESS, where Δs^{tot} reduces to the so-called housekeeping entropy production, Δs^{hk}. In general, $\Delta s^{tot} = \Delta s^{hk} + \Delta s^{ex}$, with Δs^{ex} the excess entropy production, which quantifies changes in system entropy associated to explicit time dependencies (absent in a proper NESS). Working with rates, the housekeeping EPR can be conveniently written in the form:

$$\dot{\mathcal{S}}^{hk} = \int d\mathbf{x}\, \psi(\mathbf{x})\, V^{irr}_\alpha(\boldsymbol{\epsilon}\mathbf{x}) \left(\mathbf{D}^{-1}(\boldsymbol{\epsilon}\mathbf{x})\right)_{\alpha\beta} V^{irr}_\beta(\boldsymbol{\epsilon}\mathbf{x}) \tag{190}$$

where $\mathbf{V}(\mathbf{x}) = \mathbf{j}(\mathbf{x})/\psi(\mathbf{x})$ is the phase space velocity [177]. We recall that rates like $\dot{\mathcal{S}}^{hk}$ are defined after averaging the fluctuating entropy production functional (i.e. taking here $\langle\Delta s^{hk}(\tau)\rangle$). The decomposition of the phase space velocity into $\mathbf{V}^{rev}$ and $\mathbf{V}^{irr}$ follows from that of $\mathbf{j}$. By definition, $\mathbf{D}$ is invertible only when positive definite,

7 Another way to see that the steady state of a one-dimensional system is bound to be an equilibrium state is that there cannot be non-conservative forces, i.e. $\mathbf{f} = 0$ in Eq. (170). Hence the first principle implies that $đq = 0$ and compels the entropy production rate to be zero.

from which $\dot{\mathcal{S}}^{hk} \geq 0$ follows. The equality is only realized when the irreversible phase space current vanishes and detailed balance is restored.

Another expression for $\dot{\mathcal{S}}$ is given in [177][8]:

$$\dot{\mathcal{S}} = \int d\mathbf{x}\, \psi(\mathbf{x})\, V^{irr}_{\alpha}(\mathbf{x}) \left(\mathbf{D}^{-1}(\mathbf{x})\right)_{\alpha\beta} V^{irr}_{\beta}(\mathbf{x}). \tag{191}$$

Identifying the housekeeping EPR (190) with Eq. (191), Dal Cengio et al. derived a general constraint on the phase space velocity of non-equilibrium systems whose diffusion matrix $\mathbf{D}$ is invertible and $\mathbf{x}$-independent [47]:

$$\sum_{\alpha\beta} {D^{-1}}_{\alpha\beta} V^{irr}_{\alpha}(\mathbf{x}) V^{irr}_{\beta}(\mathbf{x}) = \sum_{\alpha\beta} {D^{-1}}_{\alpha\beta} V^{irr}_{\alpha}(\boldsymbol{\epsilon}\mathbf{x}) V^{irr}_{\beta}(\boldsymbol{\epsilon}\mathbf{x}) \qquad \text{almost surely.} \tag{192}$$

Since

$$V^{irr}_{\alpha}(\mathbf{x}) \equiv \frac{j_{\alpha}(\mathbf{x})}{\psi(\mathbf{x})} = A^{irr}_{\alpha}(\mathbf{x}) + D_{\alpha\beta}\partial_{\beta}\phi \tag{193}$$

and $A^{irr}_{\alpha}(\mathbf{x}) = \epsilon_{\alpha} A^{irr}_{\alpha}(\boldsymbol{\epsilon}\mathbf{x})$, Eq. (192) implies:

$$\sum_{\alpha} \left[A^{irr}_{\alpha}(\mathbf{x}) + D_{\alpha\beta}\partial_{\beta}\phi_{+}(\mathbf{x})\right] \partial_{\alpha}\phi_{-}(\mathbf{x}) = 0 \qquad \text{almost surely.} \tag{194}$$

The functions $\phi_{\pm}$ are defined as the T-symmetric and T-antisymmetic parts of the quasi-potential $\phi(\mathbf{x}) = -\log\psi(\mathbf{x})$, respectively:

$$\phi_{+}(\mathbf{x}) = \frac{1}{2}\left[\phi(\mathbf{x}) + \phi(\boldsymbol{\epsilon}\mathbf{x})\right], \qquad \phi_{-}(\mathbf{x}) = \frac{1}{2}\left[\phi(\mathbf{x}) - \phi(\boldsymbol{\epsilon}\mathbf{x})\right]. \tag{195}$$

The parity of the variables under time reversal crucially determines the properties of the NESS distribution. If all the variables are even under time reversal, $\boldsymbol{\epsilon}\mathbf{x} = \mathbf{x}$, then $\phi_{-}(\mathbf{x}) = 0$ and Eq. (194) is trivially satisfied. In order to break the detailed balance, the remaining part of the quasipotential must satisfy $\partial_{\alpha}\phi_{+}(\mathbf{x}) \neq -{D^{-1}}_{\alpha\beta} A^{irr}_{\beta}(\mathbf{x})$. Notice that ϕ_{-} cannot be a constant function different from zero: thus, if $\mathbf{x}$ has odd coordinates, also $\partial_{\alpha}\phi_{-}$ is not the null function. In the presence of T-odd coordinates, Eq. (194) must be interpreted as an orthogonality condition, according to which $\mathbf{A}^{irr} + \mathbf{D}\left(\nabla\phi_{+}\right)$ is the tangent vector to the level curves of ϕ_{-}. The information on the NESS we can obtain from this constraint is however scarce, since ϕ_{+} is one of the unknowns we want to determine.

The conclusion derived by Dal Cengio et al. in [47] is stronger but, in our opinion, the derivation is wrong: even working with a diagonal diffusion matrix $\mathbf{D}$, we see no reason to equate all the terms in the sum in Eq. (194) to zero and conclude that in a NESS either $\phi_{-}(\mathbf{x}) = 0$ or $\nabla\phi_{+}(\mathbf{x}) = -\mathbf{D}^{-1}\mathbf{A}^{irr}(\mathbf{x})$. It is however true that both conditions cannot be realized simultaneously. Here we propose an alternative justification for this fact and a general formula for the housekeeping entropy production, which is obtained by further manipulation of the results by Dal Cengio et al..

For the sake of simplicity, let us focus on the case of diagonal $\mathbf{D}$ diffusion matrix. All the results can be generalized to the non-diagonal case, thanks to the fact that

8 Eq. (191) can be derived from Eq. (183) neglecting $\lambda(t)$ and assuming natural boundary conditions for $\psi(\mathbf{x})$. Moving from the time average in Eq. (183) to the average over the NESS probability measure $\psi(\mathbf{x})$ in Eq. (191) is justified under reasonable ergodicity assumptions.

$\mathbf{D}$ is symmetric and positive definite[9]. We start from an explicit rewriting of the entropy production rate from Eq. (191):

$$\dot{\mathcal{S}} = \int d\mathbf{x}\, e^{-\phi(\mathbf{x})} \sum_\alpha D_\alpha^{-1} \left[A_\alpha^{irr}(\mathbf{x}) + D_\alpha \partial_\alpha \phi(\mathbf{x}) \right]^2 . \tag{196}$$

Exploiting the symmetries of $\mathbf{A}^{irr}$, ϕ_+ and ϕ_- under the time reversal transformation, we can rewrite Eq. (196) as

$$\begin{aligned} \dot{\mathcal{S}} = &\int d\mathbf{x}\, 2\cosh\left(\phi_-(\mathbf{x})\right) e^{-\phi_+(\mathbf{x})} \sum_\alpha D_\alpha^{-1} \left[\left(A_\alpha^{irr}(\mathbf{x}) + D_\alpha \partial_\alpha \phi_+(\mathbf{x}) \right)^2 + D_\alpha \left(\partial_\alpha \phi_-(\mathbf{x}) \right)^2 \right] \\ &+ \int d\mathbf{x}\, 4\sinh\left(\phi_-(\mathbf{x})\right) e^{-\phi_+(\mathbf{x})} \sum_\alpha \left(A_\alpha^{irr}(\mathbf{x}) + D_\alpha \partial_\alpha \phi_+(\mathbf{x}) \right) \partial_\alpha \phi_-(\mathbf{x}). \end{aligned} \tag{197}$$

The second line of Eq. (197) is null because of Eq. (194). Hence:

$$\dot{\mathcal{S}} = \int d\mathbf{x}\, \cosh\left(\phi_-(\mathbf{x})\right) e^{-\phi_+(\mathbf{x})} \sum_\alpha \left[D_\alpha^{-1} \left(A_\alpha^{irr}(\mathbf{x}) + D_\alpha \partial_\alpha \phi_+(\mathbf{x}) \right)^2 + \left(\partial_\alpha \phi_-(\mathbf{x}) \right)^2 \right] . \tag{198}$$

From this formula for the entropy production rate it is immediate to verify $\dot{\mathcal{S}} \geq 0$. The condition to have zero entropy production rate is that both of the following conditions are verified:

$$A_\alpha^{irr}(\mathbf{x}) + D_\alpha \partial_\alpha \phi_+(\mathbf{x}) = 0 \quad \forall \alpha \tag{199}$$

and

$$\partial_\alpha \phi_-(\mathbf{x}) = 0 \quad \forall \alpha \iff \phi_-(\mathbf{x}) = 0. \tag{200}$$

Breakdown of detailed balance imposes that at least one of these terms is nonzero, together with the constraint (194).

Application to the NESS of a Langevin-Vicsek-like model

The discussion contained in the previous section highlights the importance of keeping track of the parity of the state variables involved. A simple argument based on the analysis of such properties in Langevin-Vicsek-like models can connect the breakdown of detailed balance to that of the $O(n)$ symmetry in the velocity subspace.

Let us consider active matter systems which are described as a set of N interacting ABPs, whose degrees of freedom are position and velocity coordinates $\{(\mathbf{r}_i, \mathbf{v}_i), i = 1 \dots N\}$ in a d-dimensional external space. We impose that the speed of the particles is fixed, and we restrict ourselves to $d = 2$, so that we deal with flying spins whose effective degrees of freedom are $\{(\mathbf{r}_i, \theta_i), i = 1 \dots N\}$. Let us denote by $\mathbf{\Gamma}$ the state variable of the system, and by $\psi(\mathbf{\Gamma})$ its non-equilibrium steady state distribution.

While positions and velocities are respectively even and odd under time reversal, the direction of flight of the particle has no definite parity. Under time reversal, the degrees of freedom are transformed as:

$$\epsilon \mathbf{r}_i = \mathbf{r}_i, \quad \epsilon \mathbf{v}_i = -\mathbf{v}_i, \quad \epsilon \theta_i = \theta_i + \pi. \tag{201}$$

9 We remark that it is not possible to merely diagonalize $\mathbf{D}$ through a change of basis, because we need all the coordinates of $\mathbf{x}$ to have a definite parity under time reversal.

Since not all of the coordinates of $\mathbf{\Gamma}$ are even under time reversal, we can in principle have $\phi_-(\mathbf{\Gamma}) \neq 0$ and $\nabla\phi_-(\mathbf{\Gamma}) \neq 0$ in non-equilibrium conditions. This inequality corresponds to $\psi(\{\mathbf{r}_i, \theta_i\}) \neq \psi(\{\mathbf{r}_i, \theta_i + \pi\})$, meaning that the breakdown of detailed balance causes an *explicit* symmetry breaking in the NESS.

In the following, we show that this condition is realized in a simple ABP model with Vicsek-like interactions in $d = 2$, described by the equations:

$$\dot{\mathbf{r}}_i = v_0\mathbf{e}(\theta_i), \quad \dot{\theta}_i = -Jn_{ij}(\{\mathbf{r}_i\})\sin(\theta_i - \theta_j) + \sqrt{2D}\xi_i, \tag{202}$$

where $\mathbf{e}(\theta_i) = (\cos\theta_i, \sin\theta_i))$, $n_{ij}(\{\mathbf{r}_i\})$ is assumed to be symmetric, and a sum over repeated indices is understood. In this case, the drift term decomposes into its reversible and irreversible parts as:

$$\mathbf{A}_i^{rev} = \begin{pmatrix} v_0\mathbf{e}(\theta_i) \\ 0 \end{pmatrix}, \qquad \mathbf{A}_i^{irr} = \begin{pmatrix} 0 \\ -Jn_{ij}\sin(\theta_i - \theta_j) \end{pmatrix}. \tag{203}$$

If we hypothesize that the rotational symmetry in the space of velocities is not explicitly broken by irreversibility, $\phi_-(\mathbf{\Gamma}) = 0$, then the following condition must hold:

$$\partial_{\theta_i}\phi_+(\mathbf{\Gamma}) \neq Jn_{ij}\sin(\theta_i - \theta_j) \iff \phi_+(\mathbf{\Gamma}) \neq \frac{H(\mathbf{\Gamma})}{D} + c(\{\mathbf{r}_i\}), \tag{204}$$

with $H(\mathbf{\Gamma}) = -\frac{J}{2}\sum_{ij} n_{ij}\cos(\theta_i - \theta_j)$ the classical XY Hamiltonian. However, the FPE associated to Eq. (202) under the assumption $\phi_-(\mathbf{\Gamma}) = 0$ reduces to the equation for the equilibrium process which we recover when $v_0 = 0$. Let us show it in greater detail by starting from the stationary FPE

$$v_0\sum_i \mathbf{e}(\theta_i)\cdot\nabla_i\psi = J\sum_{ij} n_{ij}\partial_{\theta_i}(\sin(\theta_i - \theta_j)\psi) + D\sum_i \partial^2_{\theta_i\theta_i}\psi, \tag{205}$$

from which a PDE for the quasi-potential can be derived:

$$v_0\sum_i \mathbf{e}(\theta_i)\cdot\nabla_i\phi = J\sum_{ij} n_{ij}\cos(\theta_i - \theta_j) + J\sum_{ij} n_{ij}\sin(\theta_i - \theta_j)\partial_{\theta_i}\phi + D\sum_i \partial^2_{\theta_i\theta_i}\phi + D\sum_i(\partial_{\theta_i}\phi)^2. \tag{206}$$

We can now apply the time-reversal operator to both the R.H.S. and L.H.S. of Eq. (206) and split the quasi-potential into its T-symmetric and T-antisymmetric parts to rewrite a set of two coupled stationary equations:

$$\begin{aligned} v_0\sum_i \mathbf{e}(\theta_i)\cdot\nabla_i\phi_- = {} & J\sum_{ij} n_{ij}\cos(\theta_i - \theta_j) + J\sum_{ij} n_{ij}\sin(\theta_i - \theta_j)\partial_{\theta_i}\phi_+ \\ & + D\sum_i \partial^2_{\theta_i\theta_i}\phi_+ + D\sum_i(\partial_{\theta_i}\phi_+)^2 + D\sum_i(\partial_{\theta_i}\phi_-)^2; \end{aligned} \tag{207}$$

$$v_0\sum_i \mathbf{e}(\theta_i)\cdot\nabla_i\phi_+ = J\sum_{ij} n_{ij}\sin(\theta_i - \theta_j)\partial_{\theta_i}\phi_- + D\sum_i \partial^2_{\theta_i\theta_i}\phi_- + 2D\sum_i(\partial_{\theta_i}\phi_+)(\partial_{\theta_i}\phi_-). \tag{208}$$

Assuming $\phi_- = 0$ in Eq.(207), the resulting PDE for $\psi = e^{\phi_+}$ is of the form:

$$0 = J\sum_{ij} n_{ij}\partial_{\theta_i}(\sin(\theta_i - \theta_j)\psi) + D\sum_i \partial^2_{\theta_i\theta_i}\psi. \tag{209}$$

This equation is solved by the Boltzmann distribution[10]

$$\psi(\mathbf{\Gamma}) = f(\{\mathbf{r}_i\}) \exp[-H(\mathbf{\Gamma})] \iff \phi_+ = \frac{H(\mathbf{\Gamma})}{D} + c(\{\mathbf{r}_i\}), \tag{210}$$

with $c(\{\mathbf{r}_i\})$ an arbitrary constant. Hence the non-equilibrium condition (204) is contradicted: the hypothesis $\phi_-(\mathbf{\Gamma}) = 0$ leads to an absurdum.

We must not confuse the condition $\phi_-(\mathbf{\Gamma}) \neq 0$ with an explicit breaking of the full rotational symmetry of the ABP system. Due to the coupling between position and velocity degrees of freedom, a reorientation of the birds' direction of motion cannot be decoupled from a rigid rotation of the whole system, so the expected condition to have a full explicit breaking of the rotational symmetry of the system is $\psi(\{\mathbf{r}_i, \theta_i\}) \neq \psi(\{\mathcal{R}_\alpha \mathbf{r}_i, \theta_i + \alpha\})$, where $\mathcal{R}_\alpha$ implements a rotation of an angle α on the two-dimensional spatial coordinates. However, this full explicit symmetry breaking is not occurring in flocking models.

4.2 ENTROPY PRODUCTION IN VICSEK-LIKE MODELS

The deep connection between irreversibility and emergence of collective motion motivates our interest in quantifying the non equilibrium nature of the process at a microscopic level. The quantity we choose to study is the entropy production rate: even though this is a non-universal, model-dependent observable, it is by definition a measure of the breakdown of detailed balance, which is the event we want to relate to the emergence of polar order in the system.

4.2.1 Stochastic thermodynamics for polar ABPs

Computing the entropy production rate for the model in Eq. (202) is a simple exercise. The structure of the equations of motion

$$\dot{\mathbf{r}}_i = v_0 \mathbf{e}(\theta_i), \quad \dot{\theta}_i = -J n_{ij}(\{\mathbf{r}_i\}) \sin(\theta_i - \theta_j) + \sqrt{2D}\xi_i, \tag{211}$$

is such that the update for the positions is deterministic, whereas, for symmetric n_{ij}, the second equation has the structure of an overdamped stochastic dynamics of a Hamiltonian system, with a gradient force term of the form $-\partial_{\theta_i} H(\mathbf{\Gamma})$, where $H(\mathbf{\Gamma})$ is the XY Hamiltonian introduced above:

$$H(\mathbf{\Gamma}) = -\frac{J}{2} \sum_{ij} n_{ij}(\{\mathbf{r}_i\}) \cos(\theta_i - \theta_j). \tag{212}$$

We highlight that this aligning force acts on the angular degrees of freedom (or, equivalently on the XY spin vectors) and depends on the particles' orientations $\{\theta_i, i = 1, \ldots, N\}$, in contrast to microscopic models for scalar active matter.

10 Uniqueness of the stationary solution of a Fokker-Planck equation is guaranteed under rather general smoothness hypotheses on the FPE stationary solution and drift term, and other requirements on the manifold where the solution is defined. A proof based on the Perron-Frobenius theorem is provided in [196]. The Boltzmann-Gibbs distribution, which we can easily construct when the force term is gradient-like, is therefore our unique solution for Eq. (209). This fact is enough to constrain $\phi_-(\mathbf{\Gamma}) \neq 0$. Let us notice that we crucially exploited the symmetry of n_{ij} in Eq. (210) to write a Boltzmann-Gibbs distribution.

The fact that the first equation of (211) is deterministic raises the apparent problem of the non invertibility of the diffusion matrix of the process, in the $\mathbf{\Gamma}$ phase space. Nonetheless, a formal Onsager-Machlup action can always be written, either employing δ-functions to define the propagator, or introducing a small regularizing diffusion term on the $\mathbf{r}_i$'s. Introducing such action

$$\mathcal{A}[\{\mathbf{r}_i(t),\theta_i(t)\}] = \lim_{\epsilon\to 0}\int_0^\tau \sum_i \left\{ \frac{1}{\epsilon^2}\left[\dot{\mathbf{r}}_i - v_0\mathbf{e}(\theta)\right]^2 + \frac{1}{4D}[\dot{\theta}_i + J\sum_j n_{ij}\sin(\theta_i-\theta_j)]^2 \right\} dt \tag{213}$$

into the definitions (179) and (183), we obtain

$$\sigma[\{\mathbf{r}_i(t),\theta_i(t)\}] = \log\frac{p(\{\mathbf{r}_i(0),\theta_i(0)\})}{p(\{\mathbf{r}_i(\tau),\theta_i(\tau)\})} - \frac{J}{D}\sum_i\int_0^\tau d\theta_i\circ n_{ij}\sin(\theta_i-\theta_j), \tag{214}$$

and thus

$$\dot{\mathcal{S}} = \frac{\langle\dot{q}\rangle}{D} = -\frac{J}{D}\sum_{ij}\langle\frac{d\theta_i}{dt}\circ n_{ij}\sin(\theta_i-\theta_j)\rangle. \tag{215}$$

Boundary terms coming from the ratio of initial and final steady state distributions have been neglected; brackets indicate averages over the non-equilibrium steady-state distributions. Recalling the definition of stochastic heat, Eq. (215) coincides with the average heat dissipated in the system, divided by the temperature (here named D). The dissipated energy is equal to the average work made by the aligning torques.

An alternative way to cope with the same problem is to restrict the process, since the beginning, to those phase space directions where $\mathbf{A}^{irr}$ has nonzero projections. Recalling Eq. (203), this operation basically consists of eliminating the spatial degrees of freedom. In this subspace, the diffusion matrix is invertible, and all the formulas above can be used, including Eq. (190) and Eq. (191) for the housekeeping entropy production rate. The effect of the eliminated spatial degrees of freedom can be taken into account by recognizing them as external parameters driving the system through a time-varying protocol. In this framework, reshuffling is seen as a quasi-static transformation implemented on the flock. At a formal level, we are just replacing $n_{ij}(\{\mathbf{r}_i\})$ with $n_{ij}(t)$ in Eq. (211), thus transforming the implicit time dependence of the adjacency matrix into explicit.

Having rephrased the problem in this way, we can borrow from a large corpus of references about stochastic thermodynamics of quasi-statically driven systems. In particular, following Crooks [60], we can proceed by discretizing over infinitesimal intervals of time the trajectories of the angular degrees of freedom — $i_n = \{\theta_i(t_n), i = 1\dots N\}$ — and the trajectories of the external variables — $\lambda_n = \{n_{ij}(t_n), i,j = 1\dots N\}$. By means of straightforward manipulations we can firstly compute the fluctuating entropy production for a trajectory of arbitrary length τ, discretized into N_s steps of amplitude Δt:

$$\sigma[\{\theta_i(t),\mathbf{r}_i(t)\}] = \log\frac{P(i_0\xrightarrow{\lambda_1}i_1\xrightarrow{\lambda_2}i_2\dots i_{N_s-1}\xrightarrow{\lambda_{N_s}}i_{N_s})}{P(i_{N_s}\xrightarrow{\lambda_{N_s}}i_{N_s-1}\dots i_2\xrightarrow{\lambda_2}i_1\xrightarrow{\lambda_1}i_0)} = \log\frac{P(i_0;\lambda_0)}{P(i_{N_s};\lambda_{N_s})} + \sum_n\Delta\sigma_n, \tag{216}$$

with

$$\Delta\sigma_n = -\frac{J}{D}\sum_i \Delta\theta_i^n n_{ij}^{n+1}\frac{1}{2}[\sin(\theta_i^n - \theta_j^n) + \sin(\theta_i^{n+1} - \theta_j^{n+1})] + o(\Delta t). \tag{217}$$

Then, taking the limits $\Delta t \to 0$, $\tau \to \infty$, Eq. (217) yields an expression for $\dot{\mathcal{S}}$ that is identical to Eq. (215). The average is now interpreted as an average over the distribution of the θ_i's and the realizations of the $\lambda(t)$ trajectories. Our transformation protocol is indeed effectively stochastic and correlated to the state of the reduced system: for this reason, we generally have $\dot{\mathcal{S}} > 0$.

Alternative formulas

Eq. (215) can be interpreted as the heat that gets dissipated in the thermal bath per unit time. Another equivalent formula for the entropy production rate of the Langevin-Vicsek model in Eq. (211) is obtained by exploiting the definition (183) and the stationarity of the process. Stratonovich calculus allows us to apply standard differentiation and integration rules, according to which Eq. (214) can be integrated by parts, leading to non-extensive (in time) boundary terms and a resulting expression for the entropy production rate:

$$\dot{\mathcal{S}} = \frac{\langle \dot{w}_{\text{resh}} \rangle}{D} = -\frac{J}{2D}\sum_{ij}\langle \dot{n}_{ij} \circ \cos(\theta_i - \theta_j)\rangle. \tag{218}$$

The derivation of Eq. (218) is based on the assumption that there exists a steady state where the properties of the interaction matrix of the flock are stable, as well as its degree of polar order. The neglected boundary terms are indeed proportional to the change in the 'internal energy' of the system between starting and final times: $-\left[\langle H\rangle(\tau) - \langle H\rangle(0)\right]/D$. The equivalence between Eq. (215) and Eq. (218) is nothing but an expression of the first principle of stochastic thermodynamics: if the flock is in a steady state, the internal energy of the system is not varied. We relabel the quantity in Eq. (218) as $\langle \dot{w}_{\text{resh}} \rangle / D$, since it can be interpreted as the average work of a fictitious external protocol designed to reshuffle particles' configurations.

The same formula is obtained in the 'external protocol' framework assuming *local detailed balance*. The local detailed balance condition reads [60]:

$$\frac{P(i_n, \lambda_{n+1})P(i_n \xrightarrow{\lambda_{n+1}} i_{n+1})}{P(i_{n+1}, \lambda_{n+1})P(i_{n+1} \xrightarrow{\lambda_{n+1}} i_n)} = 1, \tag{219}$$

where $P(i_n, \lambda_n)$ are meant to be the equilibrium Boltzmann distributions of the orientational degrees of freedom associated to fixed network configurations (fixed λ_n):

$$P_{eq}(\{\theta_i\}; \mathbf{n}) = \frac{1}{Z(\lambda)} e^{-\beta H(\{\theta_i\};\mathbf{n})}, \quad \text{with} \quad \beta = D^{-1}. \tag{220}$$

Applying the local detailed balance hypothesis (219) on Eq. (216) and taking the continuous limit yields:

$$\sigma[\{\theta_i(t), \mathbf{r}_i(t)\}] = \beta \int_0^\tau dt \sum_{i<j} \dot{n}_{ij}\left[\partial_{n_{ij}} H(\mathbf{n}) - \langle \partial_{n_{ij}} H(\mathbf{n})\rangle_{eq,\mathbf{n}}\right]. \tag{221}$$

In computing $\dot{\mathcal{S}}$, the contribution from the second term in the RHS of Eq. (221) vanishes under the assumption of stationarity, and the formula reduces to Eq. (218).

Microscopic description of neighbors' reshuffling

Eq. (218) highlights how the non-equilibrium character of this simple Vicsek-like model is entirely due to the rewiring of the interaction network. The formulas derived so far are pretty general and can be applied to any Langevin-Vicsek model with momentum-conserving interaction, be it metric or topological. Additional contributions to the entropy production rate appear if we consider non-reciprocal alignment forces, which do not obey the action-reaction principle and may generate currents in the flock even in the limit of vanishing motility. Since we are interested in quantifying the effect of self-propulsion alone, we have restricted to the case of symmetric interaction networks, to which Eqs. (215) and (218) apply. How the adjacency matrix changes over time depends however on its precise parametrization.

We focus here on the case of additive isotropic metric interactions, which by definition have the following property: $n_{ij} = n(|\mathbf{r}_i - \mathbf{r}_j|)$. Hence Eq. (218) reads:

$$\dot{\mathcal{S}} = \frac{Jv_0}{2D} \sum_{ij} \Big\langle -\frac{\partial n(r)}{\partial r}\Big|_{r=|\mathbf{r}_i-\mathbf{r}_j|} \frac{(\mathbf{e}(\theta_i) - \mathbf{e}(\theta_j)) \cdot (\mathbf{r}_i - \mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j|} \cos(\theta_i - \theta_j) \Big\rangle. \tag{222}$$

When direct alignment is short ranged, it is reasonable to assume that $n(r)$ and its derivative $\partial_r n(r)$ decay (in magnitude) sufficiently fast. A common parametrization for the Vicsek model is, for instance, $n_{ij} = \Theta(R - |\mathbf{r}_i - \mathbf{r}_j|)$, where Θ is the Heaviside function. In this case, $-\partial_r n = \delta(r - R)$, so only pairs of particles at a distance equal to the interaction radius are selected in Eq. (222).

The whole formula can be read, from a coarse-grained perspective, as a *weighted divergence*, the weight being the mutual alignment of particles' velocities. Notice that this weight has no definite sign: for almost aligned birds, $\cos(\theta_i - \theta_j) > 0$ and positive contributions to the EPR come from diverging configurations. On the contrary, anti-aligned spins, such that $\cos(\theta_i - \theta_j) < 0$, produce entropy when they move towards each other. In Fig. 20.c we show some configurations, in terms of velocity alignment and mutual displacement, associated to positive (red) and negative (blue) entropy production. The same *weighted divergence* interpretation also holds for other parametrizations of the n_{ij} matrix which bring little variations compared to the Heaviside function, like $n(r) = e^{-r/R}$.

To fix the ideas, we will only work with the former parametrization from now on. In this case, Eq. (218) is conveniently rewritten as:

$$\dot{\mathcal{S}} = \frac{Jv_0}{2D} \sum_{ij} \big\langle \delta(R - |\mathbf{r}_i - \mathbf{r}_j|) \cos(\theta_i - \theta_j) \left[\cos(\hat{\alpha}_{ij} - \varphi) - \cos\hat{\alpha}_{ij}\right] \big\rangle, \tag{223}$$

where we introduced a new variable $\hat{\alpha}_{ij}$, defined as the angle between the distance vector $\mathbf{r}_j - \mathbf{r}_i$ and the direction of motion of the i-th bird, $\mathbf{e}(\theta_i)$. Eq. (223) shows that, in the presence of metric interactions, the EPR is a rather simple quantity, expressed as the sum of two-particle functions. The violation of the time reversal symmetry is then ultimately determined by the properties of the two-body steady state distribution.

Let us call $q(\alpha, \varphi)$ the number of particles with mutual misalignment $\varphi = \theta_j - \theta_i$ placed at a distance R along a direction parametrized by a newly defined angle

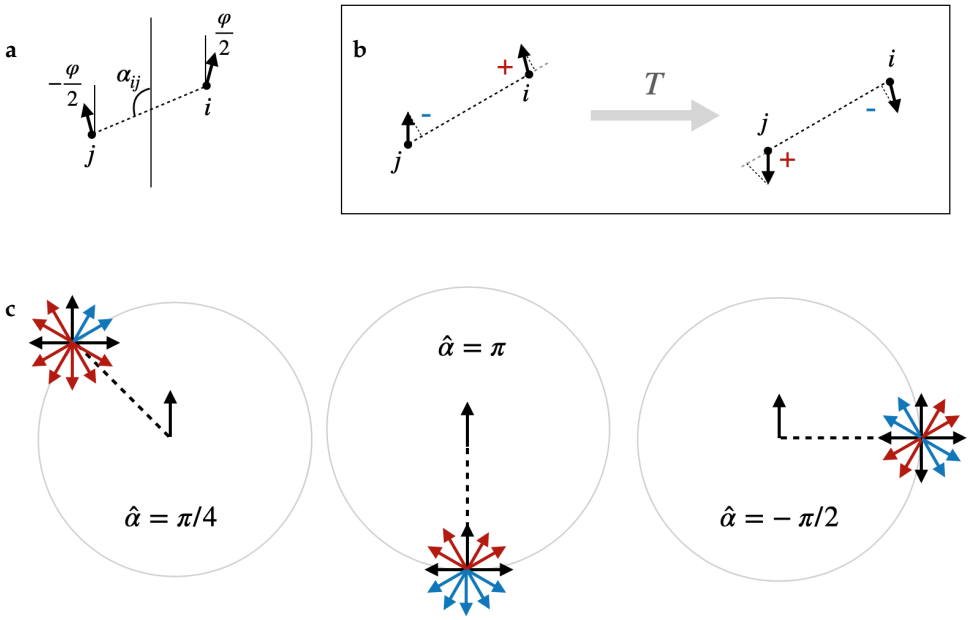

Figure 20: **a**: Definition of α and φ used in the main text: given $\hat{\alpha}_{ij}$ the angle formed by the vector $\mathbf{\Delta} = \mathbf{r}_j - \mathbf{r}_i$ with $\mathbf{v}_i$, and $\varphi = (\theta_j - \theta_i) \bmod 2\pi$, we define $\alpha_{ij} = \hat{\alpha}_{ij} - \varphi/2 + \pi\Theta(\varphi \bmod 4\pi - 2\pi)$. Notice that, because of the modulo operation in the definition of φ, $\varphi/2 \in [0, \pi)$. Since $\varphi/2$ can only span a half circle, the remaining set of configurations is parametrized by transforming α into $\alpha + \pi$. **b**: Time reversal transformation on the pair of particles (i, j): the positions are unchanged, while the directions of motion are flipped. The transformation amounts to the rotation of α of an angle π. **c**: Illustration of positive (red arrows) and negative (blue arrows) contribution to the EPR for three selected values of $\hat{\alpha}$. Black arrows denote configurations which do not produce entropy.

$\alpha_{ij} = \alpha$. It is indeed convenient to 'symmetrize' the definition of the displacement angle $\hat{\alpha}_{ij}$ and introduce the new variable:

$$\alpha_{ij} = \hat{\alpha}_{ij} - \varphi/2 + \pi\Theta(\varphi \bmod 4\pi - 2\pi)\,. \tag{224}$$

The displacement angle $\hat{\alpha}_{ij}$ is the one formed by the distance vector $\mathbf{r}_j - \mathbf{r}_i$ with $\mathbf{v}_i$ (see Fig. 20); the variable α_{ij} indicates, on the contrary, the angular displacement with respect to the average direction of the two particles[11]. The Heaviside function Θ in Eq. (224) serves us to reconcile the 2π-periodicity of the natural angular variables of the problem, $\hat{\alpha}_{ij}$, θ_i and θ_j, with the introduction of $\varphi/2$ in the definition (224). With this rewriting, Eq. (223) becomes:

$$\dot{\mathcal{S}} = \frac{Jv_0}{2D} \int_{[0,2\pi]^2} q(\alpha, \varphi)\varepsilon(\alpha, \varphi) d\alpha d\varphi, \tag{225}$$

with

$$\varepsilon(\alpha, \varphi) = \cos\varphi \left[\cos\left(\alpha - \frac{\varphi}{2}\right) - \cos\left(\alpha + \frac{\varphi}{2}\right)\right], \quad \text{for} \quad \varphi \in [0, 2\pi). \tag{226}$$

The function in Eq. (226) is represented in Fig. 21: a relevant feature is that $\varepsilon(\alpha, \varphi) = -\varepsilon(\alpha + \pi, \varphi)$. Rewriting $\dot{\mathcal{S}}$ as in Eq. (225) is helpful to notice that the

11 Only when φ is convex, $\hat{\alpha}$ is the angle formed by $\mathbf{r}_j - \mathbf{r}_i$ with respect to the direction of motion of the center of mass of this 2-body system; when φ is concave, the reference direction is the opposite one.

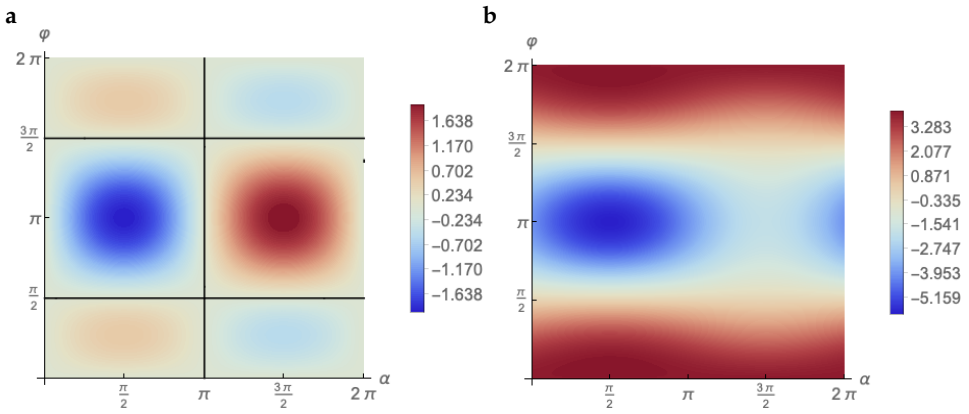

Figure 21: **a**: $\varepsilon(\alpha, \varphi)$, i.e. the intensive fluctuating entropy production rate (rescaled by J/D) as a function of α and φ. **b**: Plot of the function $\varepsilon(\alpha, \varphi) + \mu \cos \varphi$ (with $\mu = 4$). This function seems to reproduce qualitatively the numerically reconstructed $\log q(\alpha, \varphi)$ (cfr. Fig. 26) — at least in the proximity of the transition point where the EPR peaks. For a better quantitative evaluation, cfr. Fig. 28.

entropy production is wiped out by symmetry arguments: if mirror configurations like those in Fig. 20.b are equally probable, the EPR is bound to be zero. A necessary condition to have violation of the time reversal symmetry is therefore that the steady-state two-particle distribution is asymmetric under exchange of the relative positions of the two self-propelled particles, i.e. $q(\alpha, \varphi) \neq q(\alpha + \pi, \varphi)$. The positivity of the entropy production rate constrains the way this asymmetry is realized, as illustrated by numerical simulations.

PT violation

Before presenting numerical results, let us remark that asking for a violation of the T symmetry is the same as asking for a violation of the PT symmetry. As already pointed out by Cates and coworkers in a coarse-grained analysis of the stochastic field dynamics [15], in an extended flock, where the entropy production rate is computed through spatial averages of local quantities, irreversibility requires that the PT-symmetry is violated, not the T-symmetry alone. If PT symmetry holds, reversibility is restored. At the microscopic level, the same requirement is still true, and still results into a constraint on the spatial distribution of the particles which make up the active fluid. This constraint is however expressed as a condition on the local scale, whereas constraints are imposed on the shape of the mesoscopic patterns when working with stochastic field theories [15].

