The physics of flocking: Correlation as a compass from experiments to theory

Andrea Cavagna [a,*], Irene Giardina [b,a,c], Tomás S. Grigera [d,e,f]

[a] *Istituto Sistemi Complessi, Consiglio Nazionale delle Ricerche, 00185 Rome, Italy*
[b] *Dipartimento di Fisica, Università Sapienza, 00185 Rome, Italy*
[c] *INFN, Unità di Roma 1, 00185 Rome, Italy*
[d] *Instituto de Física de Líquidos y Sistemas Biológicos (IFLYSIB), CONICET y Universidad Nacional de La Plata, Calle 59 no. 789, B1900BTE La Plata, Argentina*
[e] *CCT CONICET La Plata, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina*
[f] *Departamento de Física, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Argentina*

A R T I C L E   I N F O

A B S T R A C T

Collective behavior in biological systems is a complex topic, to say the least. It runs wildly across scales in both space and time, involving taxonomically vastly different organisms, from bacteria and cell clusters, to insect swarms and up to vertebrate groups. It entails concepts as diverse as coordination, emergence, interaction, information, cooperation, decision-making, and synchronization. Amid this jumble, however, we cannot help noting many similarities between collective behavior in biological systems and collective behavior in statistical physics, even though none of these organisms remotely looks like an Ising spin. Such similarities, though somewhat qualitative, are startling, and regard mostly the emergence of global dynamical patterns qualitatively different from individual behavior, and the development of system-level order from local interactions. It is therefore tempting to describe collective behavior in biology within the conceptual framework of statistical physics, in the hope to extend to this new fascinating field at least part of the great predictive power of theoretical physics. In this review we propose that the conceptual cornerstone of this ambitious program be that of correlation. To illustrate this idea we address the case of collective behavior in bird flocks. Two key threads emerge, as two sides of one single story: the presence of scale-free correlations and the dynamical mechanism of information transfer. We discuss first static correlations in starling flocks, in particular the experimental finding of their scale-free nature, the formulation of models that account for this fact using maximum entropy, and the relation of scale-free correlations to information transfer. This is followed by a dynamic treatment of information propagation (propagation of turns across a flock), starting with a discussion of experimental results and following with possible theoretical explanations of those, which require the addition of behavioral inertia to existing theories of flocking. We finish with the definition and analysis of space–time correlations and their relevance to the detection of inertial behavior in the absence of external perturbations.

© 2017 Elsevier B.V. All rights reserved.

Contents

* Corresponding author.
  E-mail address: andrea.cavagna@roma1.infn.it (A. Cavagna).

1. Introduction

1.1. Biology, physics and the quest for universality

Biological systems are invariably complex, strongly interacting, and out of equilibrium: a nightmare for the physicist's "spherical horse" approach. The topic of this review, namely collective motion in animal groups, makes no exception. The first level of complexity in this problem is related to the strong interactions among individuals, which determine the collective behavior of the group. But it is also clear that the individuals are themselves quite complex and out of equilibrium. Besides, there is system diversity. Collective motion spans across many space and time scales, from microscopic organisms (bacteria and cell clusters), to insects swarms (mosquitoes and midges), up to vertebrate groups (bird flocks, fish schools and mammal herds), living in very different environments. How can one hope to tackle these systems other than with specialized tools for each situation? Yet many things have been understood about the problem of collective motion in biological systems by using physics-inspired approaches that tackle different groups using similar conceptual tools. One may wonder where this general understanding comes from.

A similar challenge is at the core of statistical physics, which studies systems where the large number of individual components (particles) allows the successful application of a probabilistic approach. In this way the daunting complexity of the system is tamed by focusing on a restricted set of properties: by giving up the hope to track individual particles, one can understand and even predict the behavior of collective experimentally accessible variables, like pressure, temperature and magnetization. In critical phenomena, an even more striking simplification occurs: systems whose fluctuations are correlated over long ranges display *universality*, namely the fact that their large-scale properties are independent from the microscopic interactions, and only determined by the dimensions and symmetries of the system. In this way the details become irrelevant, and a wide variety of systems can be described by very simple models, as is the case with the Ising or Heisenberg models of magnets.