The effect of the parity transformation on the microscopic degrees of freedom of the system is to flip all the orientation vectors, $\mathbf{e}(\theta_i) \mapsto -\mathbf{e}(\theta_i) = \mathbf{e}(\theta_i + \pi)$, and to reflect the spatial coordinates with respect to the origin, $\mathbf{r}_i \mapsto -\mathbf{r}_i$. Combined with the T transformation, which only maps the velocities $\mathbf{v}_i$ to $-\mathbf{v}_i$, PT amounts to

$$\mathbf{r}_i \mapsto -\mathbf{r}_i\,, \quad \theta_i \mapsto \theta_i + 2\pi = \theta_i \bmod 2\pi. \tag{227}$$

On our reduced variables α and φ, both T and PT act as follows:

$$\alpha \mapsto \alpha + \pi\,, \quad \varphi \mapsto \varphi. \tag{228}$$

The asymmetry condition on $q(\alpha, \varphi)$ that we derive from the irreversibility condition $\dot{\mathcal{S}} \geq 0$ is then the microscopic analogue of the asymmetry constraints emerging in the hydrodynamic model.

4.2.2 Numerical results

Phenomenology of the simulated model

In order to check the predictions of the previous sections, we study numerically two variants of the model in Eq. (211) for systems of N aligning ABPs constrained on a square box of linear size $L = (N/\rho_0)^{1/2}$ with periodic boundary conditions. In this section we analyze the metric variant with $n_{ij} = \Theta(R - r_{ij})$. The discussion of the Voronoi topological variant is contained in 4.2.5.

The considered continuous-time ABP model exhibits a phase transition from an isotropic disordered phase to a polar ordered phase, as in the original discrete-time model introduced by Vicsek et al. [192], but it does not reproduce some of the familiar patterns observed there, such as travelling bands. Typical configurations are shown in Fig. 22. The structures that tend to form, as the noise amplitude is reduced and the system enters into the ordered phase, are high-density polar clusters moving against an isotropic low-density background fluid. Phase coexistence between dilute and dense portions of the system is lost when the rotational diffusion coefficient of the ABPs is very low (i.e. the persistence time $\tau_p = D^{-1}$ is very high); in such conditions the whole flock consists of a single dense moving cluster.

The most important difference between the ABP model we simulate and the original model by Vicsek et al.[192], which explains the different phenomenology we observe, is in the additive nature of interactions [51]. In the original model, each bird firstly implements an *average* over the directions of motion of the surrounding particles, and then imperfectly aligns to it. This average operation makes the strength of the total alignment force roughly independent of local density, whereas in Eq. (211) local density fluctuations enhance or lower the weight of the ferromagnetic interaction over the stochastic term. The original Vicsek model is in fact better described, in the Langevin formalism, by a 'weighted' adjacency matrix $n_{ij} = w_i \Theta(R - |\mathbf{r}_i - \mathbf{r}_j|)$, with weights equal to the inverse number of neighbors: $w_i = n_i^{-1} = \left[\sum_j \Theta(R - |\mathbf{r}_i - \mathbf{r}_j|)\right]^{-1}$. This matrix cannot be built using only the information on the positions of particles i and j (although interactions are still local), it is generally non-symmetric and causes the violation of momentum conservation within the system. However, with this kind of non-additive interactions, band-like structures are recovered. A detailed analysis of the properties of continuous-time ABP models with additive versus non-additive interactions is contained in [51]. Despite the greater numerical stability of the latter, we decide to use a model with additive interactions in our analysis, in order to disentangle possible non-equilibrium effects due to the lack of momentum conservation from non-equilibrium effects resulting from self-propulsion alone.

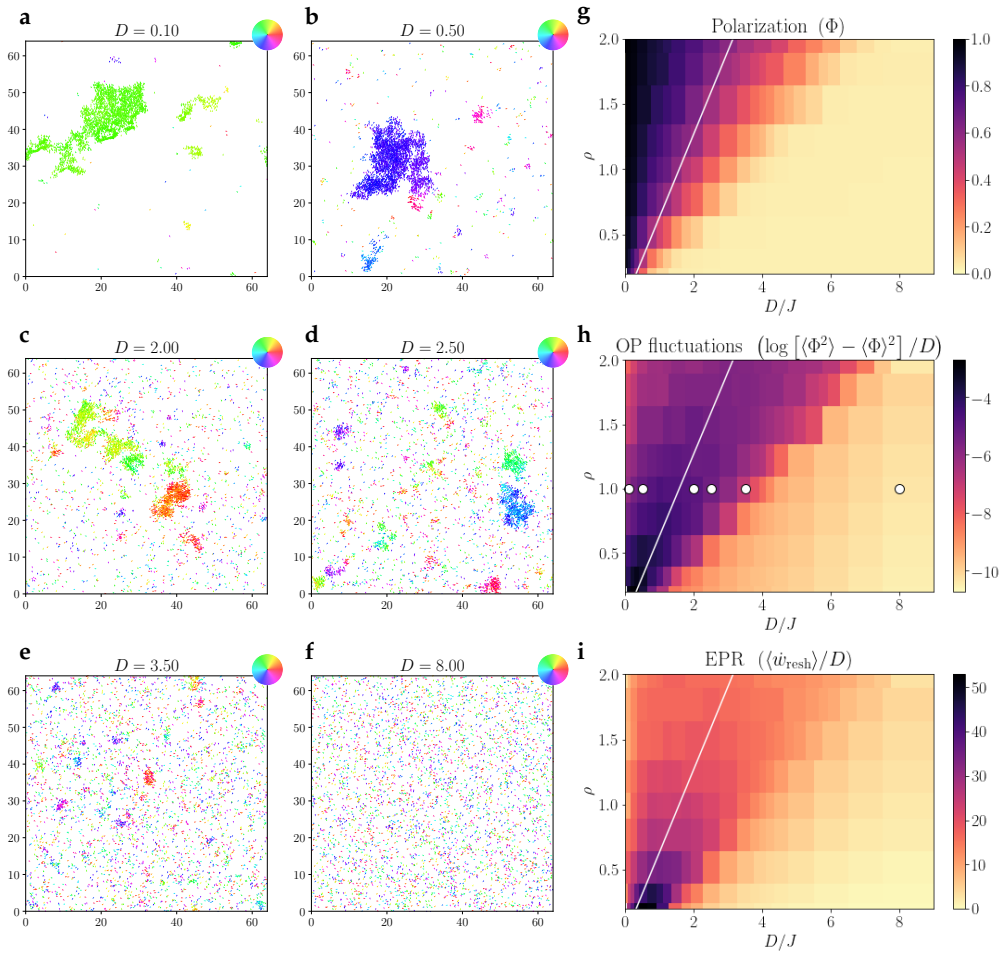

Figure 22: **a-f**: Snapshots of the SPP system (N=4096) in its (supposed) steady state, in various points of the phase diagram (marked by the white circles in Fig. h). Typical configurations in the ordered phase display dense polar clusters against a dilute isotropic background. In the deeply disordered phase, particles are oriented and displaced randomly. **g**: Order parameter $\Phi = \frac{1}{N}\left|\sum_{i=1}^{N} \mathbf{e}(\theta_i)\right|$. **h**: Phase diagram showing the quantity $\hat{\chi} = \frac{1}{D}\left(\langle\Phi^2\rangle - \langle\Phi\rangle^2\right)$, which can be considered as a proxy of the susceptibility ($\chi = -\frac{\partial\Phi}{\partial h}$, with h the magnitude of a magnetic field $\mathbf{h}$ that couples to the $\mathbf{v}_i$'s) at equilibrium. Notice that the color code is in log scale. **i**: Phase diagrams showing the entropy production rate computed as in Eq. (218). Figs. g-i have been obtained from simulations of $N = 1024$ SPPs observed over $5 \cdot 10^6$ integration steps. In all the plots the white line indicates the mean field transition line.

Global properties

Using the formulas in Eq. (215) and (218), we measured the entropy production rate across the phase diagram. Results are reported in Fig. 22. Control parameters are here the average density ρ_0 and the temperature of the thermal bath D, while we fix the value of the SPP speed to $v_0 = 0.5$. The interaction strength and the interaction range are set, respectively, to: $J = 1$, $R = 1$. Two equilibrium limits can be immediately identified, at the two extremes of the phase diagram: $J\rho_0/D \to 0$ and $J\rho_0/D \to \infty$.

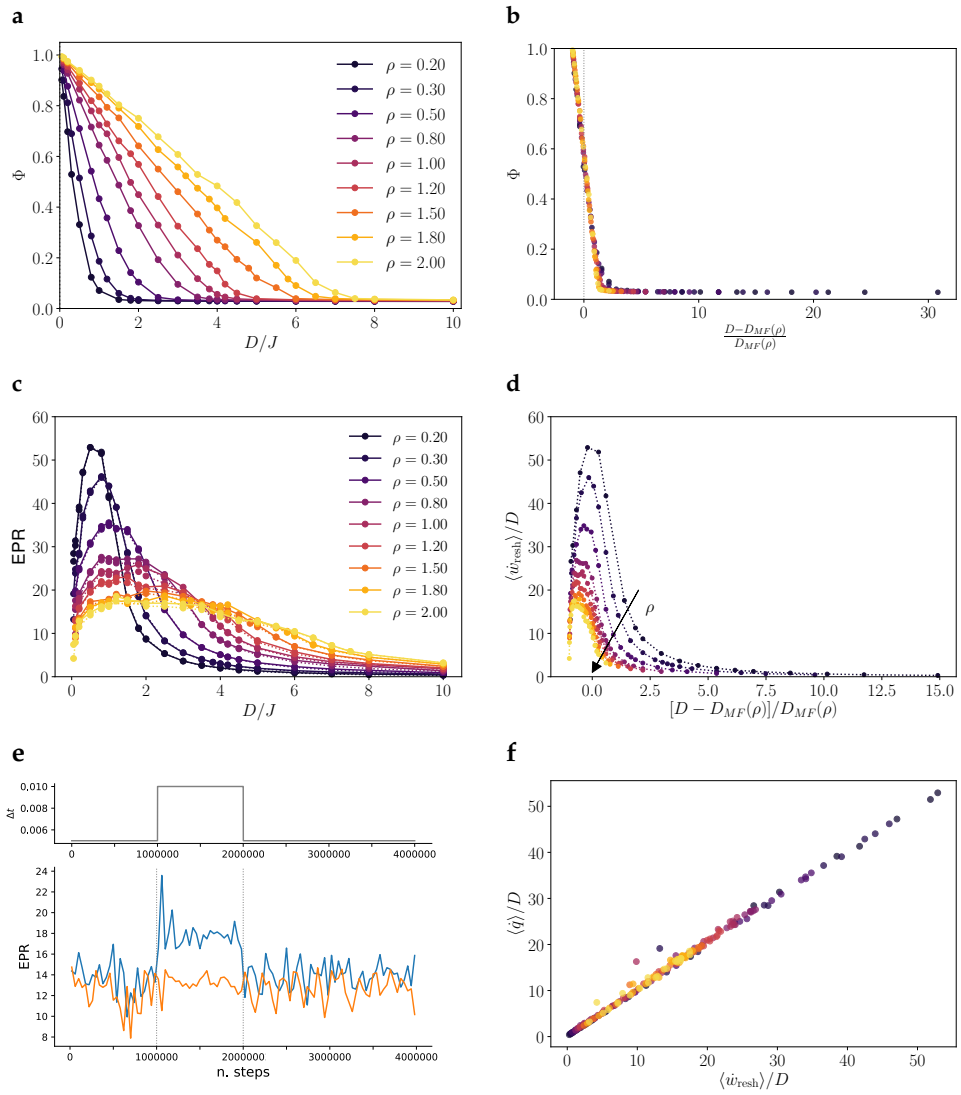

Figure 23: **a-b**: Polarization curves of a flock of $N = 1024$ birds as a function of the control parameter D/J and of the rescaled distance from the mean field critical value $D_{MF}(\rho_0) = J\pi\rho_0 R^2/2$. The mean field prediction works remarkably well, especially at higher densities, as expected. **c**: EPR estimates. Solid lines correspond to the dissipated heat interpretation — Eq. (215). Dashed lines correspond to the interpretation in terms of work of (fictitious) reshuffling forces — Eq. (218). **d**: Dashed curves are now plotted versus the rescaled distance from the mean field critical point $(D - D_{MF})/D_{MF}$. The collapse is not very good, especially for small ρ values, where the system is arguably farther from a mean field approximation. **e**: Sensitivity to the choice of the integration step Δt: blue lines are estimates of the EPR using Eq. (215), orange lines are estimates of the EPR using Eq. (218). There is a discrepancy between the two when Δt is too large, compared to the parameters of the process. Microscopic time scales are $\rho_0^{-1/2} v_0^{-1}$ and D — here $N = 1024$, $D = 5$, $\rho_0 = 1.8$. Notice that Eq. (218) (orange line) is numerically more reliable, even at a larger Δt. **f**: Parametric plot of the EPR mesured using Eq. (215)) and Eq. (218). The two quantities coincide when Δt is sufficiently small. Different colors indicate different densities, as in Fig. a.

In the non-interacting limit, $J\rho_0/D \to 0$, the system corresponds to an ideal gas of active Brownian particles where each particle can be seen as a passive tracer transported by a fluid [170]. The system is therefore in a situation of effective equilibrium[12]. The second limit, $J\rho_0/D \to \infty$ is less obvious and reveals that interactions do not bring the system out of equilibrium in a monotonic way. In a perfectly aligned system, torques do not make work, but it is not evident that the EPR must vanish, because of a D^{-1} factor in front of Eq. (215). Even using the equivalent formula (218), the argument is not conclusive: reshuffling is suppressed in the low-D limit, but we do not know how fast $\dot{n}_{ij}$ converges to zero. Our numerical simulations show that the system is asymptotically detailed balanced as $D \to 0$. Then a good reference equilibrium model in this region of the phase diagram is the XY ferromagnet (Heisenberg in $d = 3$), which has been used as an effective description for natural flocks of birds in the past years [136, 11].

From Figs. 22.g-i we see that the EPR maximum is approximately located at the mean field critical point (white lines), which provides a good estimate of the true transition point [51]. Polarization curves in Fig. 23.a-b show indeed a good collapse when plotted as a function of the distance of the control parameter D from the mean field critical value $D_{MF} = J\pi\rho_0 R^2/2$. On the contrary, EPR curves in Fig. 23.c-d, do not exhibit a perfect collapse on a single master curve. This fact suggests that the departure from equilibrium is not controlled by the polar order parameter alone. Let us notice that, since the number N of SPPs is kept constant for all the values of ρ_0, there might be some uncontrolled finite size effects. However, we believe that they are negligible, and that the decrease in non-equilibrium effects observed at larger ρ is just due to the fact that those systems are closer to a mean field situation. If the dynamics occurs at high densities on almost fully connected interaction graphs, rewiring (due to the motion of few bids on the boundaries of the interaction disk) acquires a lower and lower weight.

The final evidence from Figs. 23.c and 23.f is the agreement between dashed and solid curves, which respectively correspond to the quantities (218) and (215). This result provides a numerical validation of the hypothesis that the N-body system is in a stationary state, the only assumption we made use of.

Local origin of dissipation

As a function of N, the EPR seems to be an extensive quantity: this fact has been checked by varying the number of particles in numerical simulations, at fixed average density ρ_0 — see Fig. 24. The weighted divergence interpretation of Eq. (218) indicates indeed that heat is dissipated at a local scale — precisely the scale R of the interaction radius. Specifically, this occurs because a bird which is exiting the interaction disk (labeled j) has exchanged information with the bird in the middle

12 Let us notice that finding an equilibrium limit for $D \to \infty$ is a strict consequence of the absence of speed fluctuations. For fixed speed ABPs, the alignment force in the Langevin equation (211) is always bounded, irrespectively of D. On the contrary, if speed fluctuations are allowed, they typically depend on D in an unbounded way. In a model where the social interaction term is modified accordingly, in order to reproduce an imitation mechanism which is not limited to the flight direction but also involves the speed magnitude, we do not recover in the $D \to \infty$ limit an ideal gas behavior, and the system is in general out of equilibrium.

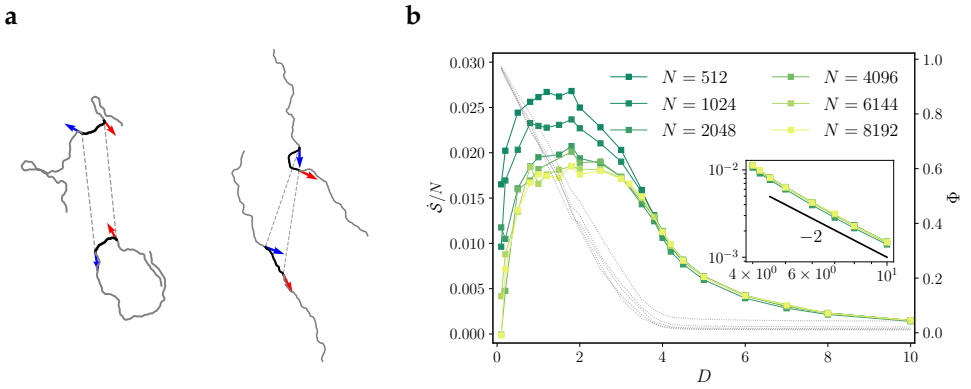

Figure 24: Positivity and extensivity of EPR. **a**: We illustrate the origin of the asymmetry between positive (red) and negative (blue) contributions to EPR. Pairs of particles which modify their status of neighbors contribute to the irreversibility of the flocking process (dashed lines indicate a distance equal to R). Pairs of particles that are exiting each other's interaction disk are typically more aligned than when they enter. For particles moving in approximately opposite directions (left), a large positive contribution to the EPR is associated to *entrance* events (red arrows, $\varepsilon_{in} = 1.76$); a smaller contribution is associated to *exit* events, where particles are less anti-aligned (red arrows, $\varepsilon_{out} = -0.92$). For pairs of particles approximately moving in the same direction (right), the EPR contribution tends to have a larger positive weight on exit events (red arrows, $\varepsilon_{out} = 0.38$ for the displayed configuration) than on entrance events (red arrows, $\varepsilon_{in} = -0.34$). On entrance events the two particles are typically less aligned because no direct interaction was occurring at previous times. The shown trajectories have been selected from numerical simulations with $N = 1024$, $D = 3$, $\rho_0 = 1$ for illustration purposes. We remark that the positivity of the EPR is only guaranteed on average on an ensemble of trajectories. The fact that all contributions come from pairs of particles at distance $R \ll L$ means that this mechanism is local ($R/L \to 0$ in the thermodynamic limit) and justifies the extensivity of the EPR. **b**: Numerical results for $\rho_0 = 1$, $v_0 = 1$, $J = 1$, and varying N.

(labeled i) and is typically more aligned to its velocity, compared to a bird j' which is entering now the interaction disk and has not interacted directly with i before. This mechanism (illustrated in Fig. 24.a) provides an explanation for several facts: (i) positivity of the EPR, as it assigns a bigger weight to positive EPR contributions than to negative ones; (ii) local origin of dissipation, jusifying the extensivity of the EPR, as confirmed by numerical results in Fig. 24.b; (iii) emerging asymmetries in the steady-state two-particle density $q(\alpha, \varphi)$, which we analyze in Section 4.2.3.

Comparison with MIPS irreversibility

Previous studies on the irreversibility of the motility-induced phase separation process have revealed that dissipation in apolar active matter is localized and particles tend to segregate in space according to their energetic role. It has been shown, both using microscopic AOUP models [130] and hydrodynamic theories [137], that the entropy production can be localized on interfaces, whereas the bulk is approximately at equilibrium. The decomposition of the global EPR as an integral of a local

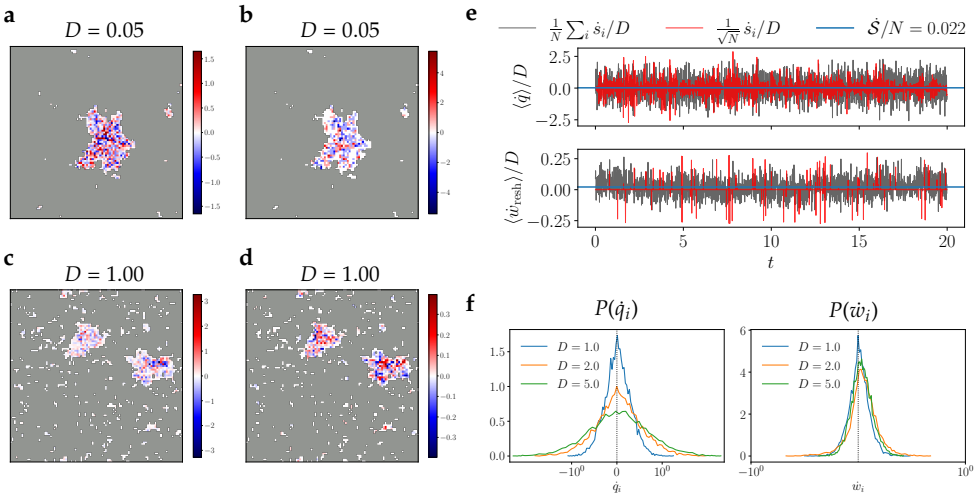

Figure 25: **a–d**: Snapshots showing the single-particle contribution to the EPR, $\dot{s}_i$. On the left (a,c) this is computed from Eq. (215) as $\dot{s}_i = \frac{\langle \dot{q}_i \rangle}{D} = -\frac{J}{D}\sum_j \langle \dot{\theta}_i \circ n_{ij} \sin(\theta_i - \theta_j) \rangle$. On the right (b,d) the single-particle contribution is computed from Eq. (218) as $\dot{s}_i = \frac{\langle \dot{w}_i \rangle}{D} = -\frac{J}{2D}\sum_j \langle \dot{n}_{ij} \circ \cos(\theta_i - \theta_j) \rangle$. In both cases no correlation between the spatial segregation of particles and their contribution to the EPR (positive/red or negative/blue) is observed. **e**: Time series of the average EPR of the flock and of a sample single particle contribution (properly rescaled). The two time series have comparable fluctuations, especially when the stochastic heat is considered. This fact shows that all the particles contribute in the same way to the global EPR. **f**: Probability density of the single-particle stochastic heat (left) or irreversible work of fictitious reshuffling forces (right). The histogram is unimodal, even when particles are organized into polar clusters, in clear contrast with the MIPS phenomenology [52].

field (in continuous theories) or as the sum of single-particle contributions (in agent-based models) is however non-unique, since local quantities are always defined up to an additive state function.

One of the single-particle quantities that characterize the irreversibility of the MIPS process is the active work per particle, i.e. the work that the active force exerts on single particles along a trajectory of length τ. Its probability distribution has been studied in the framework of the large deviation theory and it has been shown that it displays a marked bimodality in the phase-separated state, with corresponding non-convexity of the associated rate function [52]. The active work values corresponding to the minima of the rate function are typical of particles belonging to the bulk of the dense phase and to the dilute phase, respectively. Work values corresponding to the non-convex region are, on the contrary, typical of particles at the interface [52].

Motivated by such results, we try to understand whether a similar scenario is occurring in polar systems. We recall that the dissipation mechanism in our model is completely different form the one studied in [130, 52], where the self-propulsion force is treated as a T-even variable and irreversibility stems from the fact that in the forward trajectory the self-propulsion force pushes the particle, while in the time-reversed one it opposes the particle's motion. In our case, self-propulsion enters

through a reversible term, translational diffusion is absent, and the particles' motility is not altered by any excluded volume effects. Nonetheless, at low D, particles clusterize themselves and it is possible to seek for specific dissipative behaviors on the interface, the bulk, the front or the rear of the flock. From the analyses reported in Fig. 25, we deduce that this is not the case. These analyses consist in a direct inspection of the single particle contribution, computed from Eqs. (215) and (218) as $\dot{s}_i = \langle \dot{q}_i \rangle / D = -\frac{J}{D} \langle \sum_j \dot{\theta}_i \circ n_{ij} \sin(\theta_i - \theta_j) \rangle$ and $\dot{s}_i = \langle \dot{w}_i \rangle / D = -\frac{J}{2D} \langle \sum_j \dot{n}_{ij} \circ \cos(\theta_i - \theta_j) \rangle$, respectively. Looking at the movies, we could not observe any stable segregation of the flock into positively and negatively contributing assemblies of particles. on the contrary, we observed very large fluctuations of the local entropy production rate. A comparison between the time series of the average EPR (over the flock) and the single-particle contribution to the EPR (Fig. 25.e) suggests that particles do not have their own fixed energetic role, but this strongly fluctuates in time (as much as the global EPR does). Also the reconstructed probability distributions of $\langle \dot{q}_i \rangle / D$ and $\langle \dot{w}_i \rangle / D$ are unimodal rather then bimodal, in clear contrast with what has been observed in MIPS models [52].

We conclude that dissipation in polar flocks is due to a local, non localized mechanism.

4.2.3 The two-particle density

The most interesting insight derived from this analysis concerns the two-particle density. We have already observed that, in the considered active matter model, non-equilibrium stems from the combination of *interaction* and *self-propulsion*: in the absence of any of these two ingredients, the system falls in an equilibrium state. Since the interaction we implement is purely pairwise, signatures of non-equilibrium can be traced in the steady state two-particle density.

Given the distribution of the N-body system $\psi(\mathbf{\Gamma})$, where $\Gamma_i = (\mathbf{x}_i, \theta_i)$ and $\mathbf{\Gamma} = (\Gamma_1, \dots \Gamma_N)$ is a point in the system's phase space $[0, L]^{2N} \otimes [0, 2\pi]^N$, let us introduce the one-particle density

$$p_1(\Gamma_1) = \int d\Gamma_2 \dots d\Gamma_N \psi(\mathbf{\Gamma}), \tag{229}$$

and the two-particle density

$$p_2(\Gamma_1, \Gamma_2) = \int d\Gamma_3 \dots d\Gamma_N \, \psi(\mathbf{\Gamma}). \tag{230}$$

The two-particle distribution in Eq. (230) is linked to the quantity $q(\alpha, \varphi)$ introduced in Eq. (225) by the following relation:

$$\begin{aligned} q(\alpha, \varphi) = N \int & d\Gamma_1 d\Gamma_2 \, p_2(\Gamma_1, \Gamma_2) \, \delta_{2\pi}(\theta_2 - \theta_1 - \varphi) \\ & \cdot \delta_L \left(x_2 - x_1 - R \cos \left[\theta_1 + \varphi/2 + \alpha + \pi \Theta(\varphi \bmod (4\pi) - 2\pi) \right] \right) \\ & \cdot \delta_L \left(y_2 - y_1 - R \sin \left[\theta_1 + \varphi/2 + \alpha + \pi \Theta(\varphi \bmod (4\pi) - 2\pi) \right] \right), \end{aligned} \tag{231}$$

where $\delta_{2\pi}(\cdot)$ and $\delta_L(\cdot)$ are the periodically extended Dirac distributions. With this definition, $q(\alpha, \varphi)$ is 2π-periodic in both the arguments, and invariance under particles' permutations amounts to the symmetry of $q(\alpha, \varphi)$ with respect to φ, i.e.

$q(\alpha,\varphi) = q(\alpha, 2\pi - \varphi)$ in the $[0, 2\pi)$ domain. Moreover, the definition (231) corresponds to a simple operational procedure to compute $q(\alpha,\varphi)$ from the data: it is sufficient to firstly extract $\hat{\alpha}_{ij}$ and φ and work, for both angles, in the quotient set $[0, 2\pi)$; we can later compute α as $(\hat{\alpha} - \varphi/2) \bmod 2\pi$.

We have already seen that, in order to have nonzero EPR, $q(\alpha,\varphi)$ must satisfy $q(\alpha,\varphi) \neq q(\alpha+\pi,\varphi)$ for at least some sets of φ values of non-vanishing measure. Precisely, since the housekeeping entropy production is by definition non-negative, the asymmetries of $q(\alpha,\varphi)$ must be positively correlated to those of $\varepsilon(\alpha,\varphi)$. We define a 'two-particle quasi-potential' $u(\alpha,\varphi)$ associated to $q(\alpha,\varphi)$:

$$q(\alpha,\varphi) = N^2 g(R) e^{-u(\alpha,\varphi)} \tag{232}$$

and consider its T-symmetric and T-antisymmetric parts:

$$u_+(\alpha,\varphi) = \frac{1}{2}\left[u(\alpha,\varphi) + u(\alpha+\pi,\varphi)\right] \qquad u_-(\alpha,\varphi) = \frac{1}{2}\left[u(\alpha,\varphi) - u(\alpha+\pi,\varphi)\right]. \tag{233}$$

In terms of these quantities, the housekeeping EPR is computed as

$$\begin{aligned} \dot{\mathcal{S}} &= -\int d\alpha d\varphi\, e^{-u_+(\alpha,\varphi)} 2\sinh(u_-(\alpha,\varphi))\varepsilon(\alpha,\varphi) \\ &\simeq 2\int d\alpha d\varphi\, e^{-u_+(\alpha,\varphi)} \left(-u_-(\alpha,\varphi)\right)\varepsilon(\alpha,\varphi) \end{aligned} \tag{234}$$

where we exploit the assumption that $u_-(\alpha,\varphi)$ is small to go from the first to the second line. The quantities $q(\alpha,\varphi)$ and $-u_-(\alpha,\varphi)$ have been reconstructed from the numerical simulations and are shown in Fig. 26.

From a comparison between Figs. 21.b and 26, it is evident how the reconstructed $-u(\alpha,\phi)$ resembles a functional form of the kind $\lambda\varepsilon(\alpha,\varphi) + \mu\cos\varphi + \gamma$, with λ, μ, γ free parameters. We can empirically argue that most of the features of the two-particle distribution are thus captured by:

$$q(\alpha,\varphi) \sim e^{\lambda\varepsilon(\alpha,\varphi) + \mu\cos\varphi + \gamma}. \tag{235}$$

Recalling the maximum entropy construction, Eq. (235) reads as the Max-Ent distribution in the reduced two-body phase plane (α,φ) that is obtained by constraining the average values of the following physical observables: $\varepsilon(\alpha,\varphi)$ and $\cos(\varphi)$. These observables respectively correspond to measures of the irreversiblity and of the degree of local order at the scale of the interaction range. The final Lagrange multiplier γ comes from a normalization constraint. Let us recall that $q(\alpha,\varphi)$ is not normalized to one but to $Ng(R)$, where R is the interaction radius and

$$g(r) = \int d\Gamma_1 d\Gamma_2\, p_2(\Gamma_1,\Gamma_2)\,\delta(|\mathbf{r}_2 - \mathbf{r}_1| = r). \tag{236}$$

A more detailed analysis of the T-antisymmetric parts of the quasi-potential highlights the strong correlation between $u_-(\alpha,\varphi)$ and $\varepsilon(\alpha,\varphi)$. In Fig. 26 we plot the level curves of $\varepsilon(\alpha,\varphi)$ over the data obtained from our numerical simulations. There is remarkable agreement between the two, especially at intermediate angular diffusivities, $D \simeq D_{MF}$ (close to the mean field critical point), where the EPR peaks. At low D, the smaller modulations around $\varphi = \pm\pi/4$ are washed out in the data: when the system is strongly polarized, u_+ is much larger than u_-, and this might result into a

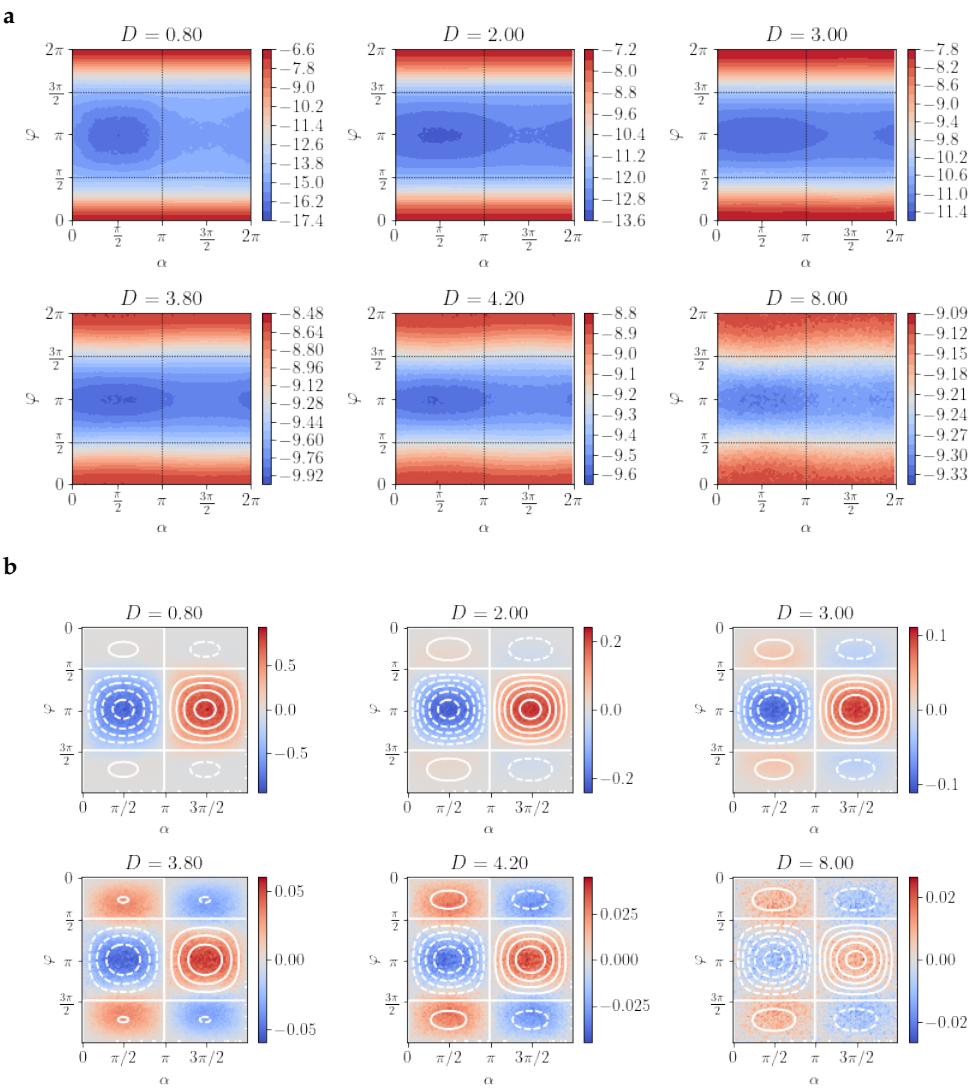

Figure 26: a: Minus two-particle quasipotential $-u(\alpha, \varphi) = \log q(\alpha, \varphi)$ reconstructed form numerical simulations ($N = 1024$). Precisely, $q(\alpha, \varphi)$ is estimated through a histogram that we build using all the particle pairs at a distance $R - \delta R < |\mathbf{r}_i - \mathbf{r}_j| < R + \delta R$, with $\delta R = 0.05$ ($R = 1$). We highlight the striking similarity between these figures and Fig. 21. Anti-symmetric part of the minus quasi-potential $-u_-(\alpha, \varphi) = \frac{1}{2}[\log q(\alpha, \varphi) - \log q(\alpha + \pi, \varphi)]$ for different values of the control parameter D. White curves indicate the level curves of the $\varepsilon(\alpha, \varphi)$ function.

systematic reconstruction error due to low statistics. In the large D limit the reconstruction of the T-antisymmetric quasi-potential $u_-(\alpha, \varphi)$ is dominated by statistical noise: as $D \to \infty$, we approach indeed an equilibrium limit, where we expect the probability density function to be symmetric, like in an ideal gas of ABPs. For a 3D representation of the reconstructed $u_-(\alpha, \varphi)$ and of their fit functions, see Fig. 27.

As regards the T-symmetric part of the quasi-potential, we reconstructed $u_+(\alpha, \varphi)$ from the simulations and fitted it to the truncated Fourier series of a symmetric φ-dependent function: $\sum_{k=0}^{4} a_k \cos(k\varphi)$. Despite there is a clear decay of the amplitude of Fourier coefficients with k, modes of order higher than 1 are needed in the ordered phase, where particles are strongly aligned and $u_+(\varphi)$ is too peaked around

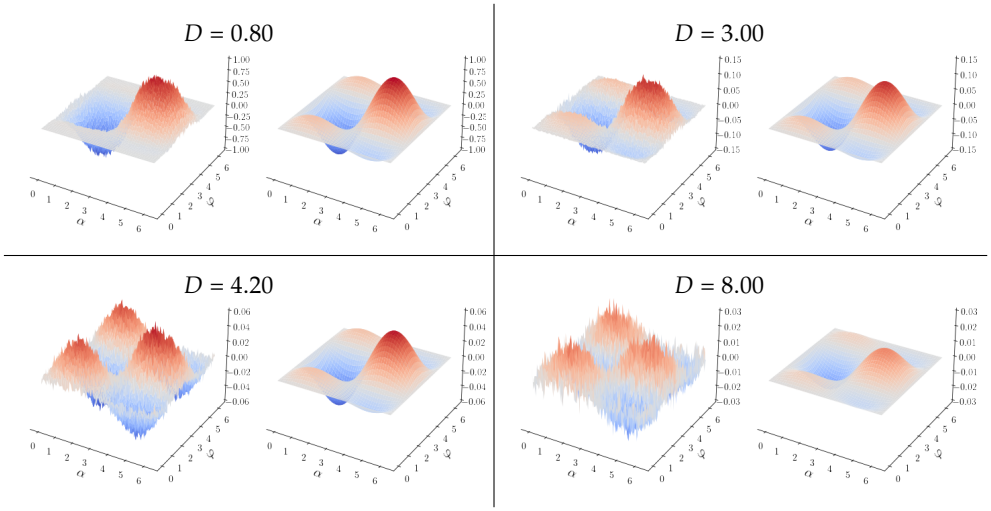

Figure 27: Three-dimensional representation of the numerically reconstructed function $-u_-(\alpha, \varphi)$ (on the left) and $\lambda\varepsilon(\alpha, \varphi)$ (on the right), with the value of λ obtained from a least square fit. The comparison between left and right-hand sides of each sub-figure reveals how good our guess — based on the heuristic interpretation in Eq. (235) — is. Residuals of this fit are shown in Fig. 28.

the origin to be described by a simple cosine function (see Fig. 28). We conclude that the heuristic functional form (235) works pretty well in the disordered phase, up to the transition, $D \gtrsim D_{MF}$. Notice that for both $D \gg D_{MF}$ and $D \ll D_{MF}$, a working equilibrium description holds and the role of the two-particle distribution becomes less important.

Finally, let us consider how this asymmetry in the pair distribution decays with the inter-particle distance r. Let us define $q(\alpha, \varphi; r)$ as in Eq. (231), but now for an

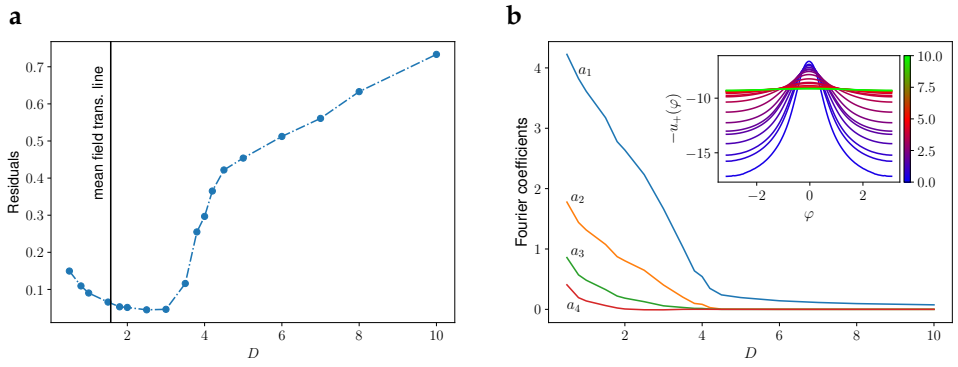

Figure 28: **a**: Residuals of the least square fit of the antisymmetric quasi-potential $u_-(\alpha, \varphi) \sim -\lambda\epsilon(\alpha, \varphi)$ shown in Figs. 26–27. Residuals are computed as $\sum_{ij}\left[u_-(\alpha_i, \varphi_j) + \lambda^*\varepsilon(\alpha_i, \varphi_j)\right]^2 / \sum_{ij}\left[u_-(\alpha_i, \varphi_j\right]^2$. The region where the fit works the best is the one where the EPR peaks and density heterogeneity is not too strong, in the near right of the mean field transition line. **b**: We fit the numerically reconstructed function $-u_+(\varphi) = -\int d\alpha\, u_+(\alpha, \varphi)$ (shown in the inset) with a truncated Fourier series $\sim \sum_{k=0}^{4} a_k \cos(k\varphi)$. The color code of the inset plot refers to the values of the D parameter. The fitted Fourier coefficients $a_1 \ldots a_4$ are plotted as functions of D.

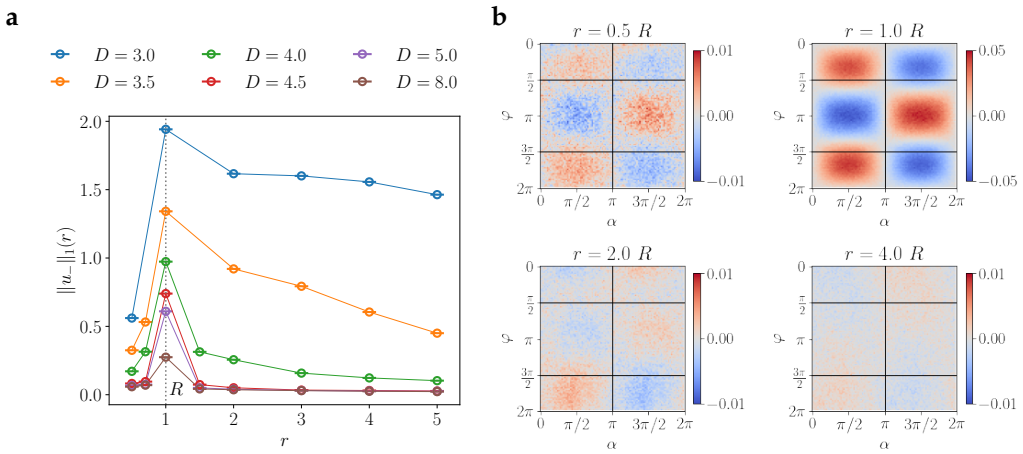

Figure 29: a: Scalar measure ($L^1([0,2\pi]^2)$ norm) of the asymmetry of the minus-quasi-potential, $u_-(\alpha,\varphi;r) = \log q(\alpha+\pi,\varphi;r) - \log q(\alpha,\varphi;r)$, for varying r and D. There is a clear peak at the interaction radius $r = R$ for all the considered values of the rotational noise variance. For lower values of D the decay to zero at large r is slower because of the presence of heterogeneous structures affecting the pair distribution. **b**: Color maps showing $-u_-(\alpha,\varphi;r)$ for $D = 4.5$. Notice the different scale for $r = R$ and $r \neq R$. All these analyses have been performed on systems of size $N = 2048$.

arbitrary r; from this function we can again define a quasi-potential and especially its T-antisymmetric part $u_-(\alpha,\varphi;r)$. In Fig. 29 we numerically reconstruct this function from our simulations and quantify its degree of asymmetry through the scalar measure

$$\|u_-\|_1(r) = \int_{[0,2\pi]^2} d\alpha d\varphi \, u_-(\alpha,\varphi;r). \tag{237}$$

When the particles' spatial distribution is not too anisotropic, this asymmetry measure presents a clear peak at $r = R$. This feature obviously depends on the precise parametrization of $n(r)$ we chose. We expect that, choosing alternative parametrizations, Eq. (237) will still give an indication of the shape of $-\partial_r n(r)$.

In summary, the pair distribution asymmetry under the time-reversal operator is not only a signature of the irreversibility of the flocking process, but can also be used to infer the nature of pairwise interactions in such high-dimensional off-equilibrium systems. This quantity is easy to reconstruct even with moderate statistics, and possibly provides a useful tool to analyze experimental datasets on diverse polar active systems.

4.2.4 Kinetic theories for the two-particle density

An interesting question is how to reproduce analytically the traits of the two-particle density discussed above. The task can be addressed using kinetic theory.

Kinetic theory approaches for active matter have been developed by adapting standard methods used in the statistical mechanics of passive systems. Let us notice that usually the interest is limited to the reconstruction of the single particle density $p_1(\Gamma_i)$, which describes the probability of finding the sample particle i in

the point Γ_i of its own phase space. This quantity is indeed sufficient to compute several average quantities of interest, which are defined as single-particle functions, like the average number density or the average kinetic energy. Such p.d.f., $p_1(\Gamma_i)$, can also be defined for N-body systems with a deterministic and time reversible dynamics; here, however, since the particles' dynamics is stochastic, a probabilistic description is already present at the microscopic level. The micro-state of the N-body system is described by the NESS density $\psi(\mathbf{\Gamma})$, and the single particle distribution, the pair distribution, and any other n-particle distribution are just obtained by marginalization.

There are several ways to model the microscopic stochastic dynamics, with as many different approaches to derive the kinetic equations. A first type of approaches consists of defining agent-based rules for the motion, diffusion and collision events which influence the evolution of each particle's state, and build from them a Boltzmann-like equation for the single particle p.d.f.. This is the case of the so-called BGL (Boltzmann-Ginzburg Landau) method [150, 7, 8], and of the Enskog-type kinetic theory [100].

Another possibility is to consider N-body ABP models, like Eq. (202), described by a set of SDEs in continuous time. This second option potentially allows us to answer broader questions, as we know the evolution of the N-body p.d.f. through the Fokker-Planck equation (205), whose marginalization yields the evolution equation for any probability density $p_n(\Gamma_1, \dots, \Gamma_n)$ with $n < N$. Kinetic theories derived using this second approach (also known as Smoluchowski approach) are therefore not limited to the reconstruction of p_1 and to the study of average single-particle observables, but may extend, for example, to average quantities depending on the two-particle density, like our entropy production rate. For this reason, we will focus on this second kind of derivation.

Formally, while the full FPE (205) is an ordinary PDE, the marginalized equation is a partial integro-differential equation, where partial derivatives of p_n mingle with collision integrals depending on higher p_m's ($m > n$). The procedure indeed provides us with a hierarchy of equations of the BBGKY kind. We report here the first two ones for the model in Eq. (211):

$$\mathcal{D}_t^{(1)} p_1(\Gamma_1) = J(N-1) \int d\Gamma_2 n_{12} \partial_{\theta_1} [\sin(\theta_1 - \theta_2) p_2(\Gamma_1, \Gamma_2)] + D \partial^2_{\theta_1 \theta_1} p_1(\Gamma_1), \quad (238)$$

$$\begin{aligned} \mathcal{D}_t^{(2)} p_2(\Gamma_1, \Gamma_2) = {} & J n_{12} \left\{ \partial_{\theta_1} \left[\sin(\theta_1 - \theta_2) p_2(\Gamma_1, \Gamma_2) \right] + \partial_{\theta_2} \left[\sin(\theta_2 - \theta_1) p_2(\Gamma_1, \Gamma_2) \right] \right\} \\ & + J(N-2) \int d\Gamma_3 \left\{ n_{13} \partial_{\theta_1} [\sin(\theta_1 - \theta_3) p_3(\Gamma_1, \Gamma_2, \Gamma_3)] \right. \\ & \left. + n_{23} \partial_{\theta_2} [\sin(\theta_2 - \theta_3) p_3(\Gamma_1, \Gamma_2, \Gamma_3)] \right\} + D \left(\partial^2_{\theta_1 \theta_1} + \partial^2_{\theta_2 \theta_2} \right) p_1(\Gamma_1). \end{aligned} \quad (239)$$

where $\mathcal{D}_t^{(1)} = \partial_t + v_0 \mathbf{e}(\theta_1) \cdot \nabla_1$ and $\mathcal{D}_t^{(2)} = \partial_t + v_0 \mathbf{e}(\theta_1) \cdot \nabla_1 + v_0 \mathbf{e}(\theta_2) \cdot \nabla_2$. The goal of any kinetic theory is to truncate this hierarchy by finding a suitable closed form approximation of the collision integrals [131].

Molecular chaos

The simplest possible closure is Boltzmann's molecular chaos hypothesis. Renaming g_n the n-body connected correlation function, the ansatz reads $g_2 = 0$. In other

words, the two-particle density factorizes into the product of single particle distributions, $p_2(\Gamma_1, \Gamma_2) = p_1(\Gamma_1)p_1(\Gamma_2)$, and $p_1(\Gamma_1)$ remains the only quantity to compute. This closure is equivalent to a mean field approximation and it is the most widely used to derive hydrodynamic equations for active matter — only one or two exceptions regard a very small number of recent works [146, 113, 53].

In order to simplify the scenario even further, we can start by assuming homogeneity. Having a system in a periodic box of linear size L, the homogeneity assumption amounts to:

$$p_1(\mathbf{x}, \theta) = \frac{1}{L^2}\phi(\theta). \tag{240}$$

The associated Boltzmann equation for $\phi(\theta)$ can be derived from Eq. (238) and easily studied by looking at the evolution of its Fourier series coefficients (details are reported in Appendix C for the sake of completeness). Numerical integration and linear stability analysis agree in predicting a transition from an isotropic phase to a polar ordered phase, occurring at the mean field transition point $D_{MF} = J\pi R^2\rho_0/2$.

Clearly, assuming that $g_2(\mathbf{x}, \theta) = 0$ and that the single particle distribution p_1 is of the form of Eq. (240), it is impossible to reproduce the desired asymmetry in the two-body p.d.f.. The same happens for any factorized p.d.f. of the kind

$$p_1(\mathbf{x}, \theta) = \rho(\mathbf{x})\phi(\theta). \tag{241}$$

If we try to compute the joint probability of a pair of particles at a distance $\mathbf{x}_2 - \mathbf{x}_1 = \mathbf{\Delta}$, i.e.

$$\tilde{p}_2(\theta_1, \theta_2, \mathbf{\Delta}) = \int d\mathbf{x}_1 d\mathbf{x}_1 p_1(\Gamma_1)p_1(\Gamma_1)\delta(\mathbf{x}_2 - \mathbf{x}_1 - \mathbf{\Delta}), \tag{242}$$

using the molecular chaos hypothesis and Eq. (241), we obtain indeed a T-symmetric p.d.f.:

$$\tilde{p}_2(\theta_1, \theta_2, \mathbf{\Delta}) = \phi(\theta_1)\phi(\theta_2) \int d\mathbf{x}_1 \rho(\mathbf{x}_1)\rho(\mathbf{x}_1 + \mathbf{\Delta}) = \tilde{p}_2(\theta_1, \theta_2, -\mathbf{\Delta}). \tag{243}$$

A first step forward may be to allow for spatial and angular coordinates to be interdependent; besides, even in standard BGL approaches, Fourier coefficients which depend on the spatial degrees of freedom are used. However, since the molecular chaos ansatz is a closure for p_1, the resulting kinetic theory is not supposed to provide in general a good prediction for the two-particle p.d.f.. We also tried to assess the degree of mutual dependence of the single particle degrees of freedom, $\mathbf{x}$ and θ, by measuring their mutual information from numerical simulations. We found that their mutual dependence is always very low, in all the regimes of interest, even though a little difference is visible when moving from the disordered to the ordered phase (additional details in App. C). This fact justifies the hypothesis that the single-particle distribution $p_1(\mathbf{x}, \theta)$ is factorized in its phase and space components.

An alternative route is to include pair correlations in the kinetic theory, as already done by Ihle and coworkers in [53] and in [113] for two variants of the Vicsek model, in discrete time with random binary collisions and in continuous time with non-additive interactions, respectively. We discuss this second option in the next paragraph.

Ring-kinetic theory

The ring-kinetic closure consists in setting the connected part of the three-particle distribution equal to zero, $g_3 = 0$, so that the hierarchy is consistently truncated at the level of the second equation — Eq. (239). In this way, correlations between pairs of colliding particles, which originate from previous collision events, are taken into account, but not higher order ones. The inclusion of such correlations is particularly relevant to describe dense systems, or systems with strong density fluctuations. Indeed, we recall that the molecular chaos ansatz typically comes along with a diluteness assumption, requiring that the time between subsequent collisions is longer than the time needed for single particle velocity distributions to decorrelate. Unlike standard Vicsek-like models[13], the model we consider has additive interactions, which are responsible for the formation of very dense clusters in the whole symmetry-broken phase (Fig. 22). The presence of these dense structures then invalidates the molecular chaos assumption in a big region of the parameter space.

Let us focus on our additive case, where closed integro-differential equations for $p_1(\Gamma_1)$ and $g_2(\Gamma_1,\Gamma_2) = p_2(\Gamma_1,\Gamma_2) - p_1(\Gamma_1)p_1(\Gamma_2)$ can be straightforwardly derived. However, these equations are intricate and in general hard to solve. Following [113], a great simplification is to invoke again the assumption of homogeneity for p_1: in this setting, the anisotropy of the two-body p.d.f. will then be entirely carried by the connected part of the two-particle density g_2. The condition of homogeneity on p_1 may seem restrictive, since it is verified only in the disordered phase, far from the transition, but it allows us to make some analytical progress by working with the reduced functions $\phi(\theta)$ — from definition (241) — and $\tilde{g}_2(\theta_1,\theta_2,\mathbf{\Delta})$, which is defined as:

$$\tilde{g}_2(\theta_1,\theta_2,\mathbf{\Delta}) = L^2 \int d\mathbf{x}_1 d\mathbf{x}_2 \delta(\mathbf{x}_2 - \mathbf{x}_1 - \mathbf{\Delta}) g_2(\Gamma_1,\Gamma_2) = L^2 \tilde{p}_2(\theta_1,\theta_2,\mathbf{\Delta}) - \phi(\theta_1)\phi(\theta_2). \tag{244}$$

The evolution equations for $\phi(\theta)$ and $\tilde{g}_2(\theta_1,\theta_2,\mathbf{\Delta})$ are obtained from Eqs. (238)–(239) using Eqs. (241) and (244) and taking the thermodynamic limit: $N \to \infty$, $L \to \infty$, $N/L^2 = \rho_0$. If we also rescale the time in order to transform it into a dimensionless variable, $t \to Dt$, we find:

$$\begin{aligned} \partial_t \phi(\theta_1) = {} & \frac{2\nu}{\pi R^2} \int d\mathbf{\Delta}\, n_{12}(\Delta) \partial_{\theta_1} \left[\sin(\theta_1 - \theta_2) \tilde{g}_2(\theta_1,\theta_2,\mathbf{\Delta}) \right] \\ & + 2\nu \int d\theta_2\, \partial_{\theta_1} \left[\sin(\theta_1 - \theta_2) \phi(\theta_1) \right] \phi(\theta_2) + \partial^2_{\theta_1\theta_1} \phi(\theta_1); \end{aligned} \tag{245}$$

13 Of course, going beyond the molecular chaos approximation is helpful to derive a more accurate kinetic theory also for Vicsek-like models with non-additive interactions, where phase coexistence and pattern formation take place across the transition point, even though in different forms. This has been done in [53, 113].

$$
\begin{aligned}
&\mathcal{D}_t^{(1-2)}\tilde{g}_2(\theta_1,\theta_2,\mathbf{\Delta}) \\
&= \frac{J}{D} n_{12}(\Delta)\left(\partial_{\theta_1}-\partial_{\theta_2}\right)\left\{\sin(\theta_1-\theta_2)\left[\tilde{g}_2(\theta_1,\theta_2,\mathbf{\Delta})+\phi(\theta_1)\phi(\theta_2)\right]\right\} \\
&\quad + 2\nu\int d\theta_3\,\phi(\theta_3)\left\{\partial_{\theta_1}\left[\sin(\theta_1-\theta_3)\tilde{g}_2(\theta_1,\theta_2,\mathbf{\Delta})\right]+(1\leftrightarrow 2)\right\} \\
&\quad + \frac{2\nu}{\pi R^2}\int d\theta_3 d\mathbf{h}\,\left\{n_{13}(h)\partial_{\theta_1}\left[\sin(\theta_1-\theta_3)\phi(\theta_1)\right]\tilde{g}_2(\theta_2,\theta_3,-\mathbf{\Delta}+\mathbf{h})+(1\leftrightarrow 2)\right\} \\
&\quad + \left(\partial^2_{\theta_1\theta_1}+\partial^2_{\theta_2\theta_2}\right)\tilde{g}_2(\theta_1,\theta_2,\mathbf{\Delta}).
\end{aligned}
\tag{246}
$$

with $\mathcal{D}_t^{(1-2)}$ the material derivative of $\tilde{g}_2$, defined as:

$$
\mathcal{D}_t^{(1-2)} = \partial_t - \frac{v_0}{D}\left[(\cos\theta_1-\cos\theta_2)\,\partial_{\Delta_x}+(\sin\theta_1-\sin\theta_2)\,\partial_{\Delta_y}\right]. \tag{247}
$$

The phrase $(1\leftrightarrow 2)$ indicates a term of the same kind as the previous one in the brackets, where particles 1 and 2 are exchanged (see App. C).

What we look for is a steady state solution for Eqs. (245)–(246). Given the structure of the equations, it is helpful to exploit the periodicity of the domain, both in the angular and spatial variables, and focus on the coefficients of the associated Fourier series:

$$
\phi(\theta) = \sum_k A_k e^{ik\theta}; \qquad \tilde{g}_2(\theta_1,\theta_2,\mathbf{\Delta}) = \sum_{k,l,m,n} F_{k,l,m,n} e^{ik\theta_1+il\theta_2+im\frac{2\pi}{L}\Delta_x+in\frac{2\pi}{L}\Delta_y}. \tag{248}
$$

Thanks to the simple step form of $n(r)$ we consider, the resulting equations for A_k and $F_{k,l,m,n}$ are easily derived (see App. C). They consist of an infinite set of coupled ODEs for the Fourier modes, with symmetries and constraints guaranteeing that:

- $A_k^* = A_{-k}$ and $F^*_{k,l,m,n} = F_{-k,-l,-m,-n}$, since ϕ and $\tilde{g}_2$ are real-valued functions;
- $F_{k,l,m,n} = F_{l,k,-m,-n}$, reflecting the invariance of the functions p_2 and g_2 under particles' labels permutation $(1\leftrightarrow 2)$;
- $\partial_t A_0 = 0$ and $\partial_t F_{0000} = 0$. As $p_1(\Gamma)$ and $p_2(\Gamma_1,\Gamma_2)$ are probability densities, they must be normalized. From the normalization of p_1 we deduce that $A_0(t) = \frac{1}{2\pi}$ at any t, while the normalization of p_2 requires $\int d\Gamma_1 d\Gamma_2\, g_2(\Gamma_1,\Gamma 2) = 0$, hence $F_{0000} = 0$ exactly.

Moreover, the connected two-particle density $g_2(\Gamma_1,\Gamma_2)$ must satisfy, by definition,

$$
\int d\Gamma_1\, g_2(\Gamma_1,\Gamma_2) = \int d\Gamma_2\, g_2(\Gamma_1,\Gamma_2) = 0, \tag{249}
$$

hence $F_{k000} = F_{0k00} = 0$ for any k. This condition is only ensured if we work in the thermodynamic limit, where these Fourier coefficients do not evolve and g_2 satisfies the normalization condition at any time. From now on we will always be implicitly referring to the thermodynamic limit.

The resulting evolution equations for the tensors of Fourier coefficients $\mathbf{A}$ and $\mathbf{F}$ are of the form:

$$
\partial_t \mathbf{A} = f(\mathbf{A}) + \mathcal{L}\mathbf{F} \tag{250}
$$

$$
\partial_t \mathbf{F} = \mathcal{M}\mathbf{F} + \mathcal{C}(\mathbf{A}) + \mathcal{H}(\mathbf{A})\mathbf{F} \tag{251}
$$

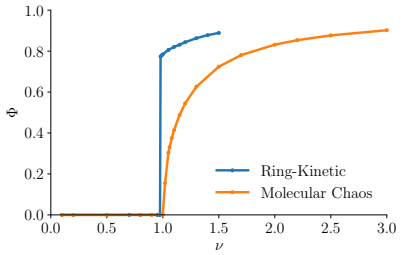

Figure 30: Order parameter $\Phi = 2\pi\Re A_1$ as a function of the control parameter of the kinetic theory ν. We compare the result of the two closures: molecular chaos ($g_2 = 0$) and ring-kinetic ($g_3 = 0$) with homogeneity assumption. For larger values of ν than those shown in the plot, the Fourier coefficients diverge, in the homogeneous ring-kinetic theory.

where $\mathcal{L}$, $\mathcal{M}$ and $\mathcal{H}$ are linear operators, some of which depend on $\mathbf{A}$ as indicated. The explicit expression of Eqs. (250)–(251) reads (see also App. C):

$$(\partial_t + k^2)A_k = \nu 2\pi k(A_{k-1}A_1 - A_{k+1}A_{-1}) + \nu 2\pi k \sum_{mn} K_{mn}\left[F_{k-1,1,m,n} - F_{k+1,-1,m,n}\right]; \quad (252)$$

$$\begin{aligned}\partial_t F_{k,l,m,n} + \frac{v_0}{2D}\frac{2\pi}{L}\left[(im+n)\left(F_{k+1,l,m,n} - F_{k,l+1,m,n}\right)\right.\\ \left. + (im-n)\left(F_{k-1,l,m,n} - F_{k,l-1,m,n}\right)\right] = \nu 2\pi k\left(F_{k-1,l,m,n}A_1 - F_{k+1,l,m,n}A_{-1}\right)\\ + \nu 2\pi l\left(F_{k,l-1,m,n}A_1 - F_{k,l+1,m,n}A_{-1}\right) + \nu K_{m,n}\left[2\pi k\left(A_{k-1}F_{1,l,m,n} - A_{k+1}F_{-1,l,m,n}\right)\right.\\ \left. + 2\pi l\left(A_{l-1}F_{k,1,-m,-n} - A_{l+1}F_{k,-1,-m,-n}\right)\right] - (k^2+l^2)F_{k,l,m,n}.\end{aligned} \quad (253)$$

Here ν is a dimensionless control parameter defined as

$$\nu = \frac{J\rho_0}{2D}\pi R^2. \quad (254)$$

Numerical integration of Eqs. (252)–(253) indicates that the system has a transition from the isotropic homogeneous phase to the polar ordered phase, occurring at ν^* slightly smaller than 1 (mean field critical value). The presence of two-body correlations in the kinetic theory facilitates the transition and transforms the mean field polarization curve into a steeper curve, which more closely resembles the behavior of the order parameter of a first order phase transition. A comparison of the two curves is shown in Fig. 30.

As regards the reconstruction of the connected two-body p.d.f., the adopted ring-kinetic theory is less satisfactory. First of all, the $\mathbf{F}$ coefficients do not seem to always reach a stationary stable fixed point over the observed temporal evolution (which is pretty long — at least more than 20 times the intervals considered for the mean field theory). Moreover, the two-body p.d.f. that we reconstruct from the (apparently stationary) numerical solutions of Eqs. (252)–(253) does not have the same features observed in the agent-based simulations of the model (see App. C). A possible explanation for this fact is that, since g_2 is the only $\mathbf{x}$-dependent function, it might be undertaking not only the description of the two-particle (connected) correlations, but also the description of spatial heterogeneities. These features should be already visible at the single-particle level, but are inevitably missed with the homogeneous

parametrization we employ. In other words, the g_2 function may be aiming here at reproducing two effects: the one we are interested in, and another one coming from a compensation for the wrong description of the marginalized p.d.f. of lower order.

Overall, these negative results suggest that the approach is not well-suited to the considered problem. The assumptions $g_3 = 0$ and that of homogeneity, despite crucial to make analytical progress, are not physically motivated in the considered system. A more detailed discussion of our numerical attempts can be found in App. C: further investigation is however in progress at the moment of writing.

Non-additive ring-kinetic theory reproduces the T-asymmetry

As a final remark, we would like to comment on the results recently obtained by Kurtsen and Ihle in the non-additive case [113]. The ring-kinetic theory developed by the authors is based on a variant of the ABP model (202) which considers non-additive, non-momentum-conserving interactions. Precisely, in this variant of the model, the metric binary adjacency matrix we adopted is replaced by $n_{ij} = \left[\sum_j \Theta(R - |\mathbf{x}_i - \mathbf{x}_j|)\right]^{-1} \cdot \Theta(R - |\mathbf{x}_i - \mathbf{x}_j|)$. With this choice of n_{ij}, collisions have in fact a multi-particle nature. Nonetheless, the authors managed to compute weights of the various collision integrals and derive a final set of equations which are now nonlinear in $\mathbf{F}$. This feature seems to stabilize the system of ODEs, which they have numerically integrated.

In addition to the quantitative agreement of their theory with the reference non-additive microscopic model, the two-particle density that Kursten and Ihle reconstruct is also in qualitative agreement with our predictions for the additive model. Precisely, they define from the connected two-body p.d.f. the quantity

$$h(r, \hat{\alpha}, \varphi) = \int d\mathbf{\Delta} d\theta_1 d\theta_2 \delta(\Delta_x - r\cos(\hat{\alpha}+\theta_1))\delta(\Delta_y - r\sin(\hat{\alpha}+\theta_1))\delta(\theta_2 - \theta_1 - \varphi)\tilde{g}_2(\theta_1, \theta_2, \mathbf{\Delta}) \tag{255}$$

and they analyze it at the onset of polar order. This function exhibits a significant and recognizable asymmetry at a distance equal to the interaction radius, $r = R$, while the same asymmetry is much reduced for $r > R$ or $r < R$. This feature reveals that pairs of particles tend to be more diverging than converging, when they are quite aligned, and more converging than diverging, when they are close to a completely anti-aligned configuration (cfr. Fig. 31).

Let us notice that we made exactly the same predictions on the basis of pure thermodynamic arguments in Fig. 20, exploiting only the EPR formula and the irreversibility of the process. It is worth remarking that the entropy production rate of the non-additive model considered by Kursten and Ihle is not exactly given by Eq. (218) (while Eq. (215) still holds): in order to compute the EPR for this model, in addition to the contribution of reshuffling, one should also include a contribution coming from the presence of irreversible currents associated to a lack of symmetry in the interaction ($n_{ij} \neq n_{ji}$). However, it is reasonable to assume that this effect is small, unless enormous and very sharp density fluctuations form within the system. A comparison between the entropy production rate of additive versus non-additive Vicsek-like ABP systems may help clarifying how different the non-equilibrium properties of these two variants of the model are.

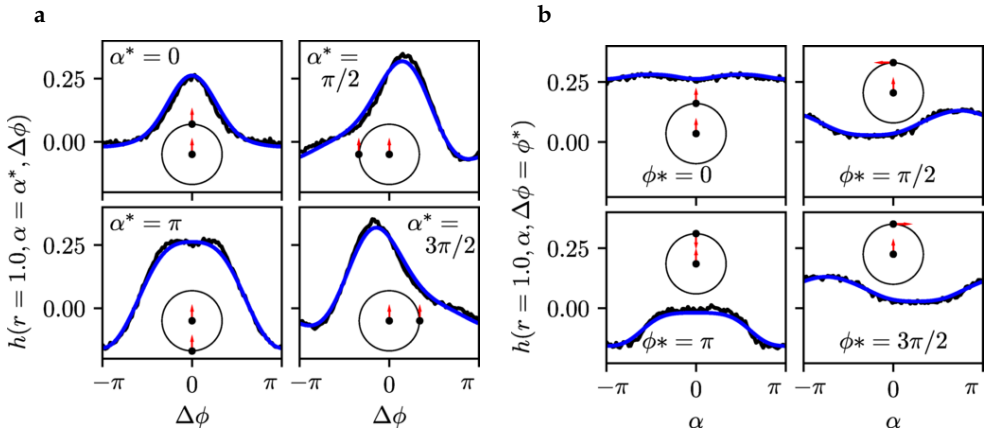

Figure 31: Figures taken from [113], representing $h(r, \hat{\alpha}, \varphi)$ at a distance equal to the interaction radius ($R = 1$). In the notation of Kursten and Ihle, α corresponds to our $\hat{\alpha}$ and $\Delta\phi$ to our φ. The black bullets indicate the relative positions of particle 1 (in the center of the circle) and particle 2 (on the circumference). **a**: Two-particle correlation function h as a function $\Delta\phi$ for selected values of $\alpha = \alpha^*$. The red arrows indicate the direction of the particle velocities at $\Delta\phi = 0$. When $\alpha^* = \pi/2$ or $\alpha^* = 3\pi/2$, the second arrow preferentially points outward. **b**: Two-particle correlation function h as a function α for selected values of $\Delta\phi = \phi^*$. The relative positions of the bullets is $\alpha = 0$. From the panels corresponding to $\phi^* = \pi/2$ and $\phi^* = 3\pi/2$, we see that peaks in the distributions are approximately located at values of α which maximize the outward projection of the red arrows along the segment connecting the bullets. Perfectly aligned particles ($\phi^* = 0$) are approximately uniformly distributed on the circle. On the contrary, completely anti-aligned particles are more likely colliding than getting apart from each other.

4.2.5 The Voronoi topological case

Model definition and numerical results

We investigate numerically another variant of the previous model, Eq. (211), where the interaction neighborhood of each bird is defined by means of a global Voronoi tessellation of the periodic domain:

$$n_{ij} = \begin{cases} 1 & \text{if} \quad i, j \text{ are centers of neighboring Voronoi cells,} \\ 0 & \text{otherwise.} \end{cases} \tag{256}$$

This kind of topological model exhibits a continuous phase transition from the isotropic state to a polar ordered state where the symmetry is spontaneously broken. In contrast to the previous case, density fluctuations do not cause an increase in the average number of neighbors per bird (always equal to 6 on the plane), so that the feedback mechanism leading to the segregation of dense ordered structures is not at play. Nonetheless, long-range order is established in systems of finite density [87].

The entropy production rate can be computed even in this case using Eqs. (215) and (218); however, the interaction matrix n_{ij} is no more a simple function of $|\mathbf{r}_i - \mathbf{r}_j|$, since determining whether two birds are Voronoi neighbors or not depends on the whole local arrangement of their positions. We apply these formulas to

measure irreversibility from numerical simulations of the model. As in the metric case, we checked that the formulas give equivalent estimates when the steady state is reached, so we only show results corresponding to Eq. (218) in Figs. 32–33.

The EPR curves exhibit a sharper peak than in the metric case, which is located at the transition between the isotropic and polar ordered phases (see Fig. 32). This is the point where reshuffling of the Voronoi network is mostly efficient. Here noise and alignment strength are balanced, so that the motion of the self-propelled particles does have enough persistence to prevent that they swirl around a fixed position, yet they do not move all in the same direction and can escape their neighborhood on relatively short time scales. For an illustration, see the sample trajectories reported in Fig. 33.

Quantifying network reshuffling

The fact that reshuffling is mostly efficient at the transition can be observed from a comparison of the auto-correlation curves of the adjacency matrix at different temperatures:

$$C_{\text{net}}(t) = \frac{1}{N(N-1)} \sum_{ij} \langle n_{ij}(t_0 + t) n_{ij}(t_0) \rangle_{t_0}, \tag{257}$$

where $\langle \cdot \rangle_{t_0}$ denotes a time average over multiple starting times. The function form of $C_{\text{net}}(t)$ is quite well described by the following formula:

$$C(t) \sim M(1 + ct^a)^{-d}. \tag{258}$$

Eq. (258) was empirically introduced in [32] to measure the *neighbor overlap* in real flocks of birds. Due to the binary nature of n_{ij}, $C_{\text{net}}(t)$ indicates the average fraction of neighbors at time $t_0 + t$ which were also neighbors at time t_0. With a crude approximation, the number of non-changing neighbors of a focal bird i can be computed assuming that the mean square distance travelled by its neighbors is $l^2 \sim t^a$ and that the diffusion is isotropic. The original M neighbors of bird i will therefore occupy a volume proportional to $(R + \hat{c}t^{a/2})^d$ at time $t_0 + t$, where R is the the radius of the volume occupied at time t_0 and $\hat{c}$ a diffusion constant. Therefore the number of non-changing neighbors is approximately estimated as the ratio between these two volumes, from which Eq. (258) derives. Let us notice that, since we only work in $d = 2$, both d and M are not parameters to fit: thanks to Euler's formula for planar graphs, the average degree of a Voronoi vertex is always $M = 6$.

A fit of Eq. (258) with free parameters a and c is shown in Fig. 33.d: the agreement is good, given the small number of fitting parameters. Both the fitted values of the rescaled diffusion constant $c = \hat{c}/R$ and exponent a show a positive correlation with the EPR of the Voronoi flock (parametric plot in Fig. 33.e), indicating that the faster is the decorrelation of the reshuffling network, the farther from equilibrium the system is. However, the fitted parameters must not be interpreted as the exact coefficients governing birds' mutual diffusion, especially in the ordered phase (low D), where the isotropy assumption through which Eq. (258) was derived breaks down and the mutual diffusion is almost one-dimensional. Moreover, reshuffling of the n_{ij}

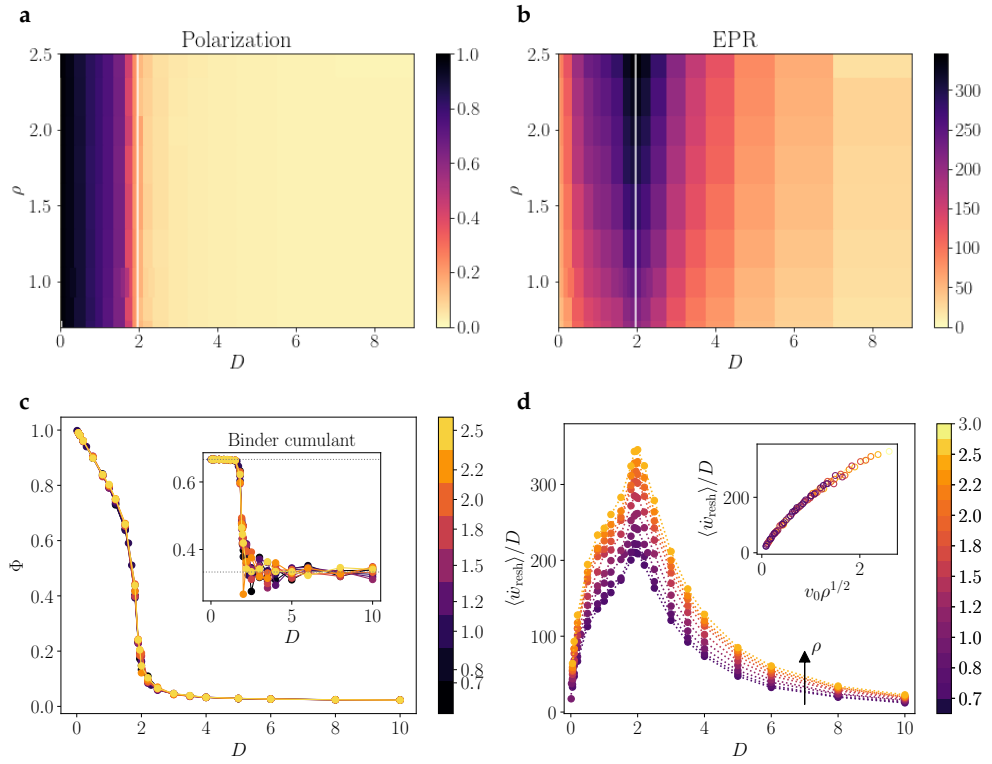

Figure 32: **a**: Order parameter Φ on the plane (D, ρ). The white line indicates the alleged critical point of the model. **b**: Entropy production rate across the phase diagram of the topological Vicsek model (computed using Eq. (218)). Irreversibility clearly peaks at the transition in this case. **c**: Polarization curves of the Voronoi ABP model as a function of the noise amplitude D, for different values of ρ (indicated by color code): without any rescaling, all of them are already overlapped on a single master curve. The Binder parameter in the inset hints at the presence of a second order phase transition. **d**: EPR curves as a function of D. The trend with ρ (color code) is opposite to the one observed in the metric variant (Fig. 23). This trend can be rationalized on the basis of dimensional arguments, according to which reshuffling in the Voronoi topological case is purely ruled by the combination $v_0 \rho^{1/2}$, as shown in the inset. All the figures are realized from simulations of $N = 2048$ boids, with average density $\rho = 1$.

matrix involves multiple scales and the argument does not take into account that the mutual diffusion exponent is, in general, non unique. A direct inspection of the mutual diffusion curves in Fig. 33.c reveals that, on the spatio-temporal scales allowed by the periodic boundary conditions, the diffusion exponents are pretty different from those extracted from the fit, with the EPR maximum point corresponding to a separating super-diffusive regime between the normal diffusive one (for $D > D_c \simeq 2$) and a ballistic one (for $D < D_c$).