In the past 30 years or so, biological systems characterized by a large number of interacting units have been studied within a statistical physics approach, and this is what we do here for the problem of flocking. A common feature to all these systems, which partly justifies such approach, is that they are all characterized by strong correlation, a key ingredient of the statical physics' aim at universality. Despite this, a possible objection is that statistical physics methods should be relevant for systems with $\sim 10^{23}$ particles, while in collective behavior, and especially in flocking, we are often looking at groups of a few thousands of individuals; scaling laws should apply at length scales of more than tens or hundreds interparticle distances, which are larger than the size of the whole group. Moreover, the "particles" in the biological problem are complex individuals which act as energy sources and can in principle move with laws completely different from the dynamical rules that apply to physical systems. So, even if one sees that a statistical physics treatment of biological systems works in a couple of cases, the question arises about how long we can get away with it before hitting the wall of the biological uniqueness of each system.

From a theoretical point of view, the statistical physics of out-of-equilibrium systems is still an open and rapidly evolving chapter of statistical mechanics and the question of under what dynamical conditions a statistico-mechanical approach works, is still largely unanswered. From an experimental point of view, our best chance is to gather an increasing amount of quantitative data where to test statistical physics concepts and tools, to prepare a firmer empirical basis upon which building a theory of strongly correlated biological systems. For this reason, the scope of this review is quite restricted: we only consider flocks, and only those aspects of this system for which experimental data is available to contrast or confirm the predictions of theories and models. On the other hand, we go into quite a bit of detail on the theoretical tools we use to interpret the experimental data. We aim to present the state of the art of our knowledge of collective motion of animal groups and its understanding in the framework of statistical physics. Our understanding and our experimental information has still many missing pieces and we make no attempt to hide this fact. We present here what we deem a useful summary of progress made over the last decade or so, which points to hope for the great predictive power of theoretical physics can be applied to this kind of biological problem.

1.2. Correlation

The overarching concept of our work is that of *correlation*. Correlation is such a fundamental concept in physics that the mean physicist would normally not bother discussing its significance. However, this is a review which aims at capturing the interest also of an audience in biology, hence a few words about what we, as physicists, mean by correlation will be useful. Collective behavior invariably appears in systems composed of many units (particles, spins, animals, cells, etc.) that interact

with each other. Interacting means exchanging information in a direct way; in physics this typically occurs through forces acting on the particles which are mutually within their interaction range; in biology we often speak of 'social forces', which is a merely nickname to describe the complex ways in which biological entities exchange direct information or directly act on each other in any way. The key word when explaining interaction is '*direct*': interaction between units *i* and *j*, whatever these units are, is *not* mediated by any third unit *k* (it could me mediated by some physical medium, though). The best way to see this is to ask: would *i* and *j* be influencing each other even if all the rest of the group disappeared? If the answer is yes, then we can say that *i* and *j* interact with each other (although the way in which *i* interacts directly with *j* may depend on whether *i* and *j* are alone, or within a group). Correlation, on the other hand, is the *indirect* result of interaction: *i* exchanges information with *j* that exchanges information with *k* that exchanges information with *l*. The result is that *i* and *l* may share some information, and perhaps 'behave' similarly even though they never talked to each other (no direct interaction). When this happens, information has spread indirectly from *i* to *l* (and vice versa) and we say that *i* and *l* are correlated.

Such indirect exchange of information comes in two types, static correlation and dynamic correlation. The first is easier to compute and to define, but the latter is more intuitive. Static correlation does not know about the dynamical mechanism through which information goes about the system, but it simply measures to what extent different parts of the system behaves similarly *at the same time*. This may seem weird, as if no time has passed one may wonder how information has been transmitted between those two positions in space. In fact, static correlation between two parts of a system always does imply the existence of dynamical phenomena that have contributed to connect these two parts is some way in the past; however, to measure static correlations we do not need to know anything about these past dynamical phenomena: we can simply average over the many possible configurations the system is found in, irrespective of how it arrived there. On the other hand, we can also compute temporal correlations, namely ask how much one part of the system at one instant in time is correlated to another part at a different time. In this case, as we shall see, correlation *is* connected to the actual propagation of information across the system.