Heuristic explanation of the EPR phase diagram

As expected, the transition point in the studied Voronoi ABP model is independent of the density of the flock, and almost no finite size effects are observed as N is varied. The EPR scales with N in an approximately linear way, indicating that entropy production is an extensive quantity and time reversal symmetry breaking

occurs at a local scale. The rescaled curves in Fig. 33.a collapse into a single master curves, corresponding to the EPR per unit bird, showing an unmistakable power law dependence in the low-temperature and high-temperature regions of the phase diagram: $\dot{\mathcal{S}}/N \sim D^{\sigma}$, with exponents $\sigma_{\text{low\_}D} \simeq 0.5$ and $\sigma_{\text{high\_}D} \simeq -2$.

A simple argument explains such scaling. Let us adopt the EPR formula (218), where $\dot{\mathcal{S}}$ is interpreted as the heat dissipated in reshuffling. The parameter D enters in Eq. (218) through the D^{-1} coefficient, through the birds' alignment $\cos(\theta_i - \theta_j)$ and through the reshuffling rate $\dot{n}_{ij}$. We approximately estimate these terms by means of their average magnitude. In particular, let us take $\langle \cos(\theta_i - \theta_j) \rangle \sim C(l)$, with C the (full) correlation function , which we assume isotropic, and l the average distance of a pair of birds which are leaving each other's Voronoi shell. As regards $\dot{n}_{ij}$, this amounts to estimating the rate of reshuffling events. We assume that the typical reshuffling rate must be related to the time needed for a bird to travel the same reference distance l.

For low D, the system is deeply ordered, so we can work in the spin-wave approximation (see Sec. 2.2.3) and only consider transverse velocity fluctuations $\boldsymbol{\pi}_i$, whose magnitude scales as $|\boldsymbol{\pi}| \sim \sqrt{1-\Phi} \sim D^{1/2}$. Mutual diffusion occurs mainly in the perpendicular direction to the collective motion of the flock. Since the rotational noise is little, we can assume that transverse motion is ballistic on the scale of the average inter-particle distance l, and that the condition $\langle \Delta \mathbf{x}_{\perp}^2(t) \rangle \sim \boldsymbol{\pi}^2 t^2 \sim l^2$ identifies a reshuffling time scale $\tau_{\text{resh}} \sim l/D^{1/2}$. Hence $\dot{n}_{ij} \sim \tau_{\text{resh}}^{-1} \sim D^{1/2}$. The correlation function is dominated in this limit by the non-connected part or, in other words, by the contribution of the parallel projections on the direction of collective motion. A low-D expansion is $C(l) \sim 1 - |\boldsymbol{\pi}|^2 \sim 1 - \alpha D$. Since $\sum_{ij} \dot{n}_{ij} = 0$ because of the Euler formula, the nonzero contribution of the EPR is expected to come from the connected fluctuations and to scale as $\dot{\mathcal{S}} \sim D^{1/2}$.

In the high D regime, where alignment interactions are negligible compared to rotational noise, the limit model is an ideal gas of free ABPs. The mean squared displacement of a free ABP is known:

$$\langle \Delta \mathbf{x}(t)^2 \rangle = \frac{2v_0^2}{D} \left[t - \frac{1}{D} \left(1 - e^{-Dt} \right) \right]. \tag{259}$$

From Eq. (259) we deduce (both in ballistic and diffusive limits) that the reshuffling rate scales as $\dot{n}_{ij} \sim D^{-1}$. At the same time, since the system is disordered, the full correlation function corresponds to the connected one: $C_c(l) \sim e^{-l/\xi(D)}$. For $D \to \infty$, $\xi(D)$ is a finite correlation length, which does not scale significantly with D; thus we can assume that the mutual alignment of the particles will be equal to some default value, independent of D. The resulting scaling for the EPR is $\dot{\mathcal{S}} \sim D^{-2}$.

Symmetry breaking

Let us notice that the argument above does not provide any cue on the symmetry breaking, hence on the positivity of the EPR. Yet, in analogy to the metric case, we can reverse the line of reasoning: knowing that the system is out of equilibrium, irreversibility must impose asymmetry constraints, at least on the N-body steady state distribution, $\psi(\boldsymbol{\Gamma}) \neq \psi(\epsilon \boldsymbol{\Gamma})$. Is it possible to find signatures of irreversibility in a lower-dimensional space, and how do they look like?

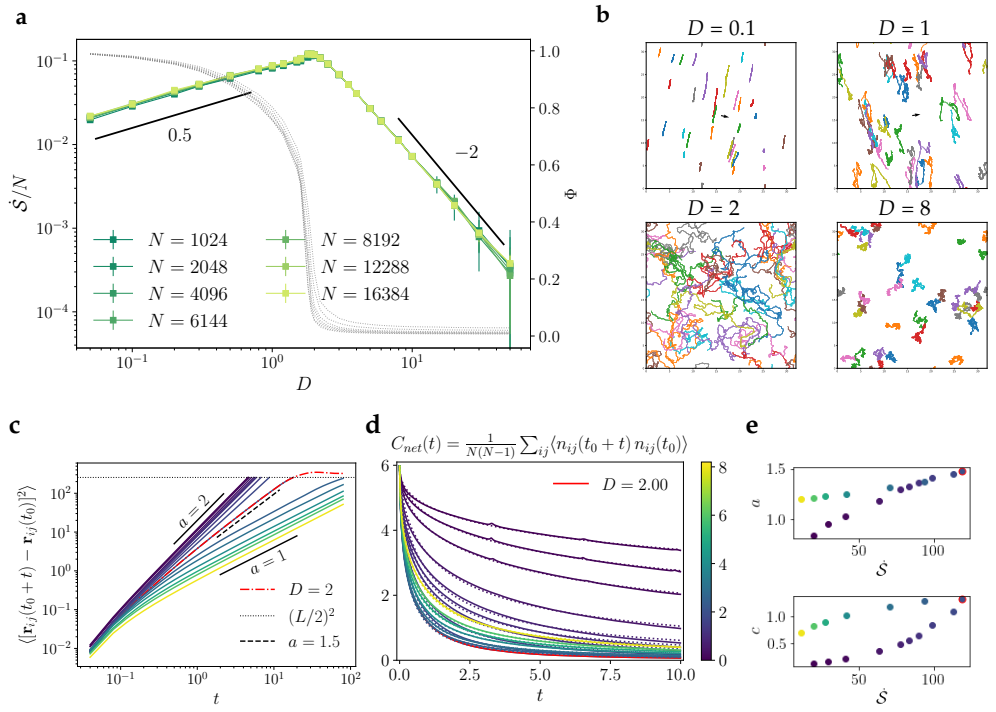

Figure 33: **a**: EPR curves for different system sizes (log scale). The curves perfectly collapse when rescaled by N. The master curve is peaked at the transition point and roughly scales as $D^{0.5}$ and D^{-2} in the low D and high D limits respectively. **b**: Trajectories of few sample particles in the periodic box where the flock is confined, in different regimes. In the ordered phase the trajectories are shown in the reference frame of the center of mass. All the trajectories span the same time interval. The observed diffusion behavior is very different: in the symmetry broken phase, 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, $N = 1024$, $\rho_0 = 1$. **e**: Mutual diffusion curves $|\Delta\mathbf{r}_{ij}|^2$, where $\Delta\mathbf{r}_{ij}(t) = \mathbf{r}_{ij}(t_0+t) - \mathbf{r}_{ij}(t_0)$ and $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$. The exponents we observe for $D < 2$ are different from those obtained through the fit: see the main text for a discussion. The dotted upper bound corresponds to the maximum distance that a pair of birds can reach before periodic boundary effects kick in. **d**: Autocorrelation function of the adjacency matrix of the flock, $C_{\text{net}}(t)$, defined as in Eq. (257). 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. (258). **e**: Parametric plot of fitted parameters, a and c, versus 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.

In contrast to the model with pairwise metric interaction, we cannot write an explicit simple expression for $\dot{n}_{ij}$. However, since the considered model is continuous in time and space and the number of particles (i.e. of Voronoi cells) is conserved, we can notice that cell rearrangements can only be mediated by two kinds of transition configurations: T1 and *rosettes* (Fig. 34.c). The T1 transition corresponds to the tem-

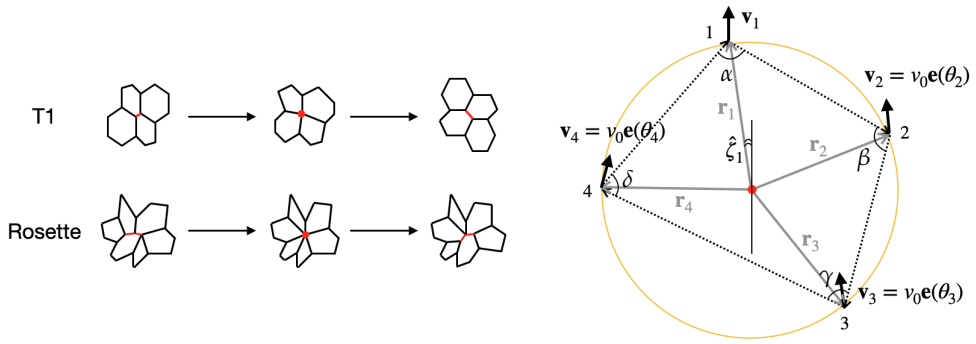

Figure 34: Reshuffling of Voronoi cells through T1 and rosette transitions. The transitional configuration associated to T1 includes a 4-fold Voronoi vertex and a set of 4 particles forming a quadrilateral inscribed in a circle. All the radii $\mathbf{r}_i$ have the same length, for $i = 1 \dots 4$, and the angles satisfy $\alpha + \gamma = \beta + \delta = 2\pi$. Conventionally, we set $\theta_1 = 0$ (vertical direction in the figure).

porary formation of a 4-fold vertex from the coalescence of two standard three-fold vertices; rosettes correspond to configurations with n-fold Voronoi vertices, with degree $n \geq 5$. In a dilute system, we can expect rosettes to be much less likely to form, and assume that reshuffling is entirely due to T1 transitions.

In order to have a 4-fold vertex in a Voronoi tessellation, four neighbor particles must form a quadrilateral inscribed in a circle, as in Fig. 34.c. The condition to be in this transition configuration is therefore that the opposite angles of the quadrilateral are supplementary: $\alpha + \gamma = \beta + \delta = \pi$. Using simple trigonometric formulas and the condition $\sin(\alpha + \gamma) = 0$, we can deduce that, crossing the T1 configuration:

$$\begin{aligned}\frac{d}{dt}\sin(\alpha+\gamma) = \frac{1}{|\mathbf{l}_1||\mathbf{l}_2||\mathbf{l}_3||\mathbf{l}_4|}\{\,&[\mathbf{l}_4\times(\mathbf{v}_2-\mathbf{v}_1)+\mathbf{l}_1\times(\mathbf{v}_4-\mathbf{v}_1)]\,(-\mathbf{l}_3\cdot\mathbf{l}_2)\\ &+[\mathbf{l}_2\times(\mathbf{v}_4-\mathbf{v}_3)+\mathbf{l}_3\times(\mathbf{v}_2-\mathbf{v}_3)]\,(-\mathbf{l}_1\cdot\mathbf{l}_4)\\ &+(\mathbf{l}_4\times\mathbf{l}_1)\,[(\mathbf{v}_3-\mathbf{v}_4)\cdot\mathbf{l}_2\\ &+(\mathbf{v}_2-\mathbf{v}_3)\cdot\mathbf{l}_3]+(\mathbf{l}_3\times\mathbf{l}_4)\,[(\mathbf{v}_1-\mathbf{v}_2)\cdot\mathbf{l}_4\\ &+(\mathbf{v}_4-\mathbf{v}_1)\cdot\mathbf{l}_1]\,\},\end{aligned} \tag{260}$$

where $\mathbf{l}_i = \mathbf{r}_{(i+1) \bmod 4} - \mathbf{r}_i$. Since the quadrilateral is inscribed in a circle ($|\mathbf{r}_i| = r_0$) and $|\mathbf{v}_i| = v_0$, Eq. (260) can also be rewritten as a function of 7 independent angular parameters which are sufficient to characterize the 4-fold vertex configuration. Given $\mathbf{r}_i = r_0(\cos\zeta_i, \sin\zeta_i)$, and $\mathbf{v}_i = v_0(\cos\theta_i, \sin\theta_i)$, the angular parameters of interest are:

- $\hat{\zeta}_1$, i.e. the angle formed by the radius $\mathbf{r}_1$ with respect to the reference direction θ_1,;
- the angles $\varphi_i = \theta_i - \theta_1$ measuring how each particle is aligned to the first bird;
- the angles between adjacent radii, ζ_{21}, ζ_{32}, ζ_{43}, where $\zeta_{ij} = \zeta_i - \zeta_j$.

The sign of the quantity in Eq.(260) determines the evolution of the 4-fold vertex. If $\frac{d}{dt}\sin(\alpha+\gamma) > 0$, $\alpha+\gamma$ shrinks and becomes smaller than π. The quadrilateral will not be inscribed in the circle anymore, and the particles 1 and 3 will become non-neighbors: $\dot{n}_{13} < 0$. The opposite fate is for the pair of particles $(2, 4)$, associated

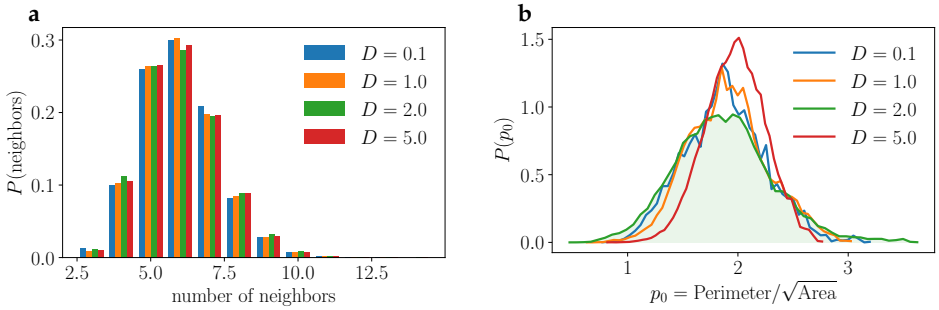

Figure 35: **a**: Distribution of the number of neighbors reconstructed from numerical simulations. **b**: Statistical characterization of the geometrical properties of the Voronoi cells through the parameter p_0, defined as the ratio between the perimeter and the square root of the area of the Voronoi cell. We recall that the transition point is at $D \simeq 2$.

to the angles β and δ. Due to the linear dependence of Eq. (260) on the velocity vectors, the effect of the time reversal operator is to change the sign of $\frac{d}{dt}\sin(\alpha+\gamma)$. Therefore, in order to have $\dot{\mathcal{S}} > 0$, the pair of neighbors $(1,3)$ such that $\dot{n}_{13} < 0$ must be more aligned than the pair $(2,4)$. In conclusion, in a metric-free model signatures of irreversibility do not directly affect the two-particle p.d.f., but must be sought in the four-particle steady-state p.d.f..

Another interesting question is whether the Voronoi interaction network presents any distinctive *static* features close to the transition point, which may help predicting its dynamical rearrangements. We briefly looked at two quantities of possible interest: the number of neighbors and the ratio between the perimeter and the square root of the area of the Voronoi cells. The distributions of both parameters are shown in Fig. 35.

The number of neighbors's distribution must have, at any D, a constrained average ($\langle n_c \rangle = 6$), but we observe that there is no relevant difference even in the fluctuations. The whole histograms are overlapping for a broad range of D values, suggesting that the efficiency of reshuffling (which is maximal at the transition) is not correlated to an excess or a lack of defects. By defects we mean here any kind of deviation from the reference hexagonal structure of the Voronoi cells.

The second parameter, defined as $p_0 = \text{Perimeter}/\sqrt{\text{Area}}$, is another indicator of the geometrical properties of the Voronoi cells. The *typical* value of this parameter (computed using the average perimeter and area, around which the cells' perimeters and areas fluctuate) acts as a control parameter for the jamming transition in self-propelled Voronoi models [9]. In our Vicsek model with Voronoi-based alignment interactions, this parameter does not govern directly the evolution of the system, in contrast to [9], but it can give an indication of how elongated the cells are in different conditions. Unfortunately, even from the analysis of this second parameter, no clear picture emerged.

4.2.6 Methods

We integrate the stochastic equations of motion using a forward Euler-Maruyama scheme, where positional and angular updates are shifted of a half-step. We obtain in this way two parallel time series $\{\mathbf{r}_i^{n-1/2}, n = 1 \dots N_s\}$ and $\{\mathbf{v}_i^n, n = 1 \dots N_s\}$, where N_s is the number of iterated steps. The entropy production rate is computed as:

$$\frac{\dot{q}}{D} = -\frac{J}{D}\frac{1}{N_s \Delta t}\sum_{n=1}^{N_s}\sum_{ij}\left(\theta_i^{n+1} - \theta_i^n\right) n_{ij}^{n+1/2}\frac{1}{2}\left[\sin(\theta_i^n - \theta_j^n) + \sin(\theta_i^{n+1} - \theta_j^{n+1})\right] \quad (261)$$

and

$$\frac{\dot{w}_{\text{resh}}}{D} = -\frac{J}{2D}\frac{1}{N_s \Delta t}\sum_{n=1}^{N_s}\sum_{ij}\left[n_{ij}^{n+1/2} - n_{ij}^{n+1/2}\right]\cos(\theta_i^n - \theta_j^n). \quad (262)$$

This prescription corresponds to the discretization adopted in Eqs. (216) and (219), through which we found the same formulas as those obtained with a continuous-time formalism.

The amplitude Δt of the integration time step is tuned according to the values of the control parameters, ρ_0, D and v_0, in order to assure that reshuffling and heat dissipation rates are correctly reproduced (see Fig. 23e-f). To speed up the thermalization, a *pre-thermalization* evolution with a time step $10\Delta t$ is implemented, followed by a thermalization run with integration step Δt. We checked that the system has reached a steady state by looking at the stationarity of the polarization and EPR time series.

Particles are constrained on a square box of linear size $L = (N/\rho_0)^{1/2}$ with periodic boundary conditions. The n_{ij} matrix is updated at each time step because we need an accurate measurement of reshuffling. In the metric case, the neighbor network is built by dividing the box in cells of size R and scanning adjacent cells at each sweep, while in the topological case we use the C++ library CGAL (5.1.1 release) for Delaunay triangulations on the torus [99].

4.3 FINAL REMARKS

4.3.1 Summary

In this ongoing project we have tried to characterize the non-equilibrium state of two paradigmatic variants of the Vicsek model through the measurement of their entropy production rate. These models describe systems of interacting ABPs which interact through short-ranged ferromagnetic interactions which are either topological or metric in nature. In both cases, the departure from equilibrium is exclusively due to a combination of interaction and motility: without the latter, the model could describe the Langevin dynamics of a Hamiltonian spin system.

Exploiting classical results in stochastic thermodynamics, we could cast the effect of motility into an effective time dependence of some parameters of the Hamiltonian (also known as *protocol*). In contrast to the standard scenario where deterministic

external protocols are employed, here they are effectively stochastic and *correlated* to the state of the system, making it hard to analytically evaluate the EPR without knowing the NESS distribution. However, we could pursue a numerical study on systems of moderate size, which has helped us to characterize the degree of irreversibility of the observed non-equilibrium dynamics across the phase diagram.

Our numerical analysis confirms the existence of two equilibrium limits, for vanishing and diverging angular diffusivities. For intermediate D values, $\dot{\mathcal{S}} > 0$ and signatures of irreversibility can be ferreted out in asymmetries of suitable steady-state distributions. The distributions to choose and the observed asymmetries reflect the properties of the interaction rule.

We consider for the metric case a sharp step-like interaction kernel, according to which a pair of ABPs interact (with equal strength) only if their distance does not exceed a given threshold R. This parametrization concentrates the dissipation on the scale of the interaction radius: positive and negative contributions to the entropy production rate respectively correspond to pairs of aligned birds which are exiting each other's interaction disk or entering into it. When particles are anti-aligned (a much less common situation at distance R) the role of colliding and diverging pairs is exchanged. Positive and negative inputs mainly cancel out, except for a residual contribution which is always positive, and maximal in polar systems that are not too far from the transition. Typical configurations for these systems are made of dense moving clusters of aligned particles, coexisting with a fraction of particles in a more dilute and isotropic phase. This fact hints at an enhancement of non-equilibrium effects when phase coexistence (a hallmark of first order transitions) is realized. Similar conclusions were obtained in [15] working with fluctuating continuous theories at the mesoscopic level.

In the topological variant of the model, each particle interacts with the first shell of Voronoi neighbors. Here again the interaction is additive and symmetric, and momentum is conserved, but reshuffling is harder to characterize, as it involves four-particle distributions, rather than two-particle distributions. Since the dynamics of these systems is less sensitive to density fluctuations, compared to the metric case, formation of dense polar clusters is suppressed — at least at the observed sizes — and the transition looks like a second order ferromagnetic transition. Correspondingly, we observe a more regular EPR phase diagram with a peak at the critical point. The power-law scaling of the EPR curve at low and high temperature is well predicted on the basis of approximate arguments.

As a final remark, let us notice that in both cases, metric and topological, the entropy production seems to be an extensive quantity: dissipation occurs at a local scale, thanks to the presence of torques which irreversibly align particles' orientations. The role of these peculiar force terms in bringing the system out of equilibrium has been already pinpointed in [175, 75], by showing that it is impossible to define the pressure as a state function. Here their peculiar role stems from the parity under time reversal of the variables on which they act. This observation is opposed to what has been observed for scalar active matter models, both at a microscopic and coarse-grained level [137, 130]. When MIPS is realized, in the absence

of explicit alignment interactions[14], non-equilibrium effects are concentrated at the interface between dense and dilute phases. The non-equilibrium terms of active ϕ^4 theories which describe the motility-induced phase separation at a coarse-grained level are indeed gradient terms which give a negligible contribution in the phase bulk, but are large at the interface.

4.3.2 Pending issues

We believe that the current study can be improved and expanded in several directions: the first pending issue is that of building an adequate kinetic theory to describe the density function $q(\alpha, \varphi)$. We recall that $q(\alpha, \varphi)$ measures the probability of finding a pair of particles with mutual misalignment φ at a distance R, along a direction parametrized by α. Our attempts to use a ring-kinetic theory with a homogeneity assumption for the additive ABP model resulted into an improved description of the transition, compared to the homogeneous mean field case, but did not allow us to reconstruct a function $q(\alpha, \varphi)$ sharing the same properties as the one we reconstructed from agent-based simulations.

A second interesting point regards the relation with the violation of the fluctuation-dissipation theorem (FDT). The fluctuation dissipation theorem links responses to small perturbations with correlations in the unperturbed steady-state of equilibrium systems. While the entropy production rate measures the breakdown of detailed balance, FDT violations may be quantified through an effective temperature as the one defined in [62]. This parameter has been largely used in the study of spin glass dynamics and was recently applied to characterize the non-equilibrium steady state of some active matter models [62, 119, 78, 124]. In the case of polar matter with pure ferromagnetic interactions, it can be used as a tool to investigate the ability of the flock to respond to directional perturbations and the impact of irreversible currents on the realization of collective turns.

Finally, it could be interesting to study, within the same framework, the role of spatial dimension, which is very important in the theory of critical phenomena, and how the presence of inertia modifies the non-equilibrium steady state of the system, by enhancing or suppressing the observed asymmetries in the pair distribution of active particles, which results into an enhancement or suppression of the breakdown of detailed balance.

14 It is worth noting that, even if microscopic MIPS models do not contain explicit alignment interactions, an effective alignment of the particles' velocity can emerge from the interplay between self-propulsion and repulsion [26].

5 CONCLUSIONS

In this thesis we have studied the microscopic dynamics of polar active systems by addressing two very different problems. The common trait of the presented work is the interest in describing emergent non-equilibrium phenomena in groups of animals which have acquired a paradigmatic role in the field of active matter. We primarily refer to starling flocks, but the presented analysis and methods can be extended also to other active systems which are supposed to be described by the same kind of models.

The first part of the work consists in the development of a new inference scheme for the collective dynamics of birds. The project was motivated by the availability of experimental data on large flocks of European starlings collected by the CoBBS group. Previous analysis of these data have revealed that the observed dynamics is better described by inertial stochastic models for the birds' velocities than by overdamped ones. The introduction of a generalized inertia allows us to explain how collective turns are performed, but also introduces significant complications for the application of statistical inference methods. Indeed, the new model enlarges the dimension of the system's phase space, which cannot be fully accessed in experiments.

The problem of reconstructing partially observed processes from time series data is common in many types of empirical studies and is well-known in the literature. Usually, it is tackled by means of computationally demanding methods. In Chapter 3 we propose a novel efficient inference scheme for generalized Langevin equations obtained as underdamped generalizations of standard, first order, SDEs. Thanks to the mathematical properties of the considered class of equations, for which memory and color kernels are explicitly known and easy to manipulate, we could derive an approximate analytical expression for the likelihood of the observed time series. This provides the solution to the considered Bayesian inference problem for a whole class of processes of physical interest.

The method we propose gives reliable and fast results, provided that the time series is noiseless and sufficiently long. We recall indeed that in any parametric inference approach the correctness of the results depends on the correctness of the underlying assumptions about the chosen model. A significant presence of noise would result into incorrect modeling, in this setting. Additionally, the parameter estimators of our inference scheme are only asymptotically consistent, in the limit of infinitely long time series. However the bias is fastly decaying with the length of the trajectory (measured in number of points) and in any case under control. This bias can be easily predicted, and an a posteriori correction may be implemented when necessary.

The analysis of the bias turns helpful for the application of the inference scheme to real data, which may come in the form of segmented collections of short time

series. In the case of starling flocks, even if individual trajectories are typically reconstructed for longer intervals, it is convenient to work with triplets of subsequent frames. A preliminary analysis of the data conducted with this method shows however that the raw trajectories do not fit the supposed inertial model. The origin for this observation is the presence of spurious signals due to measurement noise and birds' wing flapping. Therefore, the available data must be carefully pre-processed, before a sensible parameter-free model can be extracted from them.

At the moment, a definitive protocol for the above-mentioned data-set is not available, but several strategies can be employed to tackle this issue. A first possible strategy may be to apply non-specific filters to the time series, in order to de-noise the signal and/or remove the periodic components. An alternative method would consist in introducing an explicit model for the noise and infer it from the available data, along with the dynamics of the system, pursuing, for instance, an expectation-maximization principle. Finally, one can also think about using non-parametric or semi-parametric inference approaches. However, the analyzed data is high dimensional (hundreds of 3D trajectories for each flocking event) and the statistics is limited (we can rarely perform multiple observations of the same flock in the same conditions, and each time series is of the order of several hundreds of time points). Parametrization — at least a partial one — is then inescapable to build a tailored method which allows for efficient and statistically reliable inference.

An additional advantage of the parametric Bayesian method is the possibility to assign a probability score to the optimal parameter-free model that results from its application, and compare how good different types of dynamical models are in describing the available data. In our case, this means quantifying the improvement associated to switching from a first order Vicsek-like model to the inertial spin model. The information we can extract through parametric approaches is however not limited to this piece: being able to have separate estimates for all the ISM parameters — not just appearing in combinations as in previous empirical studies — opens new possibilities to check the *predictions* of the model and test diverse hypotheses.

A first example regards the second sound speed c_2, which can be numerically or analytically computed once the parameters of the inertial spin model are known. This estimate can be compared with the fitted values of the speed of propagation of the turning perturbation obtained in [3]. Moreover, the values of the ISM parameters, combined in proper ways, can be compared with previously inferred parameters of static models [11] or first order dynamical models [136], in order to understand under which conditions these simpler descriptions are effective.

New predictions, specific of the interacting inertial dynamics, can also be checked. For instance, a parameter-free model can provide an approximate quantitative prediction of the maximum distance over which real flocks can sustain an almost undamped propagation of the turning perturbation. Despite starlings can gather in large groups, containing up to hundreds of thousands of individuals, the flocks which have been experimentally observed during collective turns are typically much smaller. A possible explanation for this fact may come from constraints imposed by information transfer mechanisms. In inertial models, as the wave-vector k is re-

duced, information transmission switches from wavelike to diffusive, causing shape deformations and coupled polarity-density fluctuations, like in Toner-Tu liquids. This kind of perturbations will lead to a disruption of the flock, hindering its social and protective function [199, 157]. For this reason, the prediction of the maximal size — i.e. minimal k sustained by the finite system — which allows for undamped propagation is particularly interesting, as it is related to the ability of the flock to perform coherent turns.

These aspects make up the physical motivation to strive for the development and employment of parametric Bayesian inference methods. The work presented in Chapter 3 focuses on technical issues and served to highlight the shortcomings of the most intuitive low-order discretizations that one can apply in order to reconstruct partially observed stochastic processes from time series data. The technical difficulties raised by partial observation are not a peculiar feature of the inertial model which we use to describe bird flocks, but a ubiquitous problem in time series analysis. The origin of such difficulty can be tracked down to the lack of finite-dimensional delay vector embeddings that provide an exact Markovian description of a partially observed stochastic process. As abyproduct of this investigation, we proposed a simple formal analogy with the Renormalization Group construction in real space to provide an intuitive explanation for this fact.

In the second part of this thesis, contained in Chapter 4, we tried to address the leitmotif of quantifying the departure from equilibrium in active matter. Specifically, we considered two variants of ABP systems (with metric and topological interactions) which can be thought of as continuous-time regularizations of the original Vicsek model. In these minimal models, where each particle has a fixed speed and the interaction is a reciprocal short-ranged interaction of ferromagnetic nature, irreversibility only stems from the interplay between motility and interaction. As particles move, mutual distances — and consequently interaction networks — evolve with time, breaking the detailed balance and allowing for the onset of collective motion even in $d = 2$.

The question of measuring the distance from equilibrium can be addressed at many different scales for such spatially extended systems, and using several analytical tools. We decided to focus on the microscopic level and to pivot on the entropy production rate to characterize the non-equilibrium steady state of these systems. The entropy production rate is notoriously a model-dependent quantity, which we cannot measure in agnostic ways in experiments. Yet general observable signatures of irreversibility can be deduced solely on the basis of which degrees of freedom are selected to describe the desired process. These signatures can be used to quantify the departure from equilibrium in systems for which reasonably satisfactory models already exist, or to infer precious information about the nature of the interaction in complex off-equilibrium active systems.

For the considered class of ABP models, such signatures come in the form of asymmetries in suitably defined steady-state distributions. The time reversal symmetry breaking constrains indeed an explicit symmetry breaking in the steady state of the N-particle system. We investigated how this asymmetry is manifested in simpler and empirically accessible functions, depending on the kind of interaction

that is implemented. Specifically, we found that when the interaction is purely pairwise, this asymmetry directly affects the two-particle distribution. This prediction has been confirmed by numerical simulations of metric additive models. A similar qualitative picture emerges from the numerical study of a topological variant of the model, where each particle interacts with the first shell of Voronoi neighbors. The interaction is not purely pairwise in this second case, and the irreversibility condition directly constrains the four-particle density rather than the pair density, whose asymmetries are much more complicated to analyze. In summary, knowledge about the interaction between the constituents of the system projects the asymmetry of the steady-state probability density onto a low-dimensional space, which offers convenient tools for the analysis of experimental and numerical data.

The picture that emerges from this study points at reshuffling as an effective (self-sustained) mechanism to inject energy into the system, and dissipate it by exchanging heat with the thermal bath. We could observe that the autocorrelation function of the interaction network has the fastest decay close to the transition point, where the EPR curve peaks. In this regime, reshuffling is mostly efficient, since individual trajectories have enough persistence to quickly move away from a previously occupied point, without being followed by the rest of the flock (unlike in the deeply ordered phase). From numerical simulations we could check the existence of two equilibrium limits, for highly polarized or completely disordered flocks. While the latter is easy to predict, since the system corresponds asymptotically to an ideal gas of non-interacting ABPs, the former is less trivial because it depends on how strongly reshuffling is suppressed as the noise amplitude is lowered. An interesting aspect which connects the existence of such limits with the previous discussion about the asymmetries of the steady-state distribution is that, as the system approaches these equilibrium regimes and recovers a time-reversal symmetry, the rotational symmetry in the space of velocities is recovered. Indeed the breakdown of detailed balance comes along with the explicit breakdown of a more general symmetry, i.e. the symmetry of the NESS under independent rotations in the subspaces of particles' positions and velocities. Out of equilibrium, these two subspaces are inestricably linked; at equilibrium, the two subspaces get decoupled and Mermin-Wagner-like theorems are re-introduced, independently of whether particles are motile or not [184].

A wealth of related questions arose from this second work, and they are in part under current investigation. The first one concerns the search of a suitable kinetic theory which could predict the salient features of the two-particle density function of the metric model. A correct reproduction of this quantity must include pair correlations, i.e. the connected part of the pair distribution, which is typically neglected in the derivation of hydrodynamic theories based on the molecular chaos approximation. A step forward is represented by ring-kinetic theories. These have been already developed by other authors for a variant of the Vicsek model with binary interactions and for the non-additive ABP model, where alignment torques are normalized by the total number of interacting neighbors of each bird [53, 113]. In the latter case, the asymmetries they observe are of the same kind as the ones we

predict on the basis of thermodynamic arguments, suggesting that the signatures of irreversibility are in fact quite independent of the microscopic details.

A second question of interest is the connection between the EPR and other measures of irreversibility. Among them, the most important ones are effective temperatures, which quantify the violation of the fluctuation-dissipation theorem in non-equilibrium systems. When exposed to a turning perturbation, a flock generally responds in a different way than it would respond to internal fluctuations. We expect the deviation from the equilibrium-like response to be more and more evident as the polarization is lowered and the system approaches the region of the phase diagram where the EPR is maximal. FDT violation and EPR are indeed connected by a well-known relation, when the non-equilibrium dynamics is described in terms of a Langevin equation [94, 95].

We are also wondering about the possible effects on the EPR of speed fluctuations, non-reciprocal interactions and inertia. Despite these might look like irrelevant details, the phenomena observed at the macroscopic scale in metric versus topological models, momentum-conserving versus non-momentum conserving models, inertial versus overdamped models are significantly different, and we expect that these differences will be reflected by the EPR behavior.

In conclusion, the presented work can be seen as a tiny part of a larger effort aiming at describing the collective behavior of bird flocks (and similar animal groups ascribing to the class of polar active matter) through simple theoretical models inspired by experimental observations. We hope that the developed methodological tools and our numerical investigations on the above-mentioned reference models may help expand our understanding of these inevitably complex, out of equilibrium systems.

BIBLIOGRAPHY

[1] Alessandro Attanasi et al. "Emergence of collective changes in travel direction of starling flocks from individual birds' fluctuations." In: *Journal of The Royal Society Interface* 12.108 (2015), p. 20150319. DOI: 10.1098/rsif.2015.0319. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2015.0319.

[2] Alessandro Attanasi et al. "GReTA-A Novel Global and Recursive Tracking Algorithm in Three Dimensions." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37.12 (2015), pp. 2451–2463. DOI: 10.1109/TPAMI.2015.2414427.

[3] Alessandro Attanasi et al. "Information transfer and behavioural inertia in starling flocks." In: *Nature Physics* 10 (July 2014), p. 691. URL: https://doi.org/10.1038/nphys3035.

[4] M. Ballerini et al. "Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study." In: *Proceedings of the National Academy of Sciences* 105.4 (2008), pp. 1232–1237. ISSN: 0027-8424. DOI: 10.1073/pnas.0711437105. URL: https://www.pnas.org/content/105/4/1232.

[5] Clemens Bechinger et al. "Active particles in complex and crowded environments." In: *Rev. Mod. Phys.* 88 (4 Nov. 2016), p. 045006. DOI: 10.1103/RevModPhys.88.045006. URL: https://link.aps.org/doi/10.1103/RevModPhys.88.045006.

[6] D. Benedetto, P. Buttà, and E. Caglioti. "Some aspects of the inertial spin model for flocks and related kinetic equations." In: *Mathematical Models and Methods in Applied Sciences* 30.10 (2020), pp. 1987–2022. DOI: 10.1142/S0218202520400151. URL: https://doi.org/10.1142/S0218202520400151.

[7] Eric Bertin, Michel Droz, and Guillaume Grégoire. "Boltzmann and hydrodynamic description for self-propelled particles." In: *Phys. Rev. E* 74 (2 Aug. 2006), p. 022101. DOI: 10.1103/PhysRevE.74.022101. URL: https://link.aps.org/doi/10.1103/PhysRevE.74.022101.

[8] Eric Bertin, Michel Droz, and Guillaume Grégoire. "Hydrodynamic equations for self-propelled particles: microscopic derivation and stability analysis." In: 42.44 (Oct. 2009), p. 445001. DOI: 10.1088/1751-8113/42/44/445001. URL: https://doi.org/10.1088/1751-8113/42/44/445001.

[9] Dapeng Bi et al. "Motility-Driven Glass and Jamming Transitions in Biological Tissues." In: *Phys. Rev. X* 6 (2 Apr. 2016), p. 021011. DOI: 10.1103/PhysRevX.6.021011. URL: https://link.aps.org/doi/10.1103/PhysRevX.6.021011.

[10] William Bialek et al. "Social interactions dominate speed control in poising natural flocks near criticality." In: *Proceedings of the National Academy of Sciences* 111.20 (2014), pp. 7212–7217. ISSN: 0027-8424. DOI: 10.1073/pnas.1324045111. URL: https://www.pnas.org/content/111/20/7212.

[11] William Bialek et al. "Statistical mechanics for natural flocks of birds." In: *Proceedings of the National Academy of Sciences* 109.13 (2012), pp. 4786–4791. ISSN: 0027-8424. DOI: 10.1073/pnas.1118633109. URL: https://www.pnas.org/content/109/13/4786.

[12] Béla Bollobás. *Modern graph theory*. Graduate texts in Mathematics. Springer, 1998.

[13] J. P. Boris. "Relativistic plasma simulation-optimization of a hybrid code." In: *Fourth Conference on Numerical Simulation of Plasmas*. Ed. by J.P. Boris and R.A. Shanny. Naval Research Laboratory. 1970, pp. 3–67. URL: https://ci.nii.ac.jp/naid/10009996893/en/.

[14] Francesco Borra and Marco Baldovin. "Using machine-learning modeling to understand macroscopic dynamics in a system of coupled maps." In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 31.2 (2021), p. 023102. DOI: 10.1063/5.0036809. URL: https://doi.org/10.1063/5.0036809.

[15] Øyvind L Borthne, Étienne Fodor, and Michael E Cates. "Time-reversal symmetry violations and entropy production in field theories of polar active matter." In: *New Journal of Physics* 22.12 (Dec. 2020), p. 123012. DOI: 10.1088/1367-2630/abcd66. URL: https://doi.org/10.1088/1367-2630/abcd66.

[16] Mickael Bourgoin et al. "Kolmogorovian Active Turbulence of a Sparse Assembly of Interacting Marangoni Surfers." In: *Phys. Rev. X* 10 (2 June 2020), p. 021065. DOI: 10.1103/PhysRevX.10.021065. URL: https://link.aps.org/doi/10.1103/PhysRevX.10.021065.

[17] Edouard Brézin. *Introduction to Statistical Field Theory*. Cambridge University Press, 2010. DOI: 10.1017/CBO9780511761546.

[18] Antoine Bricard et al. "Emergence of macroscopic directed motion in populations of motile colloids." In: *Nature* 503.19 (7474 Nov. 2013), pp. 95–8. ISSN: 0028-0836. DOI: 10.1038/nature12673.

[19] P.J. Brockwell and R.A. Davis. *Introduction to Time Series and Forecasting*. Springer, Berlin, 2002.

[20] David B. Brückner, Pierre Ronceray, and Chase P. Broedersz. *Inferring the Dynamics of Underdamped Stochastic Systems*. July 2020. DOI: 10.1103/PhysRevLett.125.058103. URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.058103.

[21] David B. Brückner et al. "Learning the dynamics of cell–cell interactions in confined cell migration." In: *Proceedings of the National Academy of Sciences* 118.7 (2021). ISSN: 0027-8424. DOI: 10.1073/pnas.2016602118. URL: https://www.pnas.org/content/118/7/e2016602118.

[22] David B. Brückner et al. "Stochastic nonlinear dynamics of confined cell migration in two-state systems." In: *Nature Physics* 15.6 (2019), pp. 595–601. DOI: 10.1038/s41567-019-0445-4. URL: https://doi.org/10.1038/s41567-019-0445-4.

[23] Axel Brünger, Charles L. Brooks, and Martin Karplus. "Stochastic boundary conditions for molecular dynamics simulations of ST2 water." In: *Chemical Physics Letters* 105.5 (1984), pp. 495–500. ISSN: 0009-2614. DOI: https://doi.org/10.1016/0009-2614(84)80098-6. URL: http://www.sciencedirect.com/science/article/pii/0009261484800986.

[24] Fernando Caballero and Michael E. Cates. "Stealth Entropy Production in Active Field Theories near Ising Critical Points." In: *Phys. Rev. Lett.* 124 (24 June 2020), p. 240604. DOI: 10.1103/PhysRevLett.124.240604. URL: https://link.aps.org/doi/10.1103/PhysRevLett.124.240604.

[25] Fernando Caballero, Cesare Nardini, and Michael E Cates. "From bulk to microphase separation in scalar active matter: a perturbative renormalization group analysis." In: *Journal of Statistical Mechanics: Theory and Experiment* 2018.12 (Dec. 2018), p. 123208. DOI: 10.1088/1742-5468/aaf321. URL: https://doi.org/10.1088/1742-5468/aaf321.

[26] L. Caprini, U. Marini Bettolo Marconi, and A. Puglisi. "Spontaneous Velocity Alignment in Motility-Induced Phase Separation." In: *Phys. Rev. Lett.* 124 (7 Feb. 2020), p. 078001. DOI: 10.1103/PhysRevLett.124.078001. URL: https://link.aps.org/doi/10.1103/PhysRevLett.124.078001.

[27] Lorenzo Caprini et al. "The entropy production of Ornstein–Uhlenbeck active particles: a path integral method for correlations." In: *Journal of Statistical Mechanics: Theory and Experiment* 2019.5 (May 2019), p. 053203. DOI: 10.1088/1742-5468/ab14dd. URL: https://doi.org/10.1088/1742-5468/ab14dd.

[28] Martin Casdagli et al. "State space reconstruction in the presence of noise." In: *Physica D: Nonlinear Phenomena* 51.1 (1991), pp. 52–98. ISSN: 0167-2789. DOI: https://doi.org/10.1016/0167-2789(91)90222-U. URL: http://www.sciencedirect.com/science/article/pii/016727899190222U.

[29] Davide Cassi. "Phase transitions and random walks on graphs: A generalization of the Mermin-Wagner theorem to disordered lattices, fractals, and other discrete structures." In: *Phys. Rev. Lett.* 68 (24 June 1992), pp. 3631–3634. DOI: 10.1103/PhysRevLett.68.3631. URL: https://link.aps.org/doi/10.1103/PhysRevLett.68.3631.

[30] M. E. Cates and J. Tailleur. "When are active Brownian particles and run-and-tumble particles equivalent? Consequences for motility-induced phase separation." In: *EPL (Europhysics Letters)* 101.2 (Jan. 2013), p. 20010. DOI: 10.1209/0295-5075/101/20010. URL: https://doi.org/10.1209/0295-5075/101/20010.

[31] Michael E. Cates. *Active Field Theories*. 2019. arXiv: 1904.01330 [cond-mat.stat-mech].

[32] A. Cavagna et al. "Diffusion of individual birds in starling flocks." In: *Proceedings of the Royal Society B: Biological Sciences* 280.1756 (2013), p. 20122484. DOI: 10.1098/rspb.2012.2484. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspb.2012.2484. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2012.2484.

[33] A. Cavagna et al. "Error control in the set-up of stereo camera systems for 3d animal tracking." In: *The European Physical Journal Special Topics* 224.17 (Dec. 2015), pp. 3211–3232. ISSN: 1951-6401. DOI: 10.1140/epjst/e2015-50102-3. URL: https://doi.org/10.1140/epjst/e2015-50102-3.

[34] Andrea Cavagna, Irene Giardina, and Tomás S. Grigera. "The physics of flocking: Correlation as a compass from experiments to theory." In: *Physics Reports* 728 (2018). The physics of flocking: Correlation as a compass from experiments to theory, pp. 1–62. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2017.11.003. URL: http://www.sciencedirect.com/science/article/pii/S0370157317303575.

[35] Andrea Cavagna et al. "Dynamical maximum entropy approach to flocking." In: *Phys. Rev. E* 89 (4 Apr. 2014), p. 042707. DOI: 10.1103/PhysRevE.89.042707. URL: https://link.aps.org/doi/10.1103/PhysRevE.89.042707.

[36] Andrea Cavagna et al. "Dynamical Renormalization Group Approach to the Collective Behavior of Swarms." In: *Phys. Rev. Lett.* 123 (26 Dec. 2019), p. 268001. DOI: 10.1103/PhysRevLett.123.268001. URL: https://link.aps.org/doi/10.1103/PhysRevLett.123.268001.

[37] Andrea Cavagna et al. "Flocking and Turning: a New Model for Self-organized Collective Motion." In: *Journal of Statistical Physics* 158.3 (Feb. 2015), pp. 601–627. ISSN: 1572-9613. DOI: 10.1007/s10955-014-1119-3. URL: https://doi.org/10.1007/s10955-014-1119-3.

[38] Andrea Cavagna et al. *Natural Swarms in* **3.99** *Dimensions*. 2021. arXiv: 2107.04432 [cond-mat.stat-mech].

[39] Andrea Cavagna et al. "Nonsymmetric Interactions Trigger Collective Swings in Globally Ordered Systems." In: *Phys. Rev. Lett.* 118 (13 Mar. 2017), p. 138003. DOI: 10.1103/PhysRevLett.118.138003. URL: https://link.aps.org/doi/10.1103/PhysRevLett.118.138003.

[40] Andrea Cavagna et al. "Scale-free correlations in starling flocks." In: *Proceedings of the National Academy of Sciences* 107.26 (2010), pp. 11865–11870. ISSN: 0027-8424. DOI: 10.1073/pnas.1005766107. URL: https://www.pnas.org/content/107/26/11865.

[41] Andrea Cavagna et al. "Short-range interactions versus long-range correlations in bird flocks." In: *Phys. Rev. E* 92 (1 July 2015), p. 012705. DOI: 10.1103/PhysRevE.92.012705. URL: https://link.aps.org/doi/10.1103/PhysRevE.92.012705.

[42] Andrea Cavagna et al. "Silent Flocks: Constraints on Signal Propagation Across Biological Groups." In: *Phys. Rev. Lett.* 114 (21 May 2015), p. 218101. DOI: 10.1103/PhysRevLett.114.218101. URL: https://link.aps.org/doi/10.1103/PhysRevLett.114.218101.

[43] Andrea Cavagna et al. "SpaRTA Tracking Across Occlusions via Partitioning of 3D Clouds of Points." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.4 (2021), pp. 1394–1403. DOI: 10.1109/TPAMI.2019.2946796.

[44] Andrea Cavagna et al. "Spatio-temporal correlations in models of collective motion ruled by different dynamical laws." In: *Physical Biology* 13.6 (Nov. 2016), p. 065001. DOI: 10.1088/1478-3975/13/6/065001. URL: https://doi.org/10.1088%2F1478-3975%2F13%2F6%2F065001.

[45] Andrea Cavagna et al. "The STARFLAG handbook on collective animal behaviour: Part I, empirical methods." In: *Animal Behaviour* 76.1 (2008), pp. 217–236.

[46] Andrea Cavagna et al. "The STARFLAG handbook on collective animal behaviour: Part II, three-dimensional analysis." In: *Animal Behaviour* 76.1 (2008), pp. 237–248.

[47] Sara Dal Cengio, Demian Levis, and Ignacio Pagonabarraga. "Fluctuation–dissipation relations in the absence of detailed balance: formalism and applications to active matter." In: *Journal of Statistical Mechanics: Theory and Experiment* 2021.4 (Apr. 2021), p. 043201. DOI: 10.1088/1742-5468/abee22. URL: https://doi.org/10.1088/1742-5468/abee22.

[48] H. Chaté. "Dry Aligning Dilute Active Matter." In: *Annual Review of Condensed Matter Physics* 11 (2020), pp. 189–212.

[49] Hugues Chaté et al. "Collective motion of self-propelled particles interacting without cohesion." In: *Phys. Rev. E* 77 (4 Apr. 2008), p. 046113. DOI: 10.1103/PhysRevE.77.046113. URL: https://link.aps.org/doi/10.1103/PhysRevE.77.046113.

[50] Leiming Chen, John Toner, and Chiu Fan Lee. "Critical phenomenon of the order–disorder transition in incompressible active fluids." In: *New Journal of Physics* 17.4 (Apr. 2015), p. 042002. DOI: 10.1088/1367-2630/17/4/042002. URL: https://doi.org/10.1088/1367-2630/17/4/042002.

[51] Oleksandr Chepizhko, David Saintillan, and Fernando Peruani. "Revisiting the emergence of order in active matter." In: *Soft Matter* 17 (11 2021), pp. 3113–3120. DOI: 10.1039/D0SM01220C. URL: http://dx.doi.org/10.1039/D0SM01220C.

[52] P Chiarantoni et al. "Work fluctuations of self-propelled particles in the phase separated state." In: *Journal of Physics A: Mathematical and Theoretical* 53.36 (Aug. 2020), 36LT02. DOI: 10.1088/1751-8121/ab8f3c. URL: https://doi.org/10.1088/1751-8121/ab8f3c.

[53] Yen-Liang Chou and Thomas Ihle. "Active matter beyond mean-field: Ring-kinetic theory for self-propelled particles." In: *Phys. Rev. E* 91 (2 Feb. 2015), p. 022103. DOI: 10.1103/PhysRevE.91.022103. URL: https://link.aps.org/doi/10.1103/PhysRevE.91.022103.

[54] Quentin Clairon and Adeline Samson. "Optimal control for estimation in partially observed elliptic and hypoelliptic linear stochastic differential equations." In: *Statistical Inference for Stochastic Processes* 23.1 (2020), pp. 105–127.

[55] Robert B Cleveland et al. "STL: A seasonal-trend decomposition." In: *Journal of Official Statistics* 6 (1 1990), pp. 3–73.

[56] Luca Cocconi et al. "Entropy Production in Exactly Solvable Systems." In: *Entropy* 22.11 (2020). ISSN: 1099-4300. DOI: 10.3390/e22111252. URL: https://www.mdpi.com/1099-4300/22/11/1252.

[57] Katherine Copenhagen et al. "Topological defects promote layer formation in Myxococcus xanthus colonies." In: *Nature Physics* 17 (2 Feb. 2021), pp. 211–215. ISSN: 1745-2481. DOI: 10.1038/s41567-020-01056-4. URL: https://doi.org/10.1038/s41567-020-01056-4.

[58] Antonio Carlos Costa et al. *Maximally predictive ensemble dynamics from data*. 2021. arXiv: 2105.12811 [physics.bio-ph].

[59] Gavin E. Crooks. "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences." In: *Phys. Rev. E* 60 (3 Sept. 1999), pp. 2721–2726. DOI: 10.1103/PhysRevE.60.2721. URL: https://link.aps.org/doi/10.1103/PhysRevE.60.2721.

[60] Gavin E. Crooks. "Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems." In: *Journal of Statistical Physics* 90 (5 Mar. 1998), pp. 1481–1487. ISSN: 1572-9613. DOI: 10.1023/A:1023208217925. URL: https://doi.org/10.1023/A:1023208217925.

[61] Emanuele Crosato, Mikhail Prokopenko, and Richard E. Spinney. "Irreversibility and emergent structure in active matter." In: *Phys. Rev. E* 100 (4 Oct. 2019), p. 042613. DOI: 10.1103/PhysRevE.100.042613. URL: https://link.aps.org/doi/10.1103/PhysRevE.100.042613.

[62] Leticia F. Cugliandolo, Jorge Kurchan, and Luca Peliti. "Energy flow, partial equilibration, and effective temperatures in systems with slow dynamics." In: *Phys. Rev. E* 55 (4 Apr. 1997), pp. 3898–3914. DOI: 10.1103/PhysRevE.55.3898. URL: https://link.aps.org/doi/10.1103/PhysRevE.55.3898.

[63] Lennart Dabelow, Stefano Bo, and Ralf Eichhorn. "Irreversibility in Active Matter Systems: Fluctuation Theorem and Mutual Information." In: *Phys. Rev. X* 9 (2 Apr. 2019), p. 021009. DOI: 10.1103/PhysRevX.9.021009. URL: https://link.aps.org/doi/10.1103/PhysRevX.9.021009.

[64] Julien Deseigne, Olivier Dauchot, and Hugues Chaté. "Collective Motion of Vibrated Polar Disks." In: *Phys. Rev. Lett.* 105 (9 Aug. 2010), p. 098001. DOI: 10.1103/PhysRevLett.105.098001. URL: https://link.aps.org/doi/10.1103/PhysRevLett.105.098001.

[65] R. Di Leonardo et al. "Bacterial ratchet motors." In: *Proceedings of the National Academy of Sciences* 107.21 (2010), pp. 9541–9545. ISSN: 0027-8424. DOI: 10.1073/pnas.0910426107. URL: https://www.pnas.org/content/107/21/9541.

[66] Susanne Ditlevsen and Adeline Samson. "Hypoelliptic diffusions: filtering and inference from complete and partial observations." In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 81.2 (2019), pp. 361–384. DOI: https://doi.org/10.1111/rssb.12307. URL: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssb.12307.

[67] Christopher Dombrowski et al. "Self-Concentration and Large-Scale Coherence in Bacterial Dynamics." In: *Phys. Rev. Lett.* 93 (9 Aug. 2004), p. 098103. DOI: 10.1103/PhysRevLett.93.098103. URL: https://link.aps.org/doi/10.1103/PhysRevLett.93.098103.

[68] A. N. Drozdov and M. Morillo. "Path Integral Solution of the Kramers Problem." In: *Phys. Rev. Lett.* 77 (27 Dec. 1996), pp. 5324–5327. DOI: 10.1103/PhysRevLett.77.5324. URL: https://link.aps.org/doi/10.1103/PhysRevLett.77.5324.

[69] Freeman J. Dyson. "General theory of spin-wave interactions." In: *Physical Review* 102.5 (1956), p. 1217.

[70] Werner Ebeling, Frank Schweitzer, and Benno Tilch. "Active Brownian particles with energy depots modeling animal mobility." In: *Biosystems* 49.1 (1999), pp. 17–29. ISSN: 0303-2647. DOI: https://doi.org/10.1016/S0303-2647(98)00027-6. URL: https://www.sciencedirect.com/science/article/pii/S0303264798000276.

[71] Ola Elerian, Siddhartha Chib, and Neil Shephard. "Likelihood Inference for Discretely Observed Nonlinear Diffusions." In: *Econometrica* 69.4 (2001), pp. 959–993. DOI: https://doi.org/10.1111/1468-0262.00226. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00226.

[72] Bjorn Eraker. "MCMC Analysis of Diffusion Models with Application to Finance." In: *Journal of Business & Economic Statistics* 19.2 (Apr. 2001), pp. 177–191. URL: https://ideas.repec.org/a/bes/jnlbes/v19y2001i2p177-91.html.

[73] Federica Ferretti et al. *A Renormalization Group Approach to Connect Discrete- and Continuous-Time Descriptions of Gaussian Processes*. 2021. arXiv: 2101.06482 [stat.ML].

[74] Federica Ferretti et al. "Building General Langevin Models from Discrete Datasets." In: *Phys. Rev. X* 10 (3 July 2020), p. 031018. DOI: 10.1103/PhysRevX.10.031018. URL: https://link.aps.org/doi/10.1103/PhysRevX.10.031018.

[75] Y Fily et al. "Mechanical pressure and momentum conservation in dry active matter." In: 51.4 (Dec. 2017), p. 044003. DOI: 10.1088/1751-8121/aa99b6. URL: https://doi.org/10.1088/1751-8121/aa99b6.

[76] A. A. Filyukov and V. Ya. Karpov. "Description of steady transport processes by the method of the most probable path of evolution." In: *Journal of engineering physics* 13 (5 Nov. 1967), pp. 326–329. ISSN: 1573-871X. DOI: 10.1007/BF00832348. URL: https://doi.org/10.1007/BF00832348.

[77] R. A. Fisher. "On the mathematical foundations of theoretical statistics." In: *Philosophical Transactions* 222 (Nov. 1921). Ed. by Royal Society of London, pp. 309–368.

[78] Elijah Flenner and Grzegorz Szamel. "Active matter: Quantifying the departure from equilibrium." In: *Phys. Rev. E* 102 (2 Aug. 2020), p. 022607. DOI: 10.1103/PhysRevE.102.022607. URL: https://link.aps.org/doi/10.1103/PhysRevE.102.022607.

[79] Étienne Fodor et al. "How Far from Equilibrium Is Active Matter?" In: *Phys. Rev. Lett.* 117 (3 July 2016), p. 038103. DOI: 10.1103/PhysRevLett.117.038103. URL: https://link.aps.org/doi/10.1103/PhysRevLett.117.038103.

[80] Rudolf Friedrich et al. "Approaching complexity by stochastic methods: From biological systems to turbulence." In: *Physics Reports* 506.5 (2011), pp. 87–162. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2011.05.003. URL: https://www.sciencedirect.com/science/article/pii/S0370157311001530.

[81] Anna Frishman and Pierre Ronceray. "Learning Force Fields from Stochastic Trajectories." In: *Phys. Rev. X* 10 (2 Apr. 2020), p. 021009. DOI: 10.1103/PhysRevX.10.021009. URL: https://link.aps.org/doi/10.1103/PhysRevX.10.021009.

[82] Crispin Gardiner. *Stochastic Methods*. Vol. 13. 0172-7389 4. Springer-Verlag Berlin Heidelberg, 2009.

[83] Jacques Gautrais et al. "Analyzing fish movement as a persistent turning walker." In: *Journal of Mathematical Biology* 58.3 (2009), pp. 429–445. DOI: 10.1007/s00285-008-0198-7. URL: https://doi.org/10.1007/s00285-008-0198-7.

[84] Jacques Gautrais et al. "Deciphering Interactions in Moving Animal Groups." In: *PLOS Computational Biology* 8.9 (Sept. 2012), pp. 1–11. DOI: 10.1371/journal.pcbi.1002678. URL: https://doi.org/10.1371/journal.pcbi.1002678.

[85] Fabien Gerbal et al. "An Elastic Analysis of Listeria monocytogenes Propulsion." In: *Biophysical Journal* 79.5 (2000), pp. 2259–2275. ISSN: 0006-3495. DOI: https://doi.org/10.1016/S0006-3495(00)76473-3. URL: http://www.sciencedirect.com/science/article/pii/S0006349500764733.

[86] Daniel T. Gillespie. "Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral." In: *Phys. Rev. E* 54 (2 Aug. 1996), pp. 2084–2091. DOI: 10.1103/PhysRevE.54.2084. URL: https://link.aps.org/doi/10.1103/PhysRevE.54.2084.

[87] Francesco Ginelli and Hugues Chaté. "Relevance of Metric-Free Interactions in Flocking Phenomena." In: *Phys. Rev. Lett.* 105 (16 Oct. 2010), p. 168103. DOI: 10.1103/PhysRevLett.105.168103. URL: https://link.aps.org/doi/10.1103/PhysRevLett.105.168103.

[88] Félix Ginot et al. "Nonequilibrium Equation of State in Suspensions of Active Colloids." In: *Phys. Rev. X* 5 (1 Jan. 2015), p. 011004. DOI: 10.1103/PhysRevX.5.011004. URL: https://link.aps.org/doi/10.1103/PhysRevX.5.011004.

[89] Arnaud Gloter. "Parameter estimation for a discrete sampling of an intergrated Ornstein-Uhlenbeck process." In: *Statistics* 35.3 (2001), pp. 225–243. DOI: 10.1080/02331880108802733. URL: https://doi.org/10.1080/02331880108802733.

[90] Arnaud Gloter. "Parameter Estimation for a Discretely Observed Integrated Diffusion Process." In: *Scandinavian Journal of Statistics* 33.1 (2006), pp. 83–104. DOI: 10.1111/j.1467-9469.2006.00465.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9469.2006.00465.x.

[91] Federico S. Gnesotto et al. "Learning the non-equilibrium dynamics of Brownian movies." In: *Nature Communications* 11 (1 2020), p. 5378. DOI: 10.1038/s41467-020-18796-9. URL: https://doi.org/10.1038/s41467-020-18796-9.

[92] Gerhard Gompper et al. "The 2020 motile active matter roadmap." In: *Journal of Physics: Condensed Matter* 32.19 (Feb. 2020), p. 193001. DOI: 10.1088/1361-648x/ab6348. URL: https://doi.org/10.1088/1361-648x/ab6348.

[93] W.F. van Gunsteren and H.J.C. Berendsen. "Algorithms for brownian dynamics." In: *Molecular Physics* 45.3 (1982), pp. 637–647. DOI: 10.1080/00268978200100491. URL: https://doi.org/10.1080/00268978200100491.

[94] Takahiro Harada and Shin-ichi Sasa. "Energy dissipation and violation of the fluctuation-response relation in nonequilibrium Langevin systems." In: *Phys. Rev. E* 73 (2 Feb. 2006), p. 026131. DOI: 10.1103/PhysRevE.73.026131. URL: https://link.aps.org/doi/10.1103/PhysRevE.73.026131.

[95] Takahiro Harada and Shin-ichi Sasa. "Equality Connecting Energy Dissipation with a Violation of the Fluctuation-Response Relation." In: *Phys. Rev. Lett.* 95 (13 Sept. 2005), p. 130602. DOI: 10.1103/PhysRevLett.95.130602. URL: https://link.aps.org/doi/10.1103/PhysRevLett.95.130602.

[96] Viktor Holubec et al. "Finite-Size Scaling at the Edge of Disorder in a Time-Delay Vicsek Model." In: *Physical Review Letters* 127.25 (Dec. 2021). ISSN: 1079-7114. DOI: 10.1103/physrevlett.127.258001. URL: http://dx.doi.org/10.1103/PhysRevLett.127.258001.

[97] Illia Horenko et al. "Data-based parameter estimation of generalized multidimensional Langevin processes." In: *Phys. Rev. E* 76 (1 July 2007), p. 016706. DOI: 10.1103/PhysRevE.76.016706. URL: https://link.aps.org/doi/10.1103/PhysRevE.76.016706.

[98] J. Horowitz and C. Jarzynski. "Comment on "Failure of the Work-Hamiltonian Connection for Free-Energy Calculations"." In: *Phys. Rev. Lett.* 101 (9 Aug. 2008), p. 098901. DOI: 10.1103/PhysRevLett.101.098901. URL: https://link.aps.org/doi/10.1103/PhysRevLett.101.098901.

[99] *https://www.cgal.org*.

[100] Thomas Ihle. "Chapman–Enskog expansion for the Vicsek model of self-propelled particles." In: *Journal of Statistical Mechanics: Theory and Experiment* 2016.8 (Aug. 2016), p. 083205. DOI: 10.1088/1742-5468/2016/08/083205. URL: https://doi.org/10.1088/1742-5468/2016/08/083205.

[101] E. T. Jaynes. "Information Theory and Statistical Mechanics." In: *Phys. Rev.* 106 (4 May 1957), pp. 620–630. DOI: 10.1103/PhysRev.106.620. URL: https://link.aps.org/doi/10.1103/PhysRev.106.620.

[102] E. T. Jaynes. "Information Theory and Statistical Mechanics. II." In: *Phys. Rev.* 108 (2 Oct. 1957), pp. 171–190. DOI: 10.1103/PhysRev.108.171. URL: https://link.aps.org/doi/10.1103/PhysRev.108.171.

[103] E.T. Jaynes. "On the rationale of maximum-entropy methods." In: *Proceedings of the IEEE* 70.9 (1982), pp. 939–952. DOI: 10.1109/PROC.1982.12425.

[104] Jiteng Jia, Tomohiro Sogabe, and Moawwad El-Mikkawy. "Inversion of k-tridiagonal matrices with Toeplitz structure." In: *Computers & Mathematics with Applications* 65.1 (2013), pp. 116–125. ISSN: 0898-1221. DOI: https://doi.org/10.1016/j.camwa.2012.11.001. URL: http://www.sciencedirect.com/science/article/pii/S0898122112006475.

[105] Leo P Kadanoff. "Notes on Migdal's recursion formulas." In: *Annals of Physics* 100 (1976), pp. 359–394.

[106] R. E. Kalman. "A New Approach to Linear Filtering and Prediction Problems." In: *Journal of Basic Engineering* 82.1 (Mar. 1960), pp. 35–45. ISSN: 0021-9223. DOI: 10.1115/1.3662552. URL: https://doi.org/10.1115/1.3662552.

[107] Holger Kantz and Thomas Schreiber. *Nonlinear Time Series Analysis*. 2nd ed. Cambridge University Press, 2003. DOI: 10.1017/CBO9780511755798.

[108] B. O. Koopman. "On Distributions Admitting a Sufficient Statistic." In: *Transactions of the American Mathematical Society* 39.3 (1936), pp. 399–409. ISSN: 00029947. URL: http://www.jstor.org/stable/1989758.

[109] R Kubo. "The fluctuation-dissipation theorem." In: *Reports on Progress in Physics* 29.1 (Jan. 1966), pp. 255–284. DOI: 10.1088/0034-4885/29/1/306. URL: https://doi.org/10.1088/0034-4885/29/1/306.

[110] Arshad Kudrolli et al. "Swarming and Swirling in Self-Propelled Polar Granular Rods." In: *Phys. Rev. Lett.* 100 (5 Feb. 2008), p. 058001. DOI: 10.1103/PhysRevLett.100.058001. URL: https://link.aps.org/doi/10.1103/PhysRevLett.100.058001.

[111] Yoshiki Kuramoto. "Self-entrainment of a population of coupled non-linear oscillators." In: *International Symposium on Mathematical Problems in Theoretical Physics*. Ed. by Huzihiro Araki. Berlin, Heidelberg: Springer Berlin Heidelberg, 1975, pp. 420–422. ISBN: 978-3-540-37509-8.

[112] Rüdiger Kürsten and Thomas Ihle. "Dry Active Matter Exhibits a Self-Organized Cross Sea Phase." In: *Phys. Rev. Lett.* 125 (18 Oct. 2020), p. 188003. DOI: 10.1103/PhysRevLett.125.188003. URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.188003.

[113] Rüdiger Kürsten and Thomas Ihle. "Quantitative kinetic theory of flocking with three-particle closure." In: *Phys. Rev. E* 104 (3 Sept. 2021), p. 034604. DOI: 10.1103/PhysRevE.104.034604. URL: https://link.aps.org/doi/10.1103/PhysRevE.104.034604.

[114] Joel L. Lebowitz and Herbert Spohn. "A Gallavotti–Cohen-Type Symmetry in the Large Deviation Functional for Stochastic Dynamics." In: *Journal of Statistical Physics* 95 (1 1999), pp. 333–365. DOI: 10.1023/A:1004589714161. URL: https://doi.org/10.1023/A:1004589714161.

[115] Julian Lee and Steve Pressé. "A derivation of the master equation from path entropy maximization." In: *The Journal of Chemical Physics* 137.7 (2012), p. 074103. DOI: 10.1063/1.4743955. URL: https://doi.org/10.1063/1.4743955.

[116] B. Lehle and J. Peinke. "Analyzing a stochastic process driven by Ornstein-Uhlenbeck noise." In: *Phys. Rev. E* 97 (1 Jan. 2018), p. 012113. DOI: 10.1103/PhysRevE.97.012113. URL: https://link.aps.org/doi/10.1103/PhysRevE.97.012113.

[117] B. Lehle and J. Peinke. "Analyzing a stochastic time series obeying a second-order differential equation." In: *Phys. Rev. E* 91 (6 June 2015), p. 062113. DOI: 10.1103/PhysRevE.91.062113. URL: https://link.aps.org/doi/10.1103/PhysRevE.91.062113.

[118] Marjolein N. van der Linden et al. "Interrupted Motility Induced Phase Separation in Aligning Active Colloids." In: *Phys. Rev. Lett.* 123 (9 Aug. 2019), p. 098001. DOI: 10.1103/PhysRevLett.123.098001. URL: https://link.aps.org/doi/10.1103/PhysRevLett.123.098001.

[119] Davide Loi, Stefano Mossa, and Leticia F. Cugliandolo. "Effective temperature of active matter." In: *Phys. Rev. E* 77 (5 May 2008), p. 051111. DOI: 10.1103/PhysRevE.77.051111. URL: https://link.aps.org/doi/10.1103/PhysRevE.77.051111.

[120] Alejandra López-Pérez, Manuel Febrero-Bande, and Wencesalo González-Manteiga. "Parametric Estimation of Diffusion Processes: A Review and Comparative Study." In: *Mathematics* 9.8 (2021). ISSN: 2227-7390. DOI: 10.3390/math9080859. URL: https://www.mdpi.com/2227-7390/9/8/859.

[121] S. Machlup and L. Onsager. "Fluctuations and Irreversible Process. II. Systems with Kinetic Energy." In: *Phys. Rev.* 91 (6 Sept. 1953), pp. 1512–1515. DOI: 10.1103/PhysRev.91.1512. URL: https://link.aps.org/doi/10.1103/PhysRev.91.1512.

[122] David J.C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.

[123] Christian Maes, Frank Redig, and Annelies Van Moffaert. "On the definition of entropy production, via examples." In: *Journal of Mathematical Physics* 41.3 (2000), pp. 1528–1554. DOI: 10.1063/1.533195. URL: https://doi.org/10.1063/1.533195.

[124] Claudio Maggi et al. *Critical active dynamics is captured by a colored-noise driven field theory*. 2021. arXiv: 2108.13971 [cond-mat.soft].

[125] Claudio Maggi et al. "Universality class of the motility-induced critical point in large scale off-lattice simulations of active particles." In: *Soft Matter* 17 (14 2021), pp. 3807–3812. DOI: 10.1039/D0SM02162H. URL: http://dx.doi.org/10.1039/D0SM02162H.

[126] M. C. Marchetti et al. "Hydrodynamics of soft active matter." In: *Rev. Mod. Phys.* 85 (3 July 2013), pp. 1143–1189. DOI: 10.1103/RevModPhys.85.1143. URL: https://link.aps.org/doi/10.1103/RevModPhys.85.1143.

[127] Umberto Marini Bettolo Marconi et al. "Fluctuationâdissipation: Response theory in statistical physics." In: *Physics Reports* 461.4 (2008), pp. 111–195. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2008.02.002. URL: https://www.sciencedirect.com/science/article/pii/S0370157308000768.

[128] Umberto Marini Bettolo Marconi and Claudio Maggi. "Towards a statistical mechanical theory of active fluids." In: *Soft Matter* 11 (45 2015), pp. 8768–8781. DOI: 10.1039/C5SM01718A. URL: http://dx.doi.org/10.1039/C5SM01718A.

[129] David Martin et al. "Fluctuation-Induced Phase Separation in Metric and Topological Models of Collective Motion." In: *Phys. Rev. Lett.* 126 (14 Apr. 2021), p. 148001. DOI: 10.1103/PhysRevLett.126.148001. URL: https://link.aps.org/doi/10.1103/PhysRevLett.126.148001.

[130] David Martin et al. "Statistical mechanics of active Ornstein-Uhlenbeck particles." In: *Phys. Rev. E* 103 (3 Mar. 2021), p. 032607. DOI: 10.1103/PhysRevE.103.032607. URL: https://link.aps.org/doi/10.1103/PhysRevE.103.032607.

[131] G. Mazenko. *Nonequilibrium Statistical Mechanics*. Wiley-VCH, 2006.

[132] Oliver A. McBryan and Thomas Spencer. "On the decay of correlations inSO(n)-symmetric ferromagnets." In: *Communications in Mathematical Physics* 53.3 (1977), pp. 299–302. DOI: 10.1007/BF01609854. URL: https://doi.org/10.1007/BF01609854.

[133] Nathaniel David Mermin and Herbert Wagner. "Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic Heisenberg models." In: *Physical Review Letters* 17.22 (Nov. 1966), p. 1133.

[134] Gérard. Meurant. "A Review on the Inverse of Symmetric Tridiagonal and Block Tridiagonal Matrices." In: *SIAM Journal on Matrix Analysis and Applications* 13.3 (1992), pp. 707–728. DOI: 10.1137/0613045. URL: https://doi.org/10.1137/0613045.

[135] M. San Miguel and J. M. Sancho. "A colored-noise approach to Brownian motion in position space. Corrections to the Smoluchowski equation." In: *Journal of Statistical Physics* 22.5 (May 1980), pp. 605–624. ISSN: 1572-9613. DOI: 10.1007/BF01011341. URL: https://doi.org/10.1007/BF01011341.

[136] Thierry Mora et al. "Local equilibrium in bird flocks." In: *Nature Physics* 12 (Aug. 2016), 1153 EP -. URL: https://doi.org/10.1038/nphys3846.

[137] Cesare Nardini et al. "Entropy Production in Field Theories without Time-Reversal Symmetry: Quantifying the Non-Equilibrium Character of Active Matter." In: *Phys. Rev. X* 7 (2 Apr. 2017), p. 021007. DOI: 10.1103/PhysRevX.7.021007. URL: https://link.aps.org/doi/10.1103/PhysRevX.7.021007.

[138] H. Chau Nguyen, Riccardo Zecchina, and Johannes Berg. "Inverse statistical problems: from the inverse Ising problem to data science." In: *Advances in Physics* 66.3 (June 2017), pp. 197–261. ISSN: 1460-6976. DOI: 10.1080/00018732.2017.1341604. URL: http://dx.doi.org/10.1080/00018732.2017.1341604.

[139] Daiki Nishiguchi and Masaki Sano. "Mesoscopic turbulence and local order in Janus particles self-propelling under an ac electric field." In: *Phys. Rev. E* 92 (5 Nov. 2015), p. 052309. DOI: 10.1103/PhysRevE.92.052309. URL: https://link.aps.org/doi/10.1103/PhysRevE.92.052309.

[140] Jérémy O'Byrne et al. *Time-(ir)reversibility in active matter: from micro to macro*. 2021. arXiv: 2104.03030 [cond-mat.stat-mech].

[141] Ahmad K. Omar et al. *Tuning Nonequilibrium Phase Transitions with Inertia*. 2021. arXiv: 2108.10278 [cond-mat.soft].

[142] L. Onsager and S. Machlup. "Fluctuations and Irreversible Processes." In: *Phys. Rev.* 91 (6 Sept. 1953), pp. 1505–1512. DOI: 10.1103/PhysRev.91.1505. URL: https://link.aps.org/doi/10.1103/PhysRev.91.1505.

[143] Jeremie Palacci et al. "Living Crystals of Light-Activated Colloidal Surfers." In: *Science* 339.6122 (2013), pp. 936–940. DOI: 10.1126/science.1230020. URL: https://www.science.org/doi/abs/10.1126/science.1230020.

[144] Giorgio Parisi. *Statistical field theory*. Frontiers in Physics. Redwood City, CA: Addison-Wesley, 1988. URL: https://cds.cern.ch/record/111935.

[145] Benjamin Partridge and Chiu Fan Lee. "Critical Motility-Induced Phase Separation Belongs to the Ising Universality Class." In: *Phys. Rev. Lett.* 123 (6 Aug. 2019), p. 068002. DOI: 10.1103/PhysRevLett.123.068002. URL: https://link.aps.org/doi/10.1103/PhysRevLett.123.068002.

[146] Aurelio Patelli. "Landau kinetic equation for dry aligning active models." In: *Journal of Statistical Mechanics: Theory and Experiment* 2021.3 (Mar. 2021), p. 033210. DOI: 10.1088/1742-5468/abe410. URL: https://doi.org/10.1088/1742-5468/abe410.