We have said that correlation measures to what extents different parts of the system behaves similarly, but this is somewhat misleading and it must be specified more carefully. What is really useful to measure is how similar the *fluctuations* are, namely the deviations with respect to the mean behavior of the system. This is what physicists call the *connected* correlation function. The connected correlation function measures how much the deviation of *i* with respect to the mean behavior of the system is similar to the deviation of *j*; this deviation with respect to the mean behavior is what is called *fluctuation*. In this way, the connected correlation eliminates the somewhat trivial effect of behaving similarly just because the whole system is doing so. For example: a flock of starlings is a strongly polarized system, everybody is going (approximately) in the same direction, similarly to what happens in a ferromagnet at low temperature, where each spin points in a similar direction. The non-connected correlation, i.e. the correlation between the full velocity of *i* and that of *j* is always very large, irrespective of the distance between *i* and *j*, be it in space or time; this simply happens because the whole system has developed long-range order, so that non-connected correlation is merely a replica of the mean order of the group. Connected correlation, on the other hand, measures to what extent the *change* of direction of one individual is similar to the change of direction of another individual, which is nontrivial. The experimental determination and theoretical discussion of connected correlations, both at the static and dynamical level, will be the *fil rouge* of our work.

1.3. What the reader can find in this review

It is important to make clear from the outset that in this review we will be mostly discussing our own work, rather than actually reviewing the numerous important contributions of other groups to the physics of flocking and to collective behavior more in general. Fortunately for the reader, in the last decade or so at least two major review works appeared, the book by David Sumpter [1] and the review article by Tamas Vicsek and Anna Zafeiris [2], which, together with the most recent reviews in active matter [3,4], provide a very comprehensive and balanced view of the field. As we said, our scope is more limited: we want to focus the reader's attention on the path connecting space–time correlation to information propagation, and we aim to do so within the rather well-defined experimental arena of polarized biological systems, i.e. flocks. However, even within such a narrow scope, it was essential to discuss in detail two major pillars of the physics of flocking, namely the Vicsek model [5] and the hydrodynamic theory of Toner and Tu [6].

It is nowadays fashionable to write that the Vicsek model is too basic, lacking in biological realism. Indeed the Vicsek model is simple: particles get around with constant speed and their only interaction is the tendency to align their direction of motion with that of their local neighbors; there is no cohesion, no short-range repulsion, no speed fluctuations, no attempt to make the flock do anything smart or indeed anything at all. Yet, in our opinion, the brutal simplicity of the Vicsek model is its most precious asset: it is the quintessential statistical mechanics model, in that it captures the essential trait of the phenomenon under investigation, limiting to the bare minimum the number of details and parameters; it is not exaggerated to say that it is the Ising model of flocking. For these reasons, the Vicsek model will be the constant reference frame of our work in terms of modeling. Toner and Tu's hydrodynamic theory has a similar role. Originally derived as a continuum version of the Vicsek model, it promoted the topic of flocking from the somewhat narrow boundaries of numerical simulations of discrete models, up to the very powerful framework of field theory, critical phenomena and the renormalization group. For the first time, within the Toner–Tu theory, the physics of flocking was able to *calculate* essential things like correlation functions, critical exponents, etc. This has been an enormous step forward for the whole field. As for the Vicsek model, Toner–Tu theory will therefore be a constant conceptual reference all along this work. Part of our effort will be dedicated

to generalize the Vicsek model and the Toner–Tu theory in order to capture a key experimental phenomenon, namely the linear propagation of information within flocks. We hope it will emerge clearly that this generalization would have been impossible without these two fundamental descriptions of collective behavior. Our aim was not to oust the Vicsek model, nor the hydrodynamic theory of Toner and Tu, but simply to add a new ingredient (inertia), to make them more powerful, yet keeping the parameters complexity to the bare minimum.

1.4. Plan of the work

We will start in Sections 2.1.1 and 2.1.3 with a primer on how to compute static correlations in biological systems. The system we will study (real flocks) is not the typical statistical mechanics one, hence all classic definitions of correlations have to be modified and adapted, with some relevant technical issues to keep into consideration; to do this we will directly