[147] Jonas N. Pedersen et al. "How to connect time-lapse recorded trajectories of motile microorganisms with dynamical models in continuous time." In: *Phys. Rev. E* 94 (6 Dec. 2016), p. 062401. DOI: 10.1103/PhysRevE.94.062401. URL: https://link.aps.org/doi/10.1103/PhysRevE.94.062401.

[148] Luca Peliti. "Comment on "Failure of the Work-Hamiltonian Connection for Free-Energy Calculations"." In: *Phys. Rev. Lett.* 101 (9 Aug. 2008), p. 098903. DOI: 10.1103/PhysRevLett.101.098903. URL: https://link.aps.org/doi/10.1103/PhysRevLett.101.098903.

[149] Luca Peliti. "On the work–Hamiltonian connection in manipulated systems." In: *Journal of Statistical Mechanics: Theory and Experiment* 2008 (May 2008), P05002. DOI: 10.1088/1742-5468/2008/05/P05002.

[150] A. Peshkov et al. "Boltzmann-Ginzburg-Landau approach for continuous descriptions of generic Vicsek-like models." In: *The European Physical Journal Special Topics* 223 (7 June 2014), pp. 1315–1344. ISSN: 1951-6401. DOI: 10.1140/epjst/e2014-02193-y. URL: https://doi.org/10.1140/epjst/e2014-02193-y.

[151] Anton Peshkov et al. "Continuous Theory of Active Matter Systems with Metric-Free Interactions." In: *Phys. Rev. Lett.* 109 (9 Aug. 2012), p. 098101. DOI: 10.1103/PhysRevLett.109.098101. URL: https://link.aps.org/doi/10.1103/PhysRevLett.109.098101.

[152] E. J. G. Pitman. "Sufficient statistics and intrinsic accuracy." In: *Mathematical Proceedings of the Cambridge Philosophical Society* 32.4 (1936), pp. 567–579. DOI: 10.1017/S0305004100019307.

[153] Peter E. Kloeden Eckhard Platen and Peter E. Kloeden. *Numerical Solution of Stochastic Differential Equations*. 1st ed. Vol. 23. Stochastic Modelling and Applied Probability. Springer-Verlag Berlin Heidelberg, 1992.

[154] Yvo Pokern, Andrew M. Stuart, and Petter Wiberg. "Parameter estimation for partially observed hypoelliptic diffusions." In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71.1 (2009), pp. 49–73. DOI: https://doi.org/10.1111/j.1467-9868.2008.00689.x. URL: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2008.00689.x.

[155] M. Poujade et al. "Collective migration of an epithelial monolayer in response to a model wound." In: *Proceedings of the National Academy of Sciences* 104.41 (2007), pp. 15988–15993. ISSN: 0027-8424. DOI: 10.1073/pnas.0705062104. URL: https://www.pnas.org/content/104/41/15988.

[156] Steve Pressé et al. "Principles of maximum entropy and maximum caliber in statistical physics." In: *Rev. Mod. Phys.* 85 (3 July 2013), pp. 1115–1141. DOI: 10.1103/RevModPhys.85.1115. URL: https://link.aps.org/doi/10.1103/RevModPhys.85.1115.

[157] Andrea Procaccini et al. "Propagating waves in starling, Sturnus vulgaris, flocks under predation." In: *Animal Behaviour* 82.4 (2011), pp. 759–765. ISSN: 0003-3472. DOI: https://doi.org/10.1016/j.anbehav.2011.07.006. URL: https://www.sciencedirect.com/science/article/pii/S0003347211002867.

[158] Sriram Ramaswamy. "The Mechanics and Statistics of Active Matter." In: *Annual Review of Condensed Matter Physics* 1.1 (2010), pp. 323–345. DOI: 10.1146/annurev-conmatphys-070909-104101. URL: https://doi.org/10.1146/annurev-conmatphys-070909-104101.

[159] Hannes Risken. *The Fokker-Planck equation*. Springer, 1996.

[160] P. Romanczuk et al. "Active Brownian particles." In: *The European Physical Journal Special Topics* 202 (1 2012), pp. 1–162. DOI: 10.1140/epjst/e2012-01529-y. URL: https://doi.org/10.1140/epjst/e2012-01529-y.

[161] Olivia du Roure et al. "Force mapping in epithelial cell migration." In: *Proceedings of the National Academy of Sciences* 102.7 (2005), pp. 2390–2395. ISSN: 0027-8424. DOI: 10.1073/pnas.0408482102. URL: https://www.pnas.org/content/102/7/2390.

[162] Adeline Samson and Michèle Thieullen. "A contrast estimator for completely or partially observed hypoelliptic diffusion." In: *Stochastic Processes and their Applications* 122.7 (2012), pp. 2521–2552. ISSN: 0304-4149. DOI: https://doi.org/10.1016/j.spa.2012.04.006. URL: https://www.sciencedirect.com/science/article/pii/S0304414912000671.

[163] Tim Sanchez et al. "Spontaneous motion in hierarchically assembled active matter." In: *Nature* 491 (7424 Nov. 2012), pp. 431–434. ISSN: 1476-4687. DOI: 10.1038/nature11591. URL: https://doi.org/10.1038/nature11591.

[164] Volker Schaller et al. "Polar patterns of driven filaments." In: *Nature* 467 (7311 Sept. 2010), pp. 73–77. ISSN: 1476-4687. DOI: 10.1038/nature09312. URL: https://doi.org/10.1038/nature09312.

[165] M. Schienbein and H. Gruler. "Langevin equation, Fokker-Planck equation and cell migration." In: *Bulletin of Mathematical Biology* 55 (3 1993), pp. 585–608. DOI: 10.1007/BF02460652. URL: https://doi.org/10.1007/BF02460652.

[166] U. Seifert. "Stochastic thermodynamics: principles and perspectives." In: *The European Physical Journal B* 64 (3 Aug. 2008), pp. 423–431. ISSN: 1434-6036. DOI: 10.1140/epjb/e2008-00001-9. URL: https://doi.org/10.1140/epjb/e2008-00001-9.

[167] Udo Seifert. "Entropy Production along a Stochastic Trajectory and an Integral Fluctuation Theorem." In: *Phys. Rev. Lett.* 95 (4 July 2005), p. 040602. DOI: 10.1103/PhysRevLett.95.040602. URL: https://link.aps.org/doi/10.1103/PhysRevLett.95.040602.

[168] Udo Seifert. "Stochastic thermodynamics, fluctuation theorems and molecular machines." In: *Reports on Progress in Physics* 75.12 (Nov. 2012), p. 126001. DOI: 10.1088/0034-4885/75/12/126001. URL: https://doi.org/10.1088/0034-4885/75/12/126001.

[169] Ken Sekimoto. "Kinetic Characterization of Heat Bath and the Energetics of Thermal Ratchet Models." In: *Journal of the Physical Society of Japan* 66.5 (1997), pp. 1234–1237. DOI: 10.1143/jpsj.66.1234.

[170] Suraj Shankar and M. Cristina Marchetti. "Hidden entropy production and work fluctuations in an ideal active gas." In: *Phys. Rev. E* 98 (2 Aug. 2018), p. 020604. DOI: 10.1103/PhysRevE.98.020604. URL: https://link.aps.org/doi/10.1103/PhysRevE.98.020604.

[171] Charles Edward Shannon. "A mathematical theory of communication." In: *The Bell System Technical Journal* 27 (1948), pp. 379, 623.

[172] J. Shore and R. Johnson. "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy." In: *IEEE Transactions on Information Theory* 26.1 (1980), pp. 26–37. DOI: 10.1109/TIT.1980.1056144.

[173] Robert D. Skeel and Jesus A. Izaguirre. "An impulse integrator for Langevin dynamics." In: *Molecular Physics* 100.24 (2002), pp. 3885–3891. DOI: 10.1080/00268970210000183 21. URL: https://doi.org/10.1080/00268970210000183 21.

[174] A. P. Solon, M. E. Cates, and J. Tailleur. "Active brownian particles and run-and-tumble particles: A comparative study." In: *The European Physical Journal Special Topics* 224 (7 July 2015). ISSN: 1951-6401. DOI: 10.1140/epjst/e2015-02457-0. URL: https://doi.org/10.1140/epjst/e2015-02457-0.

[175] A. P. Solon et al. "Pressure is not a state function for generic active fluids." In: *Nature Physics* 11 (8 2015), pp. 673–678. DOI: 10.1038/nphys3377. URL: https://doi.org/10.1038/nphys3377.

[176] Alexandre P. Solon, Hugues Chaté, and Julien Tailleur. "From Phase to Microphase Separation in Flocking Models: The Essential Role of Nonequilibrium Fluctuations." In: *Phys. Rev. Lett.* 114 (6 Feb. 2015), p. 068101. DOI: 10.1103/PhysRevLett.114.068101. URL: https://link.aps.org/doi/10.1103/PhysRevLett.114.068101.

[177] Richard E. Spinney and Ian J. Ford. "Entropy production in full phase space for continuous stochastic dynamics." In: *Phys. Rev. E* 85 (5 May 2012), p. 051113. DOI: 10.1103/PhysRevE.85.051113. URL: https://link.aps.org/doi/10.1103/PhysRevE.85.051113.

[178] J. Stark. "Delay Embeddings for Forced Systems. I. Deterministic Forcing." In: *Journal of Nonlinear Science* 9.3 (1999), pp. 255–332.

[179] J. Stark et al. "Delay Embeddings for Forced Systems.II. Stochastic Forcing." In: *Journal of Nonlinear Science* 13.6 (2003), pp. 519–577.

[180] Greg J. Stephens et al. "Dimensionality and Dynamics in the Behavior of C. elegans." In: *PLOS Computational Biology* 4.4 (Apr. 2008), pp. 1–10. DOI: 10.1371/journal.pcbi.1000028. URL: https://doi.org/10.1371/journal.pcbi.1000028.

[181] Yutaka Sumino et al. "Large-scale vortex lattice emerging from collectively moving microtubules." In: *Nature* 483.7390 (2012), p. 448.

[182] B. Szabó et al. "Phase transition in the collective migration of tissue cells: Experiment and model." In: *Phys. Rev. E* 74 (6 Dec. 2006), p. 061908. DOI: 10.1103/PhysRevE.74.061908. URL: https://link.aps.org/doi/10.1103/PhysRevE.74.061908.

[183] F. Takens. "Detecting strange attractors in turbulence." In: *Lecture notes in mathematics*. Vol. 898. Springer-Verlag, 1981, pp. 366–381.

[184] Hal Tasaki. "Hohenberg-Mermin-Wagner-Type Theorems for Equilibrium Models of Flocking." In: *Phys. Rev. Lett.* 125 (22 Nov. 2020), p. 220601. DOI: 10.1103/PhysRevLett.125.220601. URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.220601.

[185] "The Fluctuation Theorem as a Gibbs Property." In: *Journal of Statistical Physics* 95 (1 Apr. 1999), pp. 367–392. ISSN: 1572-9613. DOI: 10.1023/A:1004541830999. URL: https://doi.org/10.1023/A:1004541830999.

[186] I. Theurkauff et al. "Dynamic Clustering in Active Colloidal Suspensions with Chemical Signaling." In: *Phys. Rev. Lett.* 108 (26 June 2012), p. 268303. DOI: 10.1103/PhysRevLett.108.268303. URL: https://link.aps.org/doi/10.1103/PhysRevLett.108.268303.

[187] John Toner and Yuhai Tu. "Flocks, herds, and schools: A quantitative theory of flocking." In: *Phys. Rev. E* 58 (4 Oct. 1998), pp. 4828–4858. DOI: 10.1103/PhysRevE.58.4828. URL: https://link.aps.org/doi/10.1103/PhysRevE.58.4828.

[188] John Toner and Yuhai Tu. "Long-Range Order in a Two-Dimensional Dynamical XY Model: How Birds Fly Together." In: *Phys. Rev. Lett.* 75 (23 Dec. 1995), pp. 4326–4329. DOI: 10.1103/PhysRevLett.75.4326. URL: https://link.aps.org/doi/10.1103/PhysRevLett.75.4326.

[189] John Toner, Yuhai Tu, and Sriram Ramaswamy. "Hydrodynamics and phases of flocks." In: *Annals of Physics* 318.1 (2005). Special Issue, pp. 170–244. ISSN: 0003-4916. DOI: https://doi.org/10.1016/j.aop.2005.04.011. URL: https://www.sciencedirect.com/science/article/pii/S0003491605000540.

[190] Yuhai Tu, John Toner, and Markus Ulm. "Sound Waves and the Absence of Galilean Invariance in Flocks." In: *Phys. Rev. Lett.* 80 (21 May 1998), pp. 4819–4822. DOI: 10.1103/PhysRevLett.80.4819. URL: https://link.aps.org/doi/10.1103/PhysRevLett.80.4819.

[191] Tamás Vicsek and Anna Zafeiris. "Collective motion." In: *Physics Reports* 517.3 (2012). Collective motion, pp. 71–140. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2012.03.004. URL: https://www.sciencedirect.com/science/article/pii/S0370157312000968.

[192] Tamás Vicsek et al. "Novel Type of Phase Transition in a System of Self-Driven Particles." In: *Phys. Rev. Lett.* 75 (6 Aug. 1995), pp. 1226–1229. DOI: 10.1103/PhysRevLett.75.1226. URL: https://link.aps.org/doi/10.1103/PhysRevLett.75.1226.

[193] Jose M. G. Vilar and J. Miguel Rubi. "Failure of the Work-Hamiltonian Connection for Free-Energy Calculations." In: *Phys. Rev. Lett.* 100 (2 Jan. 2008), p. 020601. DOI: 10.1103/PhysRevLett.100.020601. URL: https://link.aps.org/doi/10.1103/PhysRevLett.100.020601.

[194] Nailong Wu. *The Maximum Entropy method*. Springer Series in Information Sciences. Springer, 1997.

[195] Xingbo Yang and M. Cristina Marchetti. "Hydrodynamics of Turning Flocks." In: *Phys. Rev. Lett.* 115 (25 Dec. 2015), p. 258101. DOI: 10.1103/PhysRevLett.115.258101. URL: https://link.aps.org/doi/10.1103/PhysRevLett.115.258101.

[196] E C Zeeman. "Stability of dynamical systems." In: *Nonlinearity* 1.1 (Feb. 1988), pp. 115–155. DOI: 10.1088/0951-7715/1/1/005. URL: https://doi.org/10.1088/0951-7715/1/1/005.

[197] H. P. Zhang et al. "Collective motion and density fluctuations in bacterial colonies." In: *Proceedings of the National Academy of Sciences* 107.31 (2010), pp. 13626–13630. ISSN: 0027-8424. DOI: 10.1073/pnas.1001651107. URL: https://www.pnas.org/content/107/31/13626.

[198] Yinong Zhao et al. "Phases and homogeneous ordered states in alignment-based self-propelled particle models." In: *Phys. Rev. E* 104 (4 Oct. 2021), p. 044605. DOI: 10.1103/PhysRevE.104.044605. URL: https://link.aps.org/doi/10.1103/PhysRevE.104.044605.

[199] Francesca Zoratto et al. "Behavioural response of European starlings exposed to video playback of conspecific flocks: Effect of social context and predator threat." In: *Behavioural Processes* 103 (2014), pp. 269–277. ISSN: 0376-6357. DOI: https://doi.org/10.1016/j.beproc.2014.01.012. URL: https://www.sciencedirect.com/science/article/pii/S0376635714000266.

[200] Robert Zwanzig. *Nonequilibrium statistical mechanics*. Oxford University Press, USA, 2001.

A EXPLICIT INFERENCE FORMULAS

We report in this Appendix the explicit inference formulas used to compute the Euler-ML, Toeplitz-ML and non-Bayesian parameter estimators in the numerical analysis shown in Chapter 3.

A.1 EULER–ML ESTIMATORS

We considered three variants of the Euler-Maruyama scheme obtained by computing the derivatives in an asymmetric way centered on the pre-point (EM-fwd) or on the post-point (EM-bkd) of the time interval, or in a symmetrized way (BBK method [23]). Although the symmetric variant is known to be more stable, in the small Δt limit the three variants behave in the same way, for our inference purposes.

Given any process in the Kramers class:

$$\dot{x} = v, \qquad \dot{v} = -\eta v + f(x;\boldsymbol{\mu}) + \sigma(x;\boldsymbol{\nu})\xi \tag{263}$$

with $\xi(t)\xi(t') = \delta(t-t')$, the discrete update rules in the x space derived from the three discretizations are:

$$[\text{EM-fwd}]: \; x_{n+1} - 2x_n + x_{n-1} + \eta\Delta t(x_n - x_{n-1}) - f(x_{n-1};\boldsymbol{\mu})\Delta t^2 = \Delta t^{3/2}\sigma(x_{n-1};\boldsymbol{\nu})r_n; \tag{264}$$

$$[\text{EM-bkd}]: \; x_{n+1} - 2x_n + x_{n-1} + \eta\Delta t(x_{n+1} - x_n) - f(x_n;\boldsymbol{\mu})\Delta t^2 = \Delta t^{3/2}\sigma(x_n;\boldsymbol{\nu})r_n; \tag{265}$$

$$[\text{BBK}]: \; x_{n+1} - 2x_n + x_{n-1} + \frac{\eta\Delta t}{2}(x_{n+1} - x_{n-1}) - f(x_n;\boldsymbol{\mu})\Delta t^2 = \Delta t^{3/2}\sigma(x_n;\boldsymbol{\nu})r_n; \tag{266}$$

where the r_n's are Gaussian random variables satisfying $\langle r_n r_m\rangle = \delta_{nm}$. Let us notice that the backward discretization, Eq. (265), does not correspond to a true backward integration of the stochastic process: only the drag term is computed according to this rule, while force and diffusion terms are computed on the central time point. This choice allows for an easy calculation of the Jacobian of the transformation $r_{1:L-1} \mapsto x_{2:L}$.

From Eqs. (264)–(266) the minus log-likelihood $\mathcal{L} = -\log P(x_2 \dots x_L | x_0, x_1)$ is found:

$$[\text{EM-fwd}] \quad \mathcal{L} = \sum_{n=1}^{L-1} \frac{1}{2}\log\sigma^2(x_{n-1};\boldsymbol{\nu}) + \frac{1}{2\sigma^2(x_{n-1};\boldsymbol{\nu})\Delta t^3}\Big[x_{n+1} - 2x_n + x_{n-1} + \eta\Delta t(x_n - x_{n-1}) - f(x_{n-1};\boldsymbol{\mu})\Delta t^2\Big]^2; \tag{267}$$

$$\text{[EM-bkd]} \quad \mathcal{L} = \sum_{n=1}^{L-1} \frac{1}{2}\log\sigma^2(x_n;\boldsymbol{\nu}) - \frac{1}{2}\log(1+\eta\Delta t) + \frac{1}{2\sigma^2(x_n;\boldsymbol{\nu})\Delta t^3}\left[x_{n+1} - 2x_n + x_{n-1} + \eta\Delta t(x_{n+1}-x_n) - f(x_n;\boldsymbol{\mu})\Delta t^2\right]^2 ; \tag{268}$$

$$\text{[BBK]} \quad \mathcal{L} = \sum_{n=1}^{L-1} \frac{1}{2}\log\sigma^2(x_n;\nu) - \frac{1}{2}\log(1+\frac{\eta\Delta}{2}t) + \frac{1}{2\sigma^2(x_n;\boldsymbol{\nu})\Delta t^3}\left[x_{n+1} - 2x_n + x_{n-1} + \frac{\eta\Delta t}{2}(x_{n+1}-x_{n-1}) - f(x_{n-1};\boldsymbol{\mu})\Delta t^2\right]^2 . \tag{269}$$

In the case of a multiplicative process, it is important that $\sigma^2(x)$ is strictly positive, in order to be able to wrote the discretized path probability. The same procedure we adopted to derive Eqs. (267)–(269) can be applied to higher-dimensional processes.

The set of parameters to estimate are $\boldsymbol{\theta} = \{\eta, \boldsymbol{\mu}, \boldsymbol{\nu}\}$: once the functional dependence of $f(x;\boldsymbol{\mu})$ and $\sigma(x;\boldsymbol{\nu})$ is specified, estimators can be derived through an analytical or numerical minimization procedure (or hybrid). We report here the explicit formulas used for the processes studied in Sec. 3.3.

Linear additive processes

This model describes, in one dimension, a stochastic harmonic oscillator in thermal equilibrium with a heat bath: $f(x;\boldsymbol{\mu}) = -\omega_0^2 x$, $\sigma(x;\boldsymbol{\nu}) = \sqrt{2T\eta}$. the parameters to infer are thus $\boldsymbol{\theta} = \{\eta, \omega_0^2, T\}$. The Euler-based inference equations for the stochastic harmonic oscillator read:

- [EM-fwd]:

$$\eta^*_{\text{fwd}} = \frac{1}{\Delta t}\frac{G_s + G'_s - 2C_s + \frac{G'_s}{C''_s}(2G'_s - C''_s - F_s)}{-C_s + \frac{G'^2_s}{C''_s}} ; \tag{270}$$

$$\omega^{2*}_{0\,\text{fwd}} = \frac{1}{\Delta t^2}\frac{(2-\eta\Delta t)G'_s - (1-\eta\Delta t)C''_s - F_s}{C''_s} ; \tag{271}$$

$$T^*_{\text{fwd}} = \frac{1}{2\eta\Delta t^3}[C'_s + (2-\eta\Delta t^2)^2 C_s + (1-\eta\Delta t + \omega_0^2\Delta t^2)^2 C''_s - 2(2-\eta\Delta t)G_s + 2(1-\eta\Delta t + \omega_0^2\Delta t^2)F_s - 2(2-\eta\Delta t)(1-\eta\Delta t+\omega_0^2\Delta t^2)G'_s] . \tag{272}$$

- [EM-bkd]:

$$\eta^*_{\text{bkd}} = \frac{1}{\Delta t}\frac{C''_s + F_s - \frac{G'_s}{C_s}(G_s + G'_s)}{\frac{G_s G'_s}{C_s} - F_s} ; \tag{273}$$

$$\omega^{2*}_{0\,\text{bkd}} = \frac{1}{\Delta t^2}\frac{(2+\eta\Delta t)C_s - G'_s - (1+\eta\Delta t)G_s}{C_s} ; \tag{274}$$

$$T^*_{\text{bkd}} = \frac{1}{2\eta\Delta t^3}[(1+\eta\Delta t)^2 C'_s + (2+\eta\Delta t - \omega_0^2\Delta t^2)^2 C_s + C''_s + 2(1+\eta\Delta t)F_s - 2(1+\eta\Delta t)(2+\eta\Delta t - \omega_0^2\Delta t^2)G_s - 2(2+\eta\Delta t - \omega_0^2\Delta t^2)G'_s] ; \tag{275}$$

- [BBK]:

$$\eta^*_{\text{BBK}} = \frac{2}{\Delta t}\frac{C''_s + F_s - \frac{G'_s}{C_s}(G_s + G'_s)}{C''_s - F_s - \frac{G'_s}{C_s}(G'_s - G_s)}\,; \tag{276}$$

$$\omega_0^{2*}{}_{\text{BBK}} = \frac{1}{\Delta t^2}\frac{2C_s - \left(1 + \frac{\eta\Delta t}{2}\right)G_s - \left(1 - \frac{\eta\Delta t}{2}\right)G'_s}{C_s}\,; \tag{277}$$

$$\begin{aligned} T^*_{\text{BBK}} = &\frac{1}{2\eta\Delta t^3}[\,(1+\eta\Delta t/2)^2\,C'_s + (2-\omega_0^2\Delta t^2)^2 C_s + (1-\eta\Delta t/2)^2\,C''_s - \\ &2(2-\omega_0^2\Delta t^2)\,(1+\eta\Delta t/2)\,G_s + 2\,(1+\eta\Delta t/2)\,(1-\eta\Delta t/2)\,F_s - \\ &2(2-\omega_0^2\Delta t^2)\,(1-\eta\Delta t/2)\,G'_s]\,; \end{aligned} \tag{278}$$

where C, G and F indicate the experimental correlation functions:

$$\begin{aligned} C_s &= \frac{1}{L-1}\sum_{n=1}^{L-1} x_n x_n; \quad C'_s = \frac{1}{L-1}\sum_{n=1}^{L-1} x_{n+1}x_{n+1}; \quad C''_s = \frac{1}{L-1}\sum_{n=1}^{L-1} x_{n-1}x_{n-1}; \\ G_s &= \frac{1}{L-1}\sum_{n=1}^{L-1} x_n x_{n+1}; \quad G'_s = \frac{1}{L-1}\sum_{n=1}^{L-1} x_n x_{n-1}; \qquad F_s = \frac{1}{L-1}\sum_{n=1}^{L-1} x_{n-1}x_{n+1}. \end{aligned} \tag{279}$$

The parameters which enter in the inference formulas for ω_0^2 and T are the optimized values η^* and (η^*, ω_0^{2*}), respectively. Let us remark that the equations above have been derived by solving the saddle point equations $\nabla_{\boldsymbol{\theta}}\mathcal{L} = 0$, since the minimum point of $\mathcal{L}$ is unique when this is derived using the Euler-Maruyama discretization.

Explicit inference formulas for the inertial spin model in the SWA are obtained through an N-body generalization of the equations above. The additional difficulty we must face in the case of linear N-body dynamics is that of estimating the interaction range. Since an explicit analytical minimization of the minus-log-likelihood is not operable, a numerical approach is needed. The core of the inference algorithm for the SWA-ISM is a one-dimensional minimization routine for an effective cost function which, up to additive constants and multiplicative factors, is proportional to the reduced minus-log-likelihood $\mathcal{L}/(L-1)$. Since the coupling matrix is parametrized by a single integer parameter n_c, corresponding to the topological range of interaction, this minimization can be performed by enumeration.

In $d > 1$, out of equilibrium generalizations are also possible: the derivation of the inference equations is analogous, but the results can vary significantly. In particular, the bias is manifested in a specific way for equilibrium processes, while out of equilibrium it appears to be non-universal.

Nonlinear additive processes

Let us consider one-dimensional equilibrium processes with constant diffusion coefficient σ and polynomial force $f(x, \boldsymbol{\mu}) = \sum_\alpha \mu_\alpha x^\alpha$. We restrict in this case to the EM-fwd discretization. Let us define the following quantities, which only depend on the data sample:

$$P_\alpha = \sum_{n=1}^{L-1} x_n^\alpha (x_{n+1} - 2x_n + x_{n-1}); \qquad R_\alpha = \sum_{n=1}^{L-1} x_n^\alpha (x_n - x_{n-1}); \tag{280}$$

$$Q_{\alpha\beta} = \sum_{n=1}^{L-1} x_n^\alpha x_n^\beta; \quad K = \sum_{n=1}^{L-1}(x_n - x_{n-1})(x_{n+1} - 2x_n + x_{n-1}); \quad K_0 = \sum_{n=1}^{L-1}(x_n - x_{n-1})^2. \tag{281}$$

The EM-fwd maximum-likelihood estimators read:

$$\boldsymbol{\mu}^*_{\text{fwd}} = [\mathbf{R}\otimes\mathbf{R} - K_0\mathbf{Q}]^{-1}[K\mathbf{R} - K_0\mathbf{P}]; \tag{282}$$

$$\eta^*_{\text{fwd}} = -\frac{1}{\Delta t}\left(\frac{K}{K_0} - \frac{1}{K_0}\sum_\alpha \Big([\mathbf{R}\otimes\mathbf{R} - K_0\mathbf{Q}]^{-1}[K\mathbf{R} - K_0\mathbf{P}]\Big)_\alpha R_\alpha\right); \tag{283}$$

$$(\sigma^2)^*_{\text{fwd}} = \frac{1}{\Delta t^3}\frac{1}{L-1}\sum_{n=1}^{L-1}[x_{n+1} - 2x_n + x_{n-1} + \eta\Delta t(x_n - x_{n-1}) - \Delta t^2\sum_\alpha \mu_\alpha x_n^\alpha]^2. \tag{284}$$

For the studied case of the Brownian particle in a double well potential, the force term corresponds to $f(x) = -\kappa x - \lambda x^3$, and $\sigma^2 = 2T\eta$. The parameters to estimate are therefore only η, $T = \sigma^2/\eta$, $\kappa = -\mu_1$ and $\lambda = -\mu_3$.

A.2 TOEPLITZ-ML ESTIMATORS

Let us start from the same Kramers process as in Eq. (263) or, equivalently, from the marginalized equation

$$dx(t) = v_0 e^{-\eta(t-t_0)}dt + \left[\int_{t_0}^t e^{-\eta(t-s)}f(x(s))ds\right]dt + \int_{t_0}^t dW(s)\sigma(x(s))e^{-\eta(t-s)}. \tag{285}$$

Toeplitz-ML estimators are obtained adopting a second order Taylor-Itô expansion which is strongly convergent as $\Delta t^{3/2}$. As discussed in Section 3.3.3, this discretization method is sufficiently accurate to reproduce emerging memory and correlation effects due to the projection onto the x subspace. To carry on explicit calculations, let us make a distinction between additive processes and the more complicated multiplicative ones.

Additive processes

The formalism for the additive case has been presented for general additive processes in Section 3.3.3: we rewrite the reduced minus-log-likelihood from Eq.(103) as

$$\frac{\mathcal{L}}{L-1} = \frac{1}{2}\ln\left(2\pi\frac{\sigma^2}{3}\Delta t^3\right) + \frac{1}{2(L-1)}\sum_{k=1}^{L-1}\ln\left(2+\cos\frac{k\pi}{L}\right) + \frac{3}{L(L-1)\sigma^2\Delta t^3}\sum_{n,m=1}^{L-1}\left[(x_{n+1} + F(x_n, x_{n-1};\boldsymbol{\mu}))\tilde{A}_{nm}(x_{m+1} + F(x_m, x_{m-1};\boldsymbol{\mu}))\right], \tag{286}$$

where

$$F(x_n, x_{n-1};\boldsymbol{\mu}) = -\left(1 + e^{-\eta\Delta t}\right)x_n + e^{-\eta\Delta t}x_{n-1} - \frac{1-e^{-\eta\Delta t}}{\eta}\Delta t f(x_n;\boldsymbol{\mu}) \tag{287}$$

and

$$\tilde{A}_{nm} = \sum_{k=1}^{L-1} \frac{\sin\left(\frac{nk\pi}{L}\right)\sin\left(\frac{mk\pi}{L}\right)}{2+\cos\left(\frac{k\pi}{L}\right)}. \tag{288}$$

Given Eq. (286), it is possible to minimize it, numerically or analytically, with respect to the set of parameters $\boldsymbol{\theta} = \{\sigma, \eta, \boldsymbol{\mu}\}$. In the case of a polynomial force $f(x, \boldsymbol{\mu}) = \sum_\alpha \mu_\alpha x^\alpha$, the saddle point equations are analytically solvable. Like for the lower order scheme, it is convenient to redefine auxiliary parameters

$$\beta = e^{-\eta\Delta t}, \quad \lambda_\alpha = \frac{1-e^{-\eta\Delta t}}{\eta}\Delta t \mu_\alpha, \tag{289}$$

and data-dependent quantities:

$$K = \sum_{n,m=1}^{L-1}(x_n - x_{n-1})\tilde{A}_{nm}(x_{m+1} - x_m); \qquad K_0 = \sum_{n,m=1}^{L-1}(x_n - x_{n-1})\tilde{A}_{nm}(x_m - x_{m-1}); \tag{290}$$

$$P_\alpha = \sum_{n,m=1}^{L-1} x_n^\alpha \tilde{A}_{nm}(x_{m+1} - x_m); \qquad R_\alpha = \sum_{n,m=1}^{L-1} x_n^\alpha \tilde{A}_{nm}(x_m - x_{m-1}); \tag{291}$$

$$Q_{\alpha\gamma} = \sum_{n,m=1}^{L-1} x_n^\alpha \tilde{A}_{nm} x_m^\gamma. \tag{292}$$

The Toeplitz-ML estimators for the auxiliary parameters read:

$$\boldsymbol{\lambda}^* = [\mathbf{R}\otimes\mathbf{R} - K_0\mathbf{Q}]^{-1}[K\mathbf{R} - K_0\mathbf{P}]; \tag{293}$$

$$\beta^* = \frac{K}{K_0} - \frac{\boldsymbol{\lambda}\cdot\mathbf{R}}{K_0}; \tag{294}$$

$$(\sigma^2)^* = \frac{1}{\Delta t^3}\frac{1}{L(L-1)}\sum_{n,m=1}^{L-1}\left[(x_{n+1} + F(x_n, x_{n-1}; \boldsymbol{\mu}))\tilde{A}_{nm}(x_{m+1} + F(x_m, x_{m-1}; \boldsymbol{\mu}))\right]. \tag{295}$$

From the solutions (293)–(295), the physical parameters of the process (η, μ_α, and occasionally T, when $\sigma^2 = 2T\eta$) can then be extracted. Explicit formulas for two simple cases, the harmonic oscillator and the Brownian particle in a double well potential, can be found in [74].

Multiplicative processes

The difficulty with multiplicative processes comes from the need to ensure the invertibility of the covariance matrix. We recall that the discretization scheme we adopt is not readily extended to arbitrary processes, since a linear dependence on the velocity variables is required. Hence $\sigma(x; \boldsymbol{\nu})$ must be a function of the x coordinate only, and must be such that the covariance matrix that one can define from it is positive definite.

Let us focus on the example process with $f(x) = -kx$, $\sigma(x) = \sqrt{(a + bx^2)}$. The quantities appearing in the expression of the log-likelihood, Eq.(111), which represents the effective cost function to minimize, are:

$$F(x_n, x_{n-1}; \boldsymbol{\mu}) = x_n - e^{-\eta\Delta t}(x_n - x_{n-1}) + (1 - e^{-\eta\Delta t})\frac{k\Delta t}{\eta}x_n, \tag{296}$$

and

$$C_{nm} = \left(a + bx_n^2\right)\delta_{n,m} + \sqrt{(a + bx_n^2)\,(a + bx_m^2)}\delta_{n,m\pm1}. \tag{297}$$

Simple manipulations allow us to reduce to the minimization problem to a one-dimensional numerical optimization, since analytical formulas for the optimal values of the effective parameters $\beta = e^{-\eta\Delta t}$, $K = k\Delta t/\eta$ and b can be found:

$$\begin{aligned} b^* &= \frac{3}{L-1}\left[P_0 - \frac{P_4^2}{P_2} - \frac{(P_3P_2 - P_4P_5)\,(P_3P_2 - P_4P_5)}{P_1P_2 - P_5^2}\right]; \\ \beta^* &= \frac{P_3^*P_2^* - P_4^*P_5^*}{P_1^*P_2^* - (P_5^*)^2}; \qquad K^* = \frac{\beta^*P_5^* - P_4^*}{(1-\beta^*)P_2^*}\,. \end{aligned} \tag{298}$$

Here $\{P_i\}$ is a collection of functions of a remaining effective parameter, which we rename $\alpha = a/b$. The starred value corresponds to $P_i^* = P_i(\alpha^*)$, with α^* the ML estimator of the new parameter α. This value is obtained by numerically minimizing the following quantity:

$$\mathcal{L}(\alpha) = \frac{1}{L-1}\sum_{k=1}^{L-1}\ln\tilde{\lambda}_k + \ln\left[P_0 - \frac{P_4^2}{P_2} - \frac{(P_3P_2 - P_4P_5)\,(P_3P_2 - P_4P_5)}{P_1P_2 - P_5^2}\right], \tag{299}$$

where $\{\tilde{\lambda}_k\}$ is the set of eigevnvalues of the reduced covariance matrix

$$3C_{nm}/(b\Delta t^3) = {A^{-1}}_{nm}(\alpha) = 2\left(\alpha + x_n^2\right)\delta_{n,m} + \frac{1}{2}\sqrt{(\alpha + x_n^2)\,(\alpha + x_m^2)}\delta_{n,m\pm1}\,. \tag{300}$$

The functions $P_i(\alpha)$ are defiend as follows:

$$\begin{aligned} P_0 &= \frac{1}{L-1}\sum_{n,m=1}^{L-1}(x_{n+1} - x_n)A_{nm}(x_{m+1} - x_m)\,; \\ P_1 &= \frac{1}{L-1}\sum_{n,m=1}^{L-1}(x_n - x_{n-1})A_{nm}(x_m - x_{m-1})\,; \\ P_2 &= \frac{1}{L-1}\sum_{n,m=1}^{L-1}x_nA_{nm}x_m\,; \\ P_3 &= \frac{1}{L-1}\sum_{n,m=1}^{L-1}(x_{n+1} - x_n)A_{nm}(x_m - x_{m-1})\,; \\ P_4 &= \frac{1}{L-1}\sum_{n,m=1}^{L-1}(x_{n+1} - x_n)A_{nm}x_m\,; \\ P_5 &= \frac{1}{L-1}\sum_{n,m=1}^{L-1}(x_n - x_{n-1})A_{nm}x_m\,. \end{aligned}$$

A.3 NON-BAYESIAN ESTIMATORS

We build in this section an alternative approach to the Bayesian one, whose performance was compared to the Toeplitz method in Section 3.3.4. The idea is to exploit discrete update equations in x space to find some relations which link the parameters of the model to accurately measurable quantities. We restrict to the case of the one-dimensional stochastic harmonic oscillator, with $f(x) = -\omega_0^2x$, $\sigma = 2T\eta$.

Let us choose again the update equation corresponding to the usual continuation rule of the LI integration method [173]:

$$x_{n+1} = x_n + e^{-\eta\Delta t}(x_n - x_{n-1}) + \frac{1 - e^{-\eta\Delta t}}{\eta}\omega_0^2\Delta t\, x_n + \zeta_n, \tag{301}$$

and multiply its r.h.s. and l.h.s. by x_n, x_{n+1} and x_{n-1}; then take the average over the noise distribution. The resulting equations are:

$$\langle x_{n+1}x_n\rangle = \langle {x_n}^2\rangle + e^{-\eta\Delta t}(\langle {x_n}^2\rangle - \langle x_n x_{n-1}\rangle) + \frac{1 - e^{-\eta\Delta t}}{\eta}\omega_0^2\Delta t\langle {x_n}^2\rangle + \langle x_n\zeta_n\rangle\ ; \tag{302}$$

$$\langle x_{n+1}x_{n+1}\rangle = \langle x_n x_{n+1}\rangle + e^{-\eta\Delta t}(\langle x_n x_{n+1}\rangle - \langle x_{n-1}x_{n+1}\rangle) + \frac{1 - e^{-\eta\Delta t}}{\eta}\omega_0^2\Delta t\langle x_n x_{n+1}\rangle + \langle\zeta_n x_{n+1}\rangle\ ; \tag{303}$$

$$\langle x_{n+1}x_{n-1}\rangle = \langle x_n x_{n-1}\rangle + e^{-\eta\Delta t}(\langle x_n x_{n-1}\rangle - \langle x_{n-1}^2\rangle) + \frac{1 - e^{-\eta\Delta t}}{\eta}\omega_0^2\Delta t\langle x_n x_{n-1}\rangle\ . \tag{304}$$

Using again Eq. (301) – combined with the covariance matrix of the Gaussian variables – to compute $\langle\zeta_n x_n\rangle$ and $\langle\zeta_n x_{n+1}\rangle$, the relations we find are:

$$G_s = C_s + e^{-\eta\Delta t}(C_s - G_s') + \frac{1 - e^{-\eta\Delta t}}{\eta}\omega_0^2\Delta t\, C_s + b\ ; \tag{305}$$

$$C_s' = G_s + b + a + e^{-\eta\Delta t}(G_s - F_s + b) + \frac{1 - e^{-\eta\Delta t}}{\eta}\omega_0^2\Delta t(G_s + b)\ ; \tag{306}$$

$$F_s = G_s' + e^{-\eta\Delta t}(G_s' - C_s'') + \frac{1 - e^{-\eta\Delta t}}{\eta}\omega_0^2\Delta t\, G_s'\ . \tag{307}$$

In order to find Eqs. (305)–(307), we identified the actual correlation functions with the empirical ones, denoted with C, G and F symbols, and defined as in Eq. (279), and we hypothesized a stationarity assumption to hold. After proper manipulation, one can extract "inference relations" for b, $e^{-\eta\Delta t}$ and $\omega_0^2\Delta t$, and derive from them the physical parameters of the model. Firstly, $e^{-\eta\Delta t}$ is given as the solution of the second-degree polynomial equation:

$$\begin{aligned}&(2G_s' - C_s - C_s'')e^{-2\eta\Delta t} + \left[2G_s + C_s'' - C_s - 2F_s + 5(2G_s' - C_s - C_s'')\right]e^{-\eta\Delta t}\\ &\quad + \left[G_s - G_s' + F_s - C_s' + 5(G_s' - C_s - F_s + G_s)\right] = 0;\end{aligned} \tag{308}$$

then b and $\omega_0^2\Delta t$ are computed as follows:

$$b = G_s' - F_s + G_s - C_s + e^{-\eta\Delta t}\left(2G_s' - C_s'' - C_s\right)\ ; \tag{309}$$

$$\omega_0^2\Delta t = \frac{-\eta}{1 - e^{-\eta\Delta t}}\left[\frac{G_s - C_s - b}{C_s} - e^{-\eta\Delta t}\frac{C_s - G_s'}{C_s}\right]\ . \tag{310}$$

Notice that these inference equations are not unique. Combining the starting equations in a different way would result into slightly different inference formulas, which, however, should provide the same result if the experimental correlation functions faithfully reproduce ensemble averages at the steady state.

This strategy cannot be adapted to interacting oscillators, unless they are fully connected. The obstacle comes indeed from the parametrization of the interaction matrix. Without a priori parametrization, the issue of sufficient statistics arises:

bypassing the technical difficulties related to solving the resulting system of N^2+2 second degree equations for the unknowns b, $e^{-\eta\Delta t}$ and the independent entries of the interaction matrix $\{J_{ij}\}_{i,j=1\dots N}$, the real problem is that we have a much greater number of parameters to infer than of points in each frame. This problem becomes totally intractable if one also allows J_{ij} to evolve in time, as in active animal groups.

Assumptions about the structure of the coupling matrix J_{ij} dramatically diminish the number of parameters, and help us deal with the worry of insufficient statistics, but require an alternative strategy to estimate the interaction range, since closed-form equations are not available. However, it is still possible to approximately estimate the damping coefficient and the effective temperature of the system, assuming that all the particles are immersed in the same uniform thermal bath. Under this assumption, Eqs. (302)–(304) can be adapted to the interacting case and properly manipulated to find the following relations:

$$\begin{aligned} F_s - G'_s - G_s + C_s = e^{-\frac{\eta}{\chi}\Delta t}\left(2G'_s - C''_s - C_s\right) \\ + \frac{G'_{int} - C_{int}}{C_{int}}\left[G_s - C_s - b - e^{-\frac{\eta}{\chi}\Delta t}(C_s - G'_s)\right] - b\,; \end{aligned} \tag{311}$$

$$\begin{aligned} C'_s - 2G_s + C_s = e^{-\frac{\eta}{\chi}\Delta t}\left(G_s - F_s - C_s + G'_s\right) \\ + b\left\{4 + e^{-\frac{\eta}{\chi}\Delta t} + \frac{n_c}{C_{int}}\left[G_s - C_s - b - e^{-\frac{\eta}{\chi}\Delta t}(C_s - G'_s)\right]\right\} \\ + \frac{G_{int} - C_{int}}{C_{int}}\left[G_s - C_s - b - e^{-\frac{\eta}{\chi}\Delta t}(C_s - G'_s)\right]\,; \end{aligned} \tag{312}$$

where we have redefined $J_{ij} = J\Lambda_{ij}$ and exploited the fact that $a = 4b$, with $b = \frac{1}{6}2\frac{T\eta}{\chi^2}\Delta t^3$.

Let us define the empirical spatio-temporal correlation functions involved in these inference formulas:

- Equal-time correlations:

$$C_{ij} = \frac{1}{L-1}\sum_{n=1}^{L-1}\mathbf{x}_i^n\cdot\mathbf{x}_j^n;\quad C'_{ij} = \frac{1}{L-1}\sum_{n=1}^{L-1}\mathbf{x}_i^{n+1}\cdot\mathbf{x}_j^{n+1};\quad C''_{ij} = \frac{1}{L-1}\sum_{n=1}^{L-1}\mathbf{x}_i^{n-1}\cdot\mathbf{x}_j^{n-1}; \tag{313}$$

- One-step correlations:

$$G_{ij} = \frac{1}{L-1}\sum_{n=1}^{L-1}\mathbf{x}_i^{n+1}\cdot\mathbf{x}_j^{n};\quad G'_{ij} = \frac{1}{L-1}\sum_{n=1}^{L-1}\mathbf{x}_i^{n}\cdot\mathbf{x}_j^{n-1}; \tag{314}$$

- Two-step correlations:

$$F_{ij} = \frac{1}{L-1}\sum_{n=1}^{L-1}\mathbf{x}_i^{n+1}\cdot\mathbf{x}_j^{n-1}. \tag{315}$$

The observables appearing in Eqs. (311)–(312) are defined from (313)–(315) as in the following. We can distinguish the contribution of self-correlations, encoded by:

$$C_s = \frac{1}{N}\operatorname{Tr}\mathbf{C}\,;\quad C'_s = \tfrac{1}{N}\operatorname{Tr}\mathbf{C}'\,;\quad C''_s = \frac{1}{N}\operatorname{Tr}\mathbf{C}''\,;$$
$$G_s = \frac{1}{N}\operatorname{Tr}\mathbf{G}\,;\quad G'_s = \tfrac{1}{N}\operatorname{Tr}\mathbf{G}'\,;\quad F_s = \frac{1}{N}\operatorname{Tr}\mathbf{F}\,;$$

and that of correlations between directly interacting birds, encoded by the quantities:

$$C_{int} = \frac{\mathrm{Tr}(\mathbf{\Lambda C})}{N} \; ; \quad G_{int} = \frac{\mathrm{Tr}(\mathbf{\Lambda}^\top \mathbf{G})}{N} \; ; \quad G'_{int} = \frac{\mathrm{Tr}(\mathbf{\Lambda G}')}{N} \; .$$

In the case of the SWA equations of the inertial spin model, $\Lambda_{ij} = n_c \delta_{ij} - n_{ij}$. Notice that all of them are by definition self-averaging quantities, which obviously tend to be more and more stable as the size of the system increases.

In the absence of a proper likelihood, we cannot deal with *int* functions; however, the manipulation we carried out to derive Eqs. (311)–(312) confines them into subleading terms. This can be checked by looking at the combinations:

$$\frac{G_{int} - C_{int}}{C_{int}} \left[\left(1 - e^{-\eta \Delta t} \right) (G_s - C_s) - b \right] \simeq O(\Delta t^5) \, , \tag{316}$$

the one obtained replacing G_{int} with G'_{int}, and

$$b \cdot \frac{n_c}{C_{int}} \left[(1 - e^{-\eta \Delta t})(G_s - C_s) - b \right] \simeq O(\Delta t^6). \tag{317}$$

Under the working hypothesis that Δt is sufficiently small, we can neglect these terms and find usable relations to extract the effective parameters of the thermal bath (η, T) from the experimental self-correlations only. Precisely, η is found as a solution of the equation:

$$\begin{aligned} &(C''_s + C_s - 2G_s)e^{-2\eta\Delta t} + 2(F_s - 5G'_s - G_s + 3C_s + 2C''_s)e^{-\eta\Delta t} \\ &\quad + 4F_s - 4G'_s - 6G_s + 5C_s + C'_s = 0 \, , \end{aligned} \tag{318}$$

whereas the effective temperature is extracted from b, being:

$$b = G'_s + G_s - F_s - C_s + e^{-\eta \Delta t} \left(2G'_s - C_s - C''_s \right) \, . \tag{319}$$

B DECIMATION PROCEDURE FOR GENERAL ARMA PROCESSES

An autoregressive moving-average process of order (p, q), denoted ARMA(p, q), is a time series generated by the update equation

$$X_n = \sum_{i=1}^{p} \phi_i X_{n-i} + \sum_{i=1}^{q} \nu_i \epsilon_{n-i} + \mu \epsilon_n \tag{320}$$

with $\epsilon_n \sim \mathcal{N}(0, 1)$ I.I.D., $\nu_i, \mu \in \mathbb{R}$ [19]. The autoregressive (AR) part of the equation, of order p, can be thought of as the discretization of the deterministic part of an associated continuous-time dynamical model, and corresponds to the first RHS sum of Eq. (320). The moving average (MA) part, of order q, corresponds to the second RHS sum of Eq. (320). For the sake of simplicity, let us rename the whole random increment as $r_n = \sum_{i=1}^{q} \nu_i \epsilon_{n-i} + \mu \epsilon_n$.

In Section 3.4.1 we apply to ARMA(2,1) models a Renormalization Group procedure inspired to real space renormalization in one-dimensional systems. We will generalize here the discussion to arbitrary ARMA(p, q) models, with a special focus on the definition of the coarse-graining transformation and the derivation of the constraints on the fixed points processes imposed by the invariance condition.

Starting from Eq.(320), decimation is performed through the linear combination

$$\mathrm{Eqn}(X_n) + \sum_{i=1}^{p} (-1)^{i+1} \phi_i \mathrm{Eqn}(X_{n-i}). \tag{321}$$

which generalizes (128). Notice that this combination only depends on the AR order p. The resulting discrete-time model reads:

$$\begin{aligned} X_n = & \underbrace{\sum_{i=1}^{p} \left[1 + (-1)^i\right] \phi_i X_{n-i}}_{\text{(A)}} + \underbrace{\sum_{i=1}^{p} (-1)^{i+1} \phi_i \sum_{k=1}^{p} \phi_k X_{n-k-i}}_{\text{(B)}} \\ & + \underbrace{\mu_0 \epsilon_n + \sum_{j=1}^{q} \mu_j \epsilon_{n-j} + \sum_{i=1}^{p} (-1)^{i+1} \phi_i \left[\mu_0 \epsilon_{n-i} + \sum_{k=1}^{q} \mu_k \epsilon_{n-i-k}\right]}_{\tilde{r}_n}. \end{aligned} \tag{322}$$

Sum (B) in Eq. (322) can be rewritten, with a rearrangement of terms, as

$$\text{(B)} = \sum_{i=1}^{p} (-1)^{i+1} \phi_i \sum_{k=1}^{p} \phi_k X_{n-k-i} = \sum_{i=1}^{p} (-1)^{i+1} \left(\phi_i^2 X_{n-2i} + 2 \sum_{k=i}^{\lfloor \frac{p-1}{2} \rfloor} \phi_i \phi_{2k+1} X_{n-2k-2} \right). \tag{323}$$

Eq. (323) shows that, after the decimation, one maintains an autoregressive part of order p. The picture is not modified by sum (A), which only contributes up to an AR order $\lfloor p/2 \rfloor$.

Thus Eq. (322) can be rewritten as:

$$X_n = \sum_{i=1}^{\lfloor p/2 \rfloor} 2\phi_{2i} X_{n-2i} + \sum_{i=1}^{p} (-1)^{i+1} \left(\phi_i^2 X_{n-2i} + 2 \sum_{k=i}^{\lfloor \frac{p-1}{2} \rfloor} \phi_i \phi_{2k+1} X_{n-2k-2} \right) + \tilde{r}_n, \tag{324}$$

where it is possible to recognize a structure of the following kind

$$X_n = \sum_{i=1}^{p} \tilde{\phi}_i X_{n-2i} + \tilde{r}_n. \tag{325}$$

The process corresponds to an ARMA model with the same autoregressive order as the original one (p), but jumps cover time intervals of doubled amplitude.

Further manipulation of the sums (A) and (B) in Eq. (322) allows us to find formal expressions for the AR coefficients of the coarse grained process, $\tilde{\phi}_{i=1\dots p}$:

$$\begin{cases} \tilde{\phi}_1 = 2\phi_2 + \phi_1^2; \\ \tilde{\phi}_i = 2\phi_{2i} + (-1)^{i+1}\phi_i^2 + 2\phi_{2i-1} \sum_{k=1}^{i-1} (-1)^{k+1} \phi_k \quad \text{for} \quad 2 \le i \le \lfloor \frac{p}{2} \rfloor - 1; \\ \tilde{\phi}_i = (-1)^{i+1}\phi_i^2 \qquad\qquad\qquad\qquad\qquad\qquad\quad \text{for} \quad \lfloor \frac{p}{2} \rfloor + 1 \le i \le p. \end{cases} \tag{326}$$

The coarse-grained coefficient $\tilde{\phi}_{\lfloor p/2 \rfloor}$ takes a different form depending on p being even or odd:

$$p \text{ odd}: \qquad \tilde{\phi}_{\lfloor p/2 \rfloor} = 2\phi_{2\lfloor p/2 \rfloor} + (-1)^{1+\lfloor p/2 \rfloor} \phi_{\lfloor p/2 \rfloor}^2; \tag{327}$$

$$p \text{ even}: \qquad \tilde{\phi}_{\lfloor p/2 \rfloor} = 2\phi_{2\lfloor p/2 \rfloor} + (-1)^{1+\lfloor p/2 \rfloor} \phi_{\lfloor p/2 \rfloor}^2 + 2\phi_{2\lfloor p/2 \rfloor - 1} \sum_{k=1}^{\lfloor \frac{p-1}{2} \rfloor} (-1)^{k+1} \phi_k. \tag{328}$$

What remains to determine is whether there is an invariant MA order q associated to p. Let us restart from Eq.(322) and focus on the random term r_n. Since linear combinations of Gaussian variables are still Gaussian, one can properly redefine the ϵ_m's and rearrange the coefficients in front of them to rewrite $\tilde{r}_n = \sum_{i=0}^{\tilde{q}} \tilde{\mu}_i \epsilon_{n-2i}$, where

$$\tilde{q} = \left\lfloor \frac{p+q}{2} \right\rfloor. \tag{329}$$

We deduce there are only 2 invariant scenarios for ARMA(p,q) processes: $q = p$ or $q = p-1$. This fact tells us that partial (discrete) observation of continuous-time processes let memory emerge: each hidden degree of freedom increases by one the order of both the AR part and the MA part of the discrete model, thus introducing color, as expected [200, 97].

Once the general invariant structure under decimation of ARMA(p,q) time series is found, the RG machinery can be deployed, and a careful study of the fixed points can be carried out. In order to complete the RG step and obtain explicit flow equations, it is necessary to detail the rescaling operation. It is convenient to expand any coefficient of the ARMA(p,q) model as a power series of the time step τ (the ARMA parameters are indeed dimensionless functions of τ, whereas physical problems have dimensional parameters). Therefore, we resort to an asymptotic expansion in power series: to work exclusively with integer powers of τ, we replace the set of parameters $\{\phi_{i=1\dots p}, \nu_{i=1\dots q}, \mu\}$ with the equivalent set $\{\phi_{i=1\dots p}, \gamma_{k=1\dots q+1}\}$, where

$$\gamma_k = \mathbb{E}[r_n r_{n+k}]. \tag{330}$$

Any parameter A in the new set can be expressed as $A = \sum_{k=0}^{\infty} A_k \tau^k$. The infinite collection of coefficients A_k constitute the parameter space of our model. Upon rescaling of the time unit, each of them is now rescaled, after coarse-graining, according to its physical dimension: $2^{-k}\tilde{A}_k$, with $\tilde{A}_k$ the coarse-grained parameter. This completes the prescription to derive the RG map for the considered class of linear time series models.

C KINETIC EQUATIONS

In this Appendix we include some details about our tentative use of kinetic theory approaches to reproduce signatures of irreversibility in the two-particle density function. We firstly revise how to predict the mean field transition point by linear stability analysis of a Boltzamnn-like kinetic theory — obtained via the molecular chaos ansatz — and then discuss the shortcomings of this closure. Next, we describe our attempts to find the solution of the ring-kinetic equations associated to the additive ABP model (211).

The starting point for the construction of the kinetic theory in the Smoluchowski approach is given by the equations of motion of the microscopic model, which we report here for the readers' sake:

$$\dot{\mathbf{x}}_i = v_0 \mathbf{e}(\theta_i), \quad \dot{\theta}_i = -J n_{ij}(\{\mathbf{r}_i\}) \sin(\theta_i - \theta_j) + \sqrt{2D}\xi_i, \tag{331}$$

whose associated Fokker-Planck equation reads:

$$\partial_t \psi + v_0 \sum_i \mathbf{e}(\theta_i) \cdot \nabla_i \psi = J \sum_{ij} n_{ij} \partial_{\theta_i} \left[\sin(\theta_i - \theta_j) \psi \right] + D \sum_i \partial^2_{\theta_i \theta_i} \psi. \tag{332}$$

By marginalization of Eq. (332) we obtain the usual BBGKY hierarchy, which we need to truncate with suitable closures. The first two equations of such hierarchy are:

$$\mathcal{D}_t^{(1)} p_1(\Gamma_1) = J(N-1) \int d\Gamma_2 n_{12} \partial_{\theta_1} [\sin(\theta_1 - \theta_2) p_2(\Gamma_1, \Gamma_2)] + D \partial^2_{\theta_1 \theta_1} p_1(\Gamma_1), \tag{333}$$

$$\begin{aligned} \mathcal{D}_t^{(2)} p_2(\Gamma_1, \Gamma_2) = {} & J n_{12} \left\{ \partial_{\theta_1} \left[\sin(\theta_1 - \theta_2) p_2(\Gamma_1, \Gamma_2) \right] + \partial_{\theta_2} \left[\sin(\theta_2 - \theta_1) p_2(\Gamma_1, \Gamma_2) \right] \right\} \\ & + J(N-2) \int d\Gamma_3 \left\{ n_{13} \partial_{\theta_1} [\sin(\theta_1 - \theta_3) p_3(\Gamma_1, \Gamma_2, \Gamma_3)] \right. \\ & \left. + n_{23} \partial_{\theta_2} [\sin(\theta_2 - \theta_3) p_3(\Gamma_1, \Gamma_2, \Gamma_3)] \right\} + D \left(\partial^2_{\theta_1 \theta_1} + \partial^2_{\theta_2 \theta_2} \right) p_2(\Gamma_1, \Gamma_2), \end{aligned} \tag{334}$$

where $\mathcal{D}_t^{(1)} = \partial_t + v_0 \mathbf{e}(\theta_1) \cdot \nabla_1$ and $\mathcal{D}_t^{(2)} = \partial_t + v_0 \mathbf{e}(\theta_1) \cdot \nabla_1 + v_0 \mathbf{e}(\theta_2) \cdot \nabla_2$. We recall that we are focusing here on the metric additive case, where the adjacency matrix is defined as $n_{ij} = \Theta(R - |\mathbf{x}_i - \mathbf{x}_j|)$.

Let us firstly notice that all the equations of the hierarchy can be rescaled by D: this allows us to transform the time into a dimensionless variable through the transformation $t \to Dt$. The material derivatives are therefore redefined from now on with v_0/D instead of v_0.

C.1 MOLECULAR CHAOS

The first closure is the well-known molecular chaos ansatz, according to which two-body correlations can be neglected (at least when the system is sufficiently dilute):

$$p_2(\Gamma_1, \Gamma_2) = p_1(\Gamma_1)p_1(\Gamma_2), \quad \text{i.e.} \quad g_2(\Gamma_1, \Gamma_2) = 0. \tag{335}$$

This corresponds to a mean field approximation and truncates the BBGKY hierarchy at the level of the first equation, yielding:

$$\left[\partial_t + \frac{v_0}{D}\mathbf{e}(\theta_1)\cdot\nabla_1\right] p_1(\Gamma_1) = \frac{J}{D}(N-1)\int d\Gamma_2 n_{12}\partial_{\theta_1}[\sin(\theta_1-\theta_2)p_1(\Gamma_1)]p_1(\Gamma_2) + \partial^2_{\theta_1\theta_1}p_1(\Gamma_1). \tag{336}$$

Expanding in a Fourier series, $p_1(\mathbf{x}, \theta, t) = \sum_k A_k(\mathbf{x}, t)e^{-ik\theta}$, Eq. (336) reads:

$$\begin{aligned}&\partial_t A_k(\mathbf{x}_1) + \frac{v_0}{2D}\left[\nabla^* A_{k+1}(\mathbf{x}_1) + \nabla A_{k-1}(\mathbf{x}_1)\right] \\ &= -k^2 A_k(\mathbf{x}_1) + 2\pi k\frac{J(N-1)}{2D}\int d\mathbf{x}_2 n_{12}[A_{k-1}(\mathbf{x}_1)A_1(\mathbf{x}_2) - A_{k+1}(\mathbf{x}_1)A_{-1}(\mathbf{x}_2)].\end{aligned} \tag{337}$$

The simplest hypothesis

If we further assume independency of angular and spatial variables, $A_k(\mathbf{x}) = \rho(\mathbf{x})A_k$, then the resulting equations for the Fourier coefficients will have no advective term:

$$\partial_t A_k = 2\pi k\frac{J(N-1)}{2DL^2}w\left(A_{k-1}A_1 - A_{k+1}A_{-1}\right), \tag{338}$$

where $w = L^2\int d\mathbf{x}_1 d\mathbf{x}_2\Theta(R - |\mathbf{x}_1 - \mathbf{x}_2|)\rho(\mathbf{x}_1)\rho(\mathbf{x}_2)$. A relevant case is the homogeneous one: when $\rho(\mathbf{x}) = 1/L^2$, $w = \pi R^2$. Let us rename:

$$\nu = \frac{J\rho_0}{2D}w \quad \text{where} \quad \rho_0 = N/L^2; \tag{339}$$

this parameter will play the role of control parameter for the transition to collective motion[1].

A linear stability analysis of the isotropic solution — $A_k = \frac{1}{2\pi}\delta_{k,0}$ — reveals that the linearly unstable modes are only $k = \pm 1$, for $\nu > \nu^* = 1$. To go beyond the linear stability analysis one can integrate the equations numerically. Results of a numerical integration in the case of homogeneous spatial distribution are shown in Fig. 36. We implement a sharp truncation of the equations for the evolution of the modes, assuming that $A_k = 0$ for any $|k| > k_{\max}$; the integration algorithm we use is a simple Euler forward scheme for the resulting finite set of coupled ODEs. Furthermore, we exploit the arbitrariness of the phase offset in order to work with real coefficients. It is indeed always possible to shift the phase offset $\theta \to \theta - \theta_0$ in order to have a real-valued A_1 coefficient: $A_1 = A_{-1}$. Since A_0 is also real, we can combine Eq.(338) (the imaginary part) and the assumptions $A_0, A_1 \in \mathbb{R}$, to deduce that $A_{\pm 2} \in \mathbb{R}$. Repeating the same argument, one can easily see that taking real Fourier coefficients $A_k = A_{-k}$ for all k is not restrictive, in this factorized case.

1 We have in mind that $N \gg 1$, so that $N - 1 \simeq N, N - 2 \simeq N$.

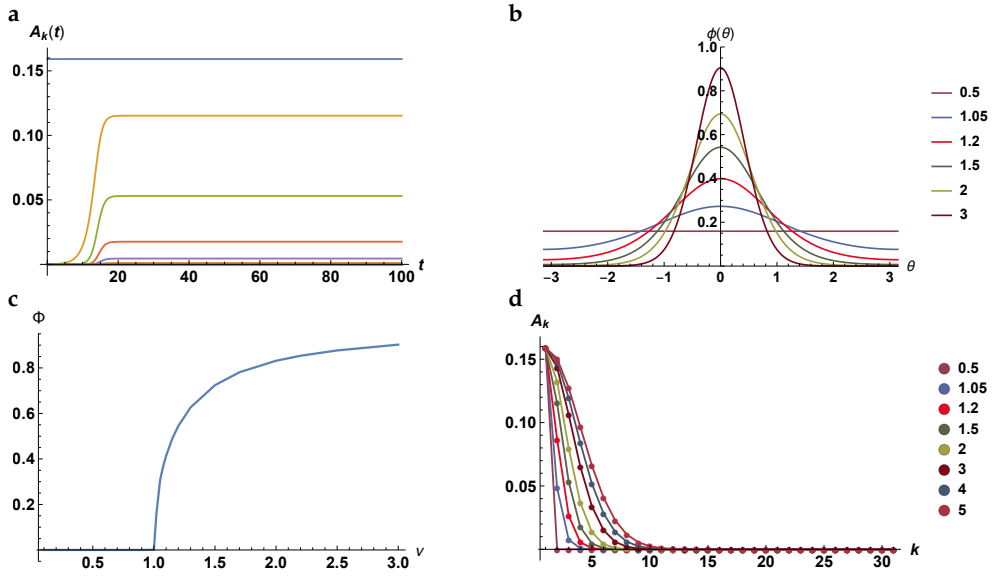

Figure 36: Numerical solution of the kinetic theory, under the molecular chaos closure and homogeneity assumption. **a**: Evolution of the Fourier coefficients $A_k(t)$ for $\nu = 1.5$. The initial condition is a small perturbation around the isotropic solution. **b**: Reconstructed single-particle p.d.f. $\phi(\theta)$ in the stationary state, for different ν values (different colors). **c**: Polarization as a function of the control parameter ν. **d**: Decay of the Fourier coefficients with increasing index k, for different ν values (colors as in Fig. b).

From this simple analysis, it is clear that this approximated description is already good to predict macrosocopic mean field properties of the system, most notably the emergence of a polarized ordered phase for high ν values. In terms of the microscopic model parameters, high ν values may be obtained having either a high interaction strength J, or high density ρ_0, or low noise amplitude D. In this regime, Fourier modes different from the zero mode acquire nonzero values; in particular, we are interested in $A_{\pm 1}$, which is related to the order parameter of the system by the following relation:

$$\langle \Phi \rangle = \int d\mathbf{x} d\theta \, \Re e^{i\theta} p_1(\mathbf{x}, \theta) = \int d\theta \cos\theta \phi(\theta) = \frac{2\pi}{2}(A_1 + A_{-1}) = 2\pi \Re A_1 \qquad (340)$$

The above-mentioned arbitrariness in the choice of the offset can be seen as the arbitrariness in the selection of the collective direction of motion. The phase offset θ_0, with respect to which all the phases θ should be measured in order to have $\Re A_1 = A_1$, indicates the direction of spontaneous symmetry breaking on the plane. We recall that when the symmetry is broken, the generators of the broken symmetry transform one 'ground state'[2] into another one, equivalent to the former. The reconstructed $\phi(\theta)$ is not indicating the global probability distribution of all the possible configurations of the system, corresponding to a mixture of pure states, but the p.d.f. of a single macrostate, where the symmetry breaking is realized in the direction θ_0.

2 The model we are considering is in fact an equilibrium model, since with this closure the effect of self-propulsion (conveyed by the advective term) is removed. It makes thus sense to talk about ground states of the equilibrium Hamiltonian, corresponding to the XY Hamiltonian.

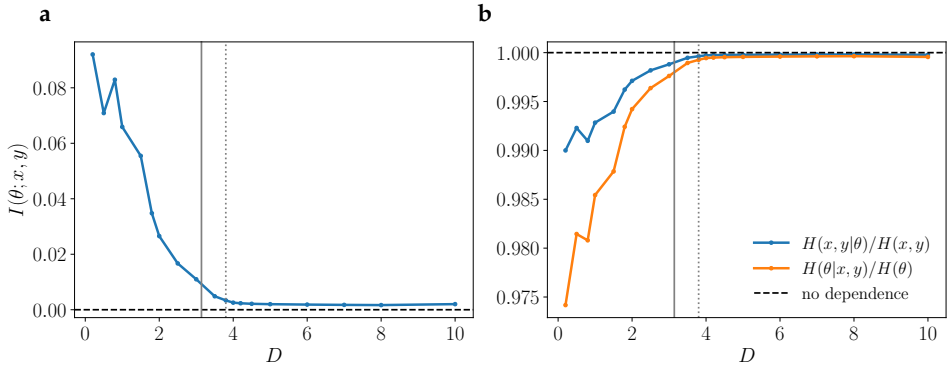

Figure 37: **a**: Mutual information between angular and positional degrees of freedom of the active Brownian particle, as a function of the angular diffusivion parameter D. **b**: Normalized conditional entropies $H(\mathbf{x}|\theta)/H(\mathbf{x})$ and $H(\theta|\mathbf{x})/H(\theta)$. In the disordered phase the curve is basically lying on the horizontal dashed line, corresponding to absence of correlations. In the ordered phase, there is a little mutual dependency between angular and spatial degrees of freedom, but the absolute values are really close to 1. In both plots the solid vertical line indicates the mean field transition point, while the dashed line is the supposed transition point of the finite-size system ($N = 1024$) that we deduced from our numerical simulations.

Beyond equilibrium closures

When the single particle p.d.f. is factorized into a spatial and angular part, we are in fact at equilibrium and there is no way to reproduce the asymmetries we observe in the two-particle correlations. The assumptions we made so far can be violated in two possible ways (or a combination of the two): (i) when $g_2 \neq 0$, i.e. the molecular chaos hypothesis is violated (discussed in C.2); (ii) when $p_1(\mathbf{x}, \theta)$ is not factorized.

Let us assume that condition (ii) is realized, while molecular chaos holds. This is the hypothesis which has been most largely used to derive hydrodynamic equations. Let us notice, however, that hydrodynamic descriptions aim at reproducing the behavior of the system on large spatio-temporal scales by considering only the evolution of slow fields: for polar active matter, they are the density field $\rho(\mathbf{x}) = 2\pi A_0(\mathbf{x})$ and the velocity field $\mathbf{v}(\mathbf{x}) = 2\pi(\Re A_1(\mathbf{x}), \Im A_1(\mathbf{x}))$. On the contrary, we are interested in reconstructing the microscopic pair distribution, on a scale $R \ll L$. What is the best approximation on this scale? Shall we rely on the molecular chaos assumption? In which conditions does hypothesis of factorization of $p_1(\mathbf{x}, \theta)$ (ii) make a difference?

Measures of the mutual information between angular and spatial variables in Fig. 37 hint at the fact that $\mathbf{x}$ and θ are approximately independent in the disordered phase — hence $p_1(\mathbf{x}, \theta)$ factorizes —, while they become slightly dependent in the ordered phase. Precisely, the mutual information is defined as:

$$I_1(\theta; x, y) = D_{KL}(p_1(x, y, \theta)||p_{\mathbf{x}}(x, y) \otimes \phi(\theta)) = H_1(x, y, \theta) - H_1(\theta|x, y) - H_1(x, y|\theta), \tag{341}$$

where $H(\cdot)$ denotes the Shannon entropy and $H(\cdot|\cdot)$ the conditional entropy. The subscript 1 refers to the single body p.d.f.. We estimate these quantities employing

the p.d.f.s reconstructed from numerical simulations of the ABP model, i.e. histograms which are therefore defined on a discrete state space. For example, the reconstructed one-particle distribution is estimated through

$$\hat{p}_1(x,y,\theta) = \frac{1}{N}\sum_{i=1}^{N}\langle \chi_{\mathcal{B}_x}(x_i(t)-x)\chi_{\mathcal{B}_y}(y_i(t)-y)\chi_{\mathcal{B}_\theta}(\theta_i(t)-\theta)\rangle_t \tag{342}$$

where $\langle\cdot\rangle_t$ denotes the empirical time average in the alleged stationary regime and $\chi_{\mathcal{B}_x}$ is the characteristic function of the bin for the x variable, centered in 0:

$$\chi_{\mathcal{B}_x} = \begin{cases} 1 & \text{if} \quad x \in [-\mathcal{B}_x/2, \mathcal{B}_x/2) \\ 0 & \text{if} \quad x \notin [-\mathcal{B}_x/2, \mathcal{B}_x/2)\,. \end{cases} \tag{343}$$

Marginalized distributions are computed in analogous ways. To realize Fig. 37 we chose non-overlapping bins of amplitude $\mathcal{B}_x = L/n^x_{bins}$, $\mathcal{B}_y = L/n^y_{bins}$, $\mathcal{B}_\theta = 2\pi/n^\theta_{bins}$, with $n^x_{bins} = n^y_{bins} = n^\theta_{bins} = 100$.

We plot in Fig. 37 both $\hat{I}_1(\theta;x,y)$ and $\hat{H}_1(\theta|x,y)/\hat{H}_1(x,y)$: this latter quantity takes values between 0 and 1 since the conditional entropy is normalized by the maximum value it can take. From this second plot we deduce that the degree of mutual dependence of the angular and spatial coordinates θ and $\mathbf{x} = (x,y)$ is pretty low, suggesting that the independence hypothesis is a quite reasonable one. Nonetheless, there is a little but visible transition exactly at the onset of polar order. Probably the two options presented above might work equally well in general, but, depending on the goals and the working regime, one can have advantages or disadvantages over the other.

As regards the molecular chaos assumption, $g_2 = 0$, two remarks are in order. First of all, we expect the correlation g_2 to be small, compared to the factorized part, especially far from the transition point (the region across the transition point is however the one of interest for us, where non-equilibrium signatures are visible). Therefore, if we are interested in p_2 and not in g_2, it is reasonable to neglect the pair correlation. At the same time, we remark that the role of g_2 is not limited to the prediction of local features of no impact on the global behavior of the system. On the contrary, the Fourier coefficients of g_2 are linked to the fluctuations of the order parameter [53], which we have shown — rescaled by D — in Fig.22. If we rewrite g_2 as:

$$g_2(\mathbf{x}_1,\theta_1,\mathbf{x}_2,\theta_2) = \sum_{kl} G_{k,l}(\mathbf{x}_1,\mathbf{x}_2)e^{-ik\theta_1 - il\theta_2}, \tag{344}$$

then the fluctuations of the order parameter correspond to:

$$\langle\Phi^2\rangle - \langle\Phi\rangle^2 = (2\pi)^2 \int d\mathbf{x}_1 d\mathbf{x}_2 \frac{G_{1,-1}(\mathbf{x}_1,\mathbf{x}_2) + G_{-1,1}(\mathbf{x}_1,\mathbf{x}_2)}{2}. \tag{345}$$

The fact that the plots in Figs. 22.h and 22.i qualitatively resemble each other is an indication that $G_{\pm1,\mp1}$ may play an important role in reproducing the non-equilibrium signatures of the considered model. For this reason, we choose to include pair correlations and consider the ring-kinetic equations.

C.2 RING KINETIC THEORY

Let us now assume $g_2 \neq 0$ but $g_3 = 0$: the ring kinetic closure moves the truncation of the hierarchy one step beyond that of molecular chaos. This operation brings in significantly more complicated collision integrals: a good strategy to handle the resulting equations is to assume again homogeneity of the single particle distribution:

$$p_1(\mathbf{x}, \theta) = \frac{1}{L^2}\phi(\theta) \tag{346}$$

and to define a reduced two-particle correlation

$$\tilde{g}_2(\theta_1, \theta_2, \mathbf{\Delta}) = L^2 \int d\mathbf{x}_1 d\mathbf{x}_2 g_2(\Gamma_1, \Gamma_2)\delta_L(\mathbf{x}_2 - \mathbf{x}_1 - \mathbf{\Delta}). \tag{347}$$

The evolution equations for these two functions, derived from Eqs. (333)–(334), read:

$$\begin{aligned}\partial_t \phi(\theta_1) = \frac{2\nu}{\pi R^2} \int d\mathbf{\Delta}\, n_{12}(\Delta)\partial_{\theta_1}\left[\sin(\theta_1 - \theta_2)\tilde{g}_2(\theta_1, \theta_2, \mathbf{\Delta})\right] \\ + 2\nu \int d\theta_2\, \partial_{\theta_1}\left[\sin(\theta_1 - \theta_2)\phi(\theta_1)\right]\phi(\theta_2) + \partial^2_{\theta_1\theta_1}\phi(\theta_1);\end{aligned} \tag{348}$$

$$\begin{aligned}\mathcal{D}_t^{(1-2)}\tilde{g}_2(\theta_1, \theta_2, \mathbf{\Delta}) = \frac{J}{D} n_{12}(\Delta)\left(\partial_{\theta_1} - \partial_{\theta_2}\right)\left\{\sin(\theta_1 - \theta_2)\left[\tilde{g}_2(\theta_1, \theta_2, \mathbf{\Delta}) + \phi(\theta_1)\phi(\theta_2)\right]\right\} \\ + 2\nu \int d\theta_3\, \phi(\theta_3)\left\{\partial_{\theta_1}\left[\sin(\theta_1 - \theta_3)\tilde{g}_2(\theta_1, \theta_2, \mathbf{\Delta})\right]\right. \\ \left. + \partial_{\theta_2}\left[\sin(\theta_2 - \theta_3)\tilde{g}_2(\theta_1, \theta_2, \mathbf{\Delta})\right]\right\} \\ + \frac{2\nu}{\pi R^2} \int d\theta_3 d\mathbf{h}\ \left\{n_{13}(h)\partial_{\theta_1}\left[\sin(\theta_1 - \theta_3)\phi(\theta_1)\right]\tilde{g}_2(\theta_2, \theta_3, -\mathbf{\Delta} + \mathbf{h})\right. \\ \left. + n_{23}(h)\partial_{\theta_2}\left[\sin(\theta_2 - \theta_3)\phi(\theta_2)\right]\tilde{g}_2(\theta_1, \theta_3, \mathbf{\Delta} + \mathbf{h})\right\} \\ + \left(\partial^2_{\theta_1\theta_1} + \partial^2_{\theta_2\theta_2}\right)\tilde{g}_2(\theta_1, \theta_2, \mathbf{\Delta}).\end{aligned} \tag{349}$$

To obtain these equations we rescaled the time unit, $t \to Dt$, and defined the dimensionless control parameter

$$\nu = \frac{J\rho_0}{2D}\pi R^2 \simeq \frac{J(N-1)}{2D}\frac{\pi R^2}{L^2}. \tag{350}$$

The latter approximated equivalence is valid in the thermodynamic regime we are interested in ($N \gg 1$). Finally, the material derivative of $\tilde{g}_2$ is defined as

$$\mathcal{D}_t^{(1-2)} = \partial_t - \frac{v_0}{D}\left[(\cos\theta_1 - \cos\theta_2)\,\partial_{\Delta_x} + (\sin\theta_1 - \sin\theta_2)\,\partial_{\Delta_y}\right]. \tag{351}$$

Eqs. (348)–(349) also have a diagrammatic expression – see Ref. [113].

To make any progress we need to expand in Fourier series:

$$\phi(\theta) = \sum_k A_k e^{-ik\theta}; \qquad \tilde{g}_2(\theta_1, \theta_2, \mathbf{\Delta}) = \sum_{kl} F_{k,l}(\mathbf{\Delta})e^{-ik\theta_1 - il\theta_2}. \tag{352}$$

The dynamic equations for the Fourier modes read:

$$(\partial_t + k^2)A_k = \nu 2\pi k(A_{k-1}A_1 - A_{k+1}A_{-1}) + \nu 2\pi k \int d\mathbf{\Delta} n(\Delta)\left[F_{k-1,1}(\mathbf{\Delta}) - F_{k+1,-1}(\mathbf{\Delta})\right]; \tag{353}$$

$$\begin{aligned}
&\partial_t F_{k,l}(\boldsymbol{\Delta}) - \frac{v_0}{D}\left[\nabla^*\left(F_{k+1,l} - F_{k,l+1}\right) + \nabla\left(F_{k-1,l} - F_{k,l-1}\right)\right] \\
&= \frac{J}{2D} n(\Delta)(k-l)\left[F_{k-1,l+1}(\boldsymbol{\Delta}) - F_{k+1,l-1}(\boldsymbol{\Delta}) + A_{k-1}A_{l+1} - A_{k+1}A_{l-1}\right] \\
&\quad + \nu 2\pi k\left[F_{k-1,l}(\boldsymbol{\Delta})A_1 - F_{k+1,l}(\boldsymbol{\Delta})A_{-1}\right] + \nu 2\pi l\left[F_{k,l-1}(\boldsymbol{\Delta})A_1 - F_{k,l+1}(\boldsymbol{\Delta})A_{-1}\right] \\
&\quad + \frac{2\nu}{R^2}\int d\mathbf{h} n(h)\left\{k\left[A_{k-1}F_{1,l}(\boldsymbol{\Delta}-\mathbf{h}) - A_{k+1}F_{-1,l}(\boldsymbol{\Delta}-\mathbf{h})\right]\right. \\
&\qquad \left. + l\left[A_{l-1}F_{k,1}(\boldsymbol{\Delta}+\mathbf{h}) - A_{l+1}F_{k,-1}(\boldsymbol{\Delta}+\mathbf{h})\right]\right\} - (k^2+l^2)F_{k,l}(\boldsymbol{\Delta}),
\end{aligned} \tag{354}$$

where $\nabla = \partial_{\Delta_x} + i\partial_{\Delta_y}$, $\nabla^* = \partial_{\Delta_x} - i\partial_{\Delta_y}$. Although simpler, these equations are still integro-differential: in order to remove the spatial integrals we can either approximate the collision kernel in such a way that it yields punctual local collisions, $n(\Delta) \sim \delta(\Delta)$, or exploit the PBC of the domain, in the same way as for the angular variables (cfr. [113]).

Following this second approach, we define new tensor-like Fourier coefficients $F_{k,l,m,n}$ s.t.

$$F_{k,l}(\boldsymbol{\Delta}) = \sum_{m,n} F_{k,l,m,n} e^{-im\frac{2\pi}{L}\Delta_x - in\frac{2\pi}{L}\Delta_y}. \tag{355}$$

Let us define:

$$K_{mn} = \frac{1}{\pi R^2}\int_0^R dr \int_0^{2\pi} d\alpha\, r e^{ir\frac{2\pi}{L}(m\cos\alpha + n\sin\alpha)} = {}_0\tilde{F}_1\left(2; -(m^2+n^2)\frac{\pi^2 R^2}{L^2}\right), \tag{356}$$

where ${}_0\tilde{F}_1(\cdot;\cdot)$ denotes the regularized confluent hypergeometric function, ${}_0\tilde{F}_1(a;z) = \frac{{}_0F_1(a;z)}{\Gamma(a)}$, which can be also rewritten as follows:

$$\begin{aligned}
K_{mn} &= {}_0\tilde{F}_1\left(2; -(m^2+n^2)\frac{\pi^2R^2}{L^2}\right) = \frac{L}{\pi R}\frac{J_1\left(2\frac{\pi R}{L}\sqrt{m^2+n^2}\right)}{\sqrt{m^2+n^2}} \\
&= J_0\left(\frac{2\pi R}{L}\sqrt{m^2+n^2}\right) + J_2\left(\frac{2\pi R}{L}\sqrt{m^2+n^2}\right),
\end{aligned} \tag{357}$$

being $J_n(z)$ the Bessel function of the first kind. Since K_{mn} only depends to (m,n) through $\sqrt{m^2+n^2}$, we have the useful properties:

$$K_{m,n} = K_{-m,n} = K_{m,-n} = K_{-m,-n}. \tag{358}$$

Exploiting the series representation of the Heaviside function

$$n(\Delta) = H(R - |\boldsymbol{\Delta}|) = \sum_{m,n}\frac{\pi R^2}{L^2}K_{m,n}e^{-im\frac{2\pi}{L}\Delta_x - in\frac{2\pi}{L}\Delta_y}, \tag{359}$$

we obtain from Eqs. (353)–(354):

$$(\partial_t + k^2)A_k = \nu 2\pi k(A_{k-1}A_1 - A_{k+1}A_{-1}) + \nu 2\pi k\sum_{mn} K_{mn}\left[F_{k-1,1,m,n} - F_{k+1,-1,m,n}\right]; \tag{360}$$

$$\begin{aligned}
&\partial_t F_{klmn} + \frac{v_0}{2D}\frac{2\pi}{L}\left[(im+n)\left(F_{k+1,l,m,n} - F_{k,l+1,m,n}\right) + (im-n)\left(F_{k-1,l,m,n} - F_{k,l-1,m,n}\right)\right] \\
&= \frac{\nu}{N}(k-l)\sum_{s,t} K_{s,t}\left(F_{k-1,l+1,m-s,n-t} - F_{k+1,l-1,m-s,n-t}\right) \\
&\quad + \nu 2\pi k\left(F_{k-1,l,m,n}A_1 - F_{k+1,l,m,n}A_{-1}\right) + \nu 2\pi l\left(F_{k,l-1,m,n}A_1 - F_{k,l+1,m,n}A_{-1}\right) \\
&\quad + \frac{\nu}{N}(k-l)K_{mn}\left(A_{k-1}A_{l+1} - A_{k+1}A_{l-1}\right) + \nu K_{m,n}\left[2\pi k\left(A_{k-1}F_{1,l,m,n} - A_{k+1}F_{-1,l,m,n}\right)\right. \\
&\qquad \left. + 2\pi l\left(A_{l-1}F_{k,1,-m,-n} - A_{l+1}F_{k,-1,-m,-n}\right)\right] - (k^2+l^2)F_{klmn}.
\end{aligned} \tag{361}$$

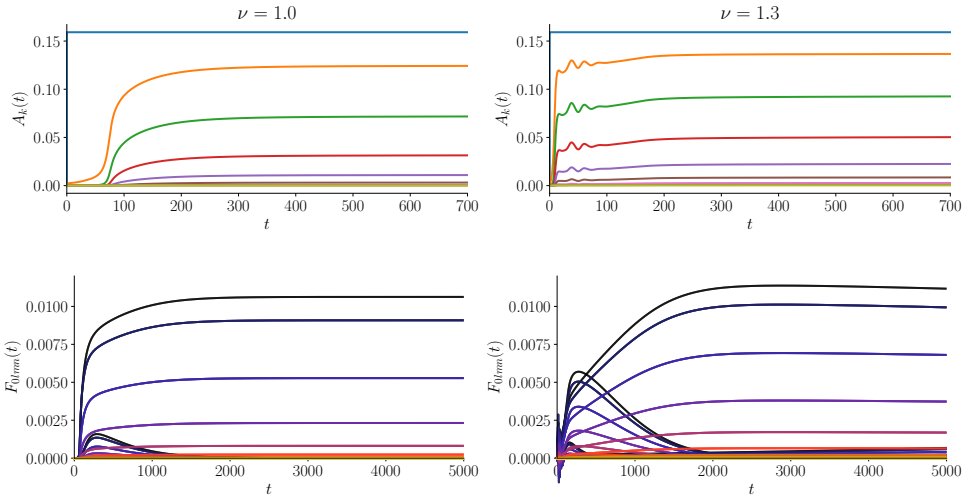

Figure 38: Numerical integration of the ring-kinetic equations (360)–(361), for different values of ν in the ordered phase. In the first row, we plot the coefficients $A_k(t)$ of te single particle distribution; in the second row we plota subset of coefficients for the two-particle p.d.f. $\tilde{g}_2$, $F_{0lmn}(t)$. The color code in the second row is based on the value of $|l|$, ranging from o (black) to 8 (yellow). In all the numerical integrations of the ring-kinetic equations, we chose $n_{\max} = m_{\max} = 10$; for $\nu \leq 0.9$, $k_{\max} = 5$, whereas for $\nu > 0.9$, $k_{\max} = 8$. T time intervals of the plots in the first and second row are very different, in order to highlight the short-time dynamics of A_k. The coefficient of A_k are approximately stationary from $t = 500$ to $t = 5000$.

The constraints imposed by reality and normalization conditions, in addition to the invariance under particles' label permutations, are:

- $A_k^* = A_{-k}$, $F^*_{k,l,m,n} = F_{-k,-l,-m,-n}$;
- $A_0 = \frac{1}{2\pi}$, $F_{0000} = 0$;
- $F_{k,l,m,n} = F_{l,k,-m,-n}$.

Eqs.(360)–(361) are compatible with these conditions. By definition, g_2 must also satisfy:

$$\int d\Gamma_2\, g_2(\Gamma_1, \Gamma_2) = 0 \implies \int d\theta_2 d\boldsymbol{\Delta}\, \tilde{g}_2(\theta_1, \theta_2, \boldsymbol{\Delta}) = \int d\theta_1 d\boldsymbol{\Delta}\, \tilde{g}_2(\theta_1, \theta_2, \boldsymbol{\Delta}) = 0, \quad (362)$$

hence

- $F_{k,0,0,0} = F_{0,k,0,0} = 0\ \forall k$.

If $F_{k,l,m,n}$ evolves according to Eq. (361), this requires:

$$\frac{\nu}{N} k \sum_{s,t} K_{s,t} \left(F_{k-1,1,-s,-t} - F_{k+1,-1,-s,-t} \right) = -\frac{\nu}{N} k \left(A_{k-1} A_1 - A_{k+1} A_{-1} \right). \quad (363)$$

Assuming $\frac{\nu}{N} k \neq 0$ and plugging the resulting equation into Eq.(360), one obtains a simple diffusion equation for $\phi(\theta)$: $(\partial_t + k^2) A_k = 0$ which does not allow for non-trivial steady state solutions. However, let us notice that Eqs. (360) and (361) have been obtained in the $N \to \infty$ limit; terms of order $1/N$ have already been neglected, assuming that $N - 1 \simeq N - 2 \simeq N$. To be consistent, we should work in the $N \to \infty$

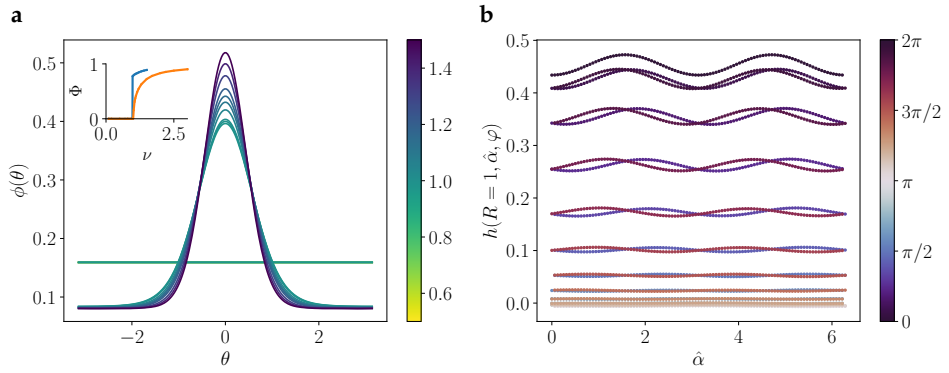

Figure 39: **a**: Single particle distribution $\phi(\theta)$ reconstructed from the stationary Fourier coefficients A_k, for varying ν values (specified in the colorbar). The inset shows the order parameter curve (corresponding to $2\pi\Re A_1 = \int d\theta\phi(\theta)\cos(\theta)$, comparing the results of the ring kinetic theory and of the mean field theory. **b**: Reconstructed function $h(\hat{\alpha}, \varphi)$ at $\nu = 1.1$ (to be compared with Fig. 31 from [113]): we show for selected values of φ the dependency on $\hat{\alpha}$ (as defined in Fig. 20). The obtained result is unsatisfactory under several respects. Firstly, the asymmetry we look for is not reproduced; on the contrary, if we compare the curves for $\varphi = \pm\varphi/4$ (the pair at height ~ 0.25), converging configurations seem to be more probable than divergig ones, in clear contrast with thermodynamic constraints. Moreover, the curves do not have the right periodicity in $\hat{\alpha}$ (expected periodicity of 2π). However, a possible explanation for this result may be that since $h(\hat{\alpha}, \varphi)$ is the only function which contains information about the spatial structure, it primarily tends to reproduce the fact that a particle which travels in a fixed direction has a greater probability of finding another close particle (typically well aligned, at $R = 1$) on one of its sides ($\hat{\alpha} = \pm\pi/2$), rather than in front or at the back ($\hat{\alpha} = 0$ or $\hat{\alpha} = \pi$).

limit, where Eq. (362) is immediately satisfied: let us then remove the corresponding terms form Eq. (361).

Eq. (360) and the thermodynamic limit of Eq. (361) define an infinite set of coupled ODEs, which have a linear dependency on $\mathbf{F}$ but nonlinear on $\mathbf{A}$. We are in principle interested in the stable equilibrium points of such equations. In order to study them, we have integrated numerically a truncated set of equations (separating real and imaginary parts[3]), obtained imposing that $A_k = 0$ and $F_{k,l,m,n} = 0$ for all $|k| > k_{max}$, $|l| > k_{max}$, $|m| > m_{max}$ or $|n| > n_{max}$. In our numerical attempts we an initial condition describing a little perturbation around the homogeneous isotropic solution: $A_k = \frac{1}{2\pi}\delta_{k0} + \alpha(1 - \delta_{k0})$, with $|\alpha| \ll \frac{1}{2\pi}$ ($\alpha \in \mathbb{C}$), and $F_{k,l,m,n} = \epsilon \ll \frac{1}{(2\pi)^2}$ ($\epsilon \in \mathbb{R}$) — except in the cases $k = l = m = n = 0$, or $k = m = n = 0$, or $l = m = n = 0$, where $F_{k,l,m,n} = 0$.

From numerical integration of Eqs. (360)–(361), we observe that, whenever $\nu < \nu^* \simeq 0.98$, the homogeneous isotropic solution $(A_k, F_{k,l,m,n}) = (\frac{1}{2\pi}\delta_{k,0}, 0)$ is asymptotically stable, while it becomes unstable for $\nu > \nu^*$. It is worth noticing that our initial conditions are chosen in such a way that $F_{k,l,m,n} \ll A_k$. Recalling

3 We could not find an easy way to remove the degeneracy associated to the SSB in this second case. Since complex $F_{k,l,m,n}$ coefficients appear in the equation for A_k, requiring that $A_{\pm 1}$ is real does not guarantee that $A_k \in \mathbb{R}$ for all the subsequent k values.

that $\nu^* = 1$ is the value at which the instability of A_1 emerges in the absence of correlations, the picture seems consistent with an instability led by the coefficients of the single-particle p.d.f.: as A_1 becomes significantly different from zero, the terms $\nu 2\pi k \left(F_{k-1,l,m,n}A_1 - F_{k+1,l,m,n}A_{-1}\right) + \nu 2\pi l \left(F_{k,l-1,m,n}A_1 - F_{k,l+1,m,n}A_{-1}\right)$ in Eq.(361) are no more negligible (as they are at the beginning) and boost the growth of $\mathbf{F}$ coefficients. In turn, as the coefficients of the two-particle p.d.f. evolve, the Fourier modes A_k grow and reach a new stable stationary point. We cannot exclude that other routes to instability may be covered with different initial conditions; a direct inspection of the linearized equations is not easy to carry on for such a high-dimensional tensor system. The difference between the stationary state reached in the molecular chaos approximation and in the ring-kinetic theory is visible in Fig. 30 the main text. We show in Fig. 38 the temporal evolution of A_k and F_{klmn} for two selected values of ν.

When $\nu > 1.5$ the numerical solution diverges. This fact is not worrying per se, because it can simply be an indication that, when the system is deeply ordered, higher order correlations described by the third and following equations of the BBGKY hierarchy are non-negligible. Moreover, the homogeneity assumption clearly breaks down in this regime, as it is visible in Fig. 22, so we expect the theory to become inappropriate. Our hope is to describe with the ring-kinetic theory the asymmetries in the pair correlation function which emerge close to the onset of order. A more problematic issue comes then from the fact that the function $q(\alpha, \varphi)$ that we can build using the asymptotic values of the $\mathbf{F}$ and $\mathbf{A}$ coefficients does not match the one we reconstructed from numerical simulations, for ν close to 1. The resulting plot is shown in Fig. 39. However, a ring kinetic theory developed by Kursten and Ihle in [113] for the non-additive model leads to results that are in qualitative agreement with our findings, as discussed in the main text. What changes in this theory is that the PDE is nonlinear in $\mathbf{F}$ because of the presence of $\mathbf{F}$-dependent weights in the collision integrals.

Numerical reconstruction of the reduced two-particle density

Once the stable stationary solutions of the set of ODEs for the truncated series of Fourier coefficients is found, we use them to reconstruct the quantity of interest for us, i.e. the reduced two-particle density $q(\alpha, \varphi)$. The tricky function to reconstruct is the non-symmetrized version of the connected pair distribution $h(r, \hat{\alpha}, \varphi)$, which contains the entire spatial dependence. Knowing this function, we can compute $q(\hat{\alpha}, \varphi) = h(R, \hat{\alpha}, \varphi) + \int d\theta_1 \phi(\theta_1)\phi(\theta_1 + \varphi)$, and the desired symmetrized function $q(\alpha, \varphi)$ is eventually obtained by using the definition of α in Eq. (224).

Exploiting the definition of h and the Fourier series expansion of $\tilde{g}_2$, $h(r, \hat{\alpha}, \varphi)$ reads:

$$\begin{aligned} h(r, \hat{\alpha}, \varphi) &= \int d\theta_1 d\theta_2 d\mathbf{\Delta}\, \tilde{g}_2(\theta_1, \theta_2, \mathbf{\Delta}) \delta(\theta_2 - \theta_1 - \varphi) \delta(\Delta_x - r\cos(\hat{\alpha} + \theta_1)) \delta(\Delta_y - r\cos(\hat{\alpha} + \theta_1)) \\ &= \sum_{klmn} 2\pi F_{klmn} e^{-il\varphi + i(k+l)\hat{\alpha}} \left(\frac{-n - im}{\sqrt{m^2 + n^2}}\right)^{k+l} J_{k+l}\left(\frac{2\pi r}{L}\sqrt{m^2 + n^2}\right), \end{aligned} \tag{364}$$

where $J_n(x)$ is the Bessel function of the first kind:

$$J_n(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\tau\, e^{i(x\sin\tau - n\tau)}. \tag{365}$$

Let us employ the symmetries of the Fourier coefficients and of the Bessel functions to rewrite the sum in Eq. (364) as a sum which only involves real numbers.

Thanks to the invariance of the two-particle density under particle relabeling, $F_{k,l,m,n} = F_{l,k,-m,-n}$, so:

$$\begin{aligned} h(r,\hat{\alpha},\varphi) = \sum_{klmn} 2\pi F_{k,l,m,n} \left(\frac{-n-im}{\sqrt{m^2+n^2}}\right)^{k+l} J_{k+l}\left(\frac{2\pi r}{L}\sqrt{m^2+n^2}\right) e^{i(k+l)(\hat{\alpha}-\varphi/2)} \\ \cdot \frac{1}{2}\left[e^{i(k-l)\varphi/2} + (-1)^{k+l} e^{-i(k-l)\varphi/2}\right]. \end{aligned} \tag{366}$$

Notice that $k+l$ and $k-l$ are either both odd or both even, so we can split the sum as follows (depending on the parity of the indexes $k+l$):

$$\begin{aligned} &\frac{1}{2\pi}h(r,\hat{\alpha},\varphi) \\ &= \sum_{k+l \text{ even}} \sum_{mn} F_{k,l,m,n} \left(\frac{-n-im}{\sqrt{m^2+n^2}}\right)^{k+l} J_{k+l}\left(\frac{2\pi r}{L}\sqrt{m^2+n^2}\right) e^{i(k+l)(\hat{\alpha}-\varphi/2)} \cos\left[(k-l)\frac{\varphi}{2}\right] \\ &+ \sum_{k+l \text{ odd}} \sum_{mn} F_{k,l,m,n} \left(\frac{-n-im}{\sqrt{m^2+n^2}}\right)^{k+l} J_{k+l}\left(\frac{2\pi r}{L}\sqrt{m^2+n^2}\right) e^{i(k+l)(\hat{\alpha}-\varphi/2)} i\sin\left[(k-l)\frac{\varphi}{2}\right]. \end{aligned} \tag{367}$$

We now exploit the following property of the Bessel functions of the first kind of integer order:

$$J_{2\nu}(x) = J_{-2\nu}(x), \quad x\in\mathbb{R},\ \nu\in\mathbb{Z}; \qquad J_{2\nu+1}(x) = -J_{-(2\nu+1)}(x), \quad x\in\mathbb{R},\ \nu\in\mathbb{Z}. \tag{368}$$

Let us indicate explicitly the real and imaginary parts of the Fourier coefficients:

$$F_{k,l,m,n} = f_{k,l,m,n} + i g_{k,l,m,n}, \tag{369}$$

so that each term in the sums above can be explicitly rewritten as

$$\begin{aligned} k+l \text{ even}: \quad & (f_{k,l,m,n} + i \underbrace{g_{k,l,m,n}}_{\text{changes sign}}) \left(\cos[(k+l)\tau_{nm}] + i\underbrace{\sin[(k+l)\tau_{nm}]}_{\text{changes sign}}\right) \\ & \cdot J_{k+l}\left(\frac{2\pi r}{L}\sqrt{m^2+n^2}\right)\left\{\cos\left[(k+l)\left(\hat{\alpha}-\frac{\varphi}{2}\right)\right]\right. \\ & \left.+ i\underbrace{\sin\left[(k+l)\left(\hat{\alpha}-\frac{\varphi}{2}\right)\right]}_{\text{changes sign}}\right\}\cos\left[(k-l)\frac{\varphi}{2}\right] \end{aligned} \tag{370}$$

or

$$\begin{aligned} k+l \text{ odd}: \quad & (f_{k,l,m,n} + i \underbrace{g_{k,l,m,n}}_{\text{changes sign}}) \left(\underbrace{\cos[(k+l)\tau_{nm}]}_{\text{changes sign}} + i\sin[(k+l)\tau_{nm}]\right) \\ & \cdot \underbrace{J_{k+l}\left(\frac{2\pi r}{L}\sqrt{m^2+n^2}\right)}_{\text{changes sign}}\left\{\cos\left[(k+l)\left(\hat{\alpha}-\frac{\varphi}{2}\right)\right]\right. \\ & \left.+ i\underbrace{\sin\left[(k+l)\left(\hat{\alpha}-\frac{\varphi}{2}\right)\right]}_{\text{changes sign}}\right\}\underbrace{i\sin\left[(k-l)\frac{\varphi}{2}\right]}_{\text{changes sign}}. \end{aligned} \tag{371}$$

We marked down how the sign of the various pieces change under the conjugate transformation $(k,l,m,n) \to (-k,-l,-m,-n)$. The angle τ_{nm} is defined by

$$\cos(\tau_{nm}) = \frac{-n}{\sqrt{m^2+n^2}}; \qquad \sin(\tau_{nm}) = \frac{-m}{\sqrt{m^2+n^2}}. \tag{372}$$

Under the conjugate transformation it transforms as $\tau_{-n,-m} = \tau_{nm} + \pi$.

Only products involving an even number of changes of sign do survive in the sum, so we have:

$$\begin{aligned} h(r,\hat{\alpha},\varphi) = 2\pi & \sum_{\substack{k,l,m,n: \\ k+l \text{ even}}} J_{k+l}\left(\frac{2\pi r}{L}\sqrt{m^2+n^2}\right)\cos\left[(k-l)\frac{\varphi}{2}\right] \\ & \cdot\left\{f_{k,l,m,n}\cos\left[(k+l)\left(\hat{\alpha}-\frac{\varphi}{2}+\tau_{nm}\right)\right] - g_{k,l,m,n}\sin\left[(k+l)\left(\hat{\alpha}-\frac{\varphi}{2}+\tau_{nm}\right)\right]\right\} \\ & - 2\pi \sum_{\substack{k,l,m,n: \\ k+l \text{ odd}}} J_{k+l}\left(\frac{2\pi r}{L}\sqrt{m^2+n^2}\right)\sin\left[(k-l)\frac{\varphi}{2}\right] \\ & \cdot\left\{g_{k,l,m,n}\cos\left[(k+l)\left(\hat{\alpha}-\frac{\varphi}{2}+\tau_{nm}\right)\right] + f_{k,l,m,n}\sin\left[(k+l)\left(\hat{\alpha}-\frac{\varphi}{2}+\tau_{nm}\right)\right]\right\}. \end{aligned} \tag{373}$$

Using Eq. (373), we can finally compute:

$$q(\hat{\alpha},\varphi) = h(R,\hat{\alpha},\varphi) + 2\pi\sum_k |A_k|^2 e^{-ik\varphi}. \tag{374}$$

D | DESCRIPTION OF THE DATASET

In this appendix we give a more detailed description of the analyzed dataset and of the experimental protocol, with a focus on the possible sources of 'measurement errors'. We argue in the main text (Section 3.5.2) that there must be spurious contributions in the reconstructed trajectories, which have a comparable effect to the one that superimposing an uncorrelated noise series to the model time series has.

The dataset

The analyzed data belong to the second collection campaign of the CoBBS group, which was conducted mainly in the winter season 2011/2012. In all data collection campaigns, bird flocks are always observed in natural conditions during their aerial display before roosting. In nature, birds easily gather in large groups of more than thousands or even tens of thousands of individuals, a fact that creates severe issues for the reconstruction of 3D individual coordinates from such dense clouds of moving animals. However, starting from the STARFLAG project, a modern methodology for the observation of large groups of birds outside the lab has been developed. An overview of the experimental setup and reconstruction procedure is contained in [45, 46] and in [43, 2] (the latter regarding dynamic reconstruction).

The dataset of the second campaign comprises flocks mostly performing collective turns: the same data were indeed used in [3], where the Inertial Spin Model was proposed. The size of these flocks, in the range of hundreds, is smaller than those collected during the first experimental campaign (STARFLAG 2007-2010). This is due to the different apparatus used in the two campaigns. The first experimental campaign was performed with Canon D1 Mark ii cameras shooting interlaced at 10 fps, with a resolution of 8.2 MP and with a sensor size of 28.7mm x 19.1 mm. Such a low frame rate represented a severe limitation for the *dynamic* reconstruction of the flocks. The empirical study of the first campaign was indeed limited to static reconstruction of birds' relative positions. To overcome this limitation, the second experimental campaign was carried out with much faster cameras IDT M5, shooting at a maximum speed of 170 fps, but with a much lower resolution (4MP) and with a smaller sensor (16mm x 8.1 mm), which results in a dramatic reduction of the size of the field of view and consequently of the potentially collectable flocks.

Few details about the observed flocking events, e.g. concerning the size of the flock and the length of the observation, are summarized in Table 2. The analyzed data were obtained using the GReTA algorithm for 3D dynamical reconstruction [2].

Table 2: The analyzed dataset consists of the output of the reconstruction algorithm GReTA. For each flocking event, we have 3D trajectories of maximum length t_{max}. Time is measured in number of recorded steps. Not all of the trajectories are of the same length, due to reconstruction errors which may create breaks. N_{full} birds are recorded for the full length t_{max} of the observation; $\bar{N}_{in}$ is the average number of birds appearing simultaneously in a single frame. Finally, $\bar{t}_{in}$ is the average length of reconstructed trajectories or subtrajectories, without breaks in the middle.

Event	obs. rate	t_{max}	N_{full}	$\bar{N}_{in}$	$\bar{t}_{in}$
20120209_ACQ1	170 (fps)	600	231	373	536
20111214_ACQ4F1	170 (fps)	761	49	126	579
20111214_ACQ4F2	170 (fps)	708	45	115	507
20111215_ACQ1	170 (fps)	976	123	305	753
20111220_ACQ2	170 (fps)	300	95	145	230
20111222_ACQ1	170 (fps)	609	59	62	596
20111201_ACQ3F4	170 (fps)	400	189	354	320
20111201_ACQ3F1	170 (fps)	500	62	112	410
20110211_ACQ1	80 (fps)	360	420	564	334
20110208_ACQ3	80 (fps)	440	133	166	412
20110217_ACQ2	80 (fps)	150	301	378	142
20111124_ACQ1	170 (fps)	310	97	120	302
20111125_ACQ1	170 (fps)	955	44	49	900
20111125_ACQ2	170 (fps)	761	259	424	621

Stereometry experiments

In order to understand the origin of possible measurement errors, whose relevance in the inference procedure was highlighted by the trends in Figs. 14–15, it is useful to pause over some technicalities concerning the experimental protocol. A graphical abstract of the entire strategy is provided in Fig. 40.

Flocks are observed at a distance, using the technique of stereometry and employing still cameras with high digital resolution. These cameras are professionally calibrated in the lab to estimate their internal parameters (focal length, position of the image center and distortion coefficients), and on a daily basis in the field to estimate the external parameters of the system (mutual orientation and position of the cameras). The 2D images shot by each of them are segmented to isolate the birds from the background. Several specific problems related to the recognition of individual birds may arise in this segmentation step and are discussed in [45]. What is of interest for us is that the experimental setup is chosen and designed in such a way that each bird covers at least 4 pixels on the CCD (but typically many more). The position of the bird on the 2D photo is identified with the center of mass of the intensity distribution, after background subtraction and blob splitting. A *blob* is an ensemble of two or more birds that coalesce and appear as a unique feature in the 2D picture of the camera.

Due to the finite resolution of digital images, the segmentation process introduces unavoidable errors that affect the reconstructed position of the birds. The upper bound for this segmentation error may be computed from 3D reconstruction

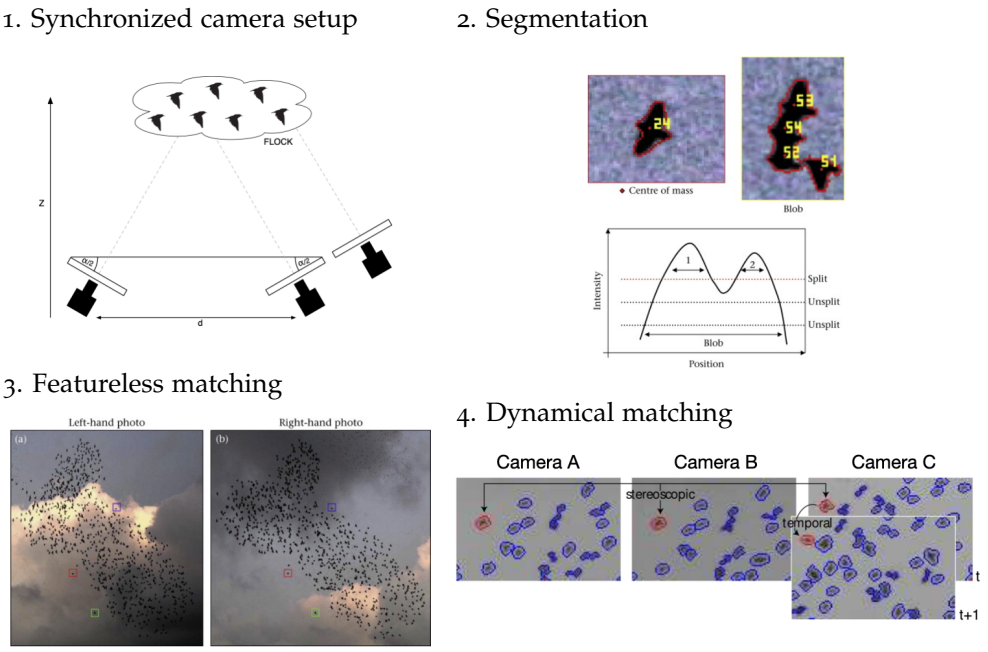

Figure 40: Graphical abstract of the experimental procedure. The figures are reproduced with permission from [45, 2]. The first step involves the true experiment, whose difficulties are in the calibration and synchronization of the cameras. Three cameras, placed as shown in the figure, are used for stereometric photography (six in the first type of experiments in order to double the shooting rate). Image segmentation is performed on each frame recorded by each camera. The 3D static reconstruction is done in step 3 using only stereometric photography. The dynamic reconstruction of birds' trajectories is performed in a recursive way combining stereometry with bird matching in subsequent frames.

formulae [33], using standard error propagation and assuming a segmentation error of 1px. In our experimental setup (baseline of the system equal to d=25m, birds at z=125m from the cameras and 28mm lenses) this upper limit corresponds to a maximum error of 15cm in the reconstructed positions. It is worth noting that this error refers to absolute positions, whereas for all practical purposes concerning this work, we are interested in birds' velocities, computed as the difference between positions in consecutive frames. When comparing the images of the same bird in consecutive frames, or more in general in frames close in time, its shape essentially does not change, hence the images are affected by segmentation errors of the same order, which actually compensate for each other, reducing the overall error on the reconstructed velocity.

Other sources of error may be related to the calibration of the cameras, which are tested [33] and checked to be negligible compared to the major error source coming from segmentation.

C'è in noi un bisogno radicale di riconoscimento che non ha niente a che fare con l'ammirazione, la stima, la fama. È come un bisogno di benedizione, di parentela o almeno di familiarità, di iniziazione superata, di passaggio a stirpe che ti sceglie all'improvviso e ti dà il nome.

Chandra Candiani

ACKNOWLEDGEMENTS

I would like to start by thanking the referees, Profs. Greg Stephens and Julien Tailleur for having kindly accepted to examine my manuscript.

This PhD was not as I had imagined it. Whether it is good or bad news that it came to an end is hard to tell, but it is certain that its completion would have not been possible without the support — both human and scientific — of a number of people who is impossible to acknowledge in full detail.

I need to firstly thank my supervisor Prof. Irene Giardina, and Profs. Aleksandra Walczak and Thierry Mora, who hosted me in their lab at École Normale Supérieure in Paris to collaborate on the presented projects. I am grateful to Irene for the trust and independence I received over these three years of work, for the freedom I had to propose ideas and follow them, and even to crash into my own limits, mistakes or inadequacies. I am indebted to Aleks and Thierry for giving me the opportunity to repeatedly join the Statbiophys group at ENS and for involving me in all their activities, as well as for being examples of restless commitment to their job.

I really thank the past and present members of the experimental team in Rome (in the first place Dr. Stefania Melillo), who collected and processed the raw data, and allowed me to analyze what resulted from their long and hard work.

My thanks go then to all the colleagues of the PhD program in Rome and of the Statbiophys group at ENS, with whom I shared any kind of physics- and non physics-related discussion, litres of coffee, some research activity and many (too many) complaints: you are too many, beyond those I cannot avoid mentioning — Marco, Victor, Natanael, Giulia, Luca, Antonio, Giulio — but thanks (and apologies) to you all!

I also want to undistinguishably acknowledge all those people — professors, students, senior and junior researchers — that I had the chance to meet over these years, virtually or in person, with whom I could exchange ideas or from whom I could take inspiration, knowledge and example.

Thanks to my closest friends, always willing to lend an ear or trying to remind me not to take myself too seriously, no matter if from a thousand km distance.

Finally, I'd like to thank my family for their unconditional tacit support, and especially my sister, to whom this thesis is dedicated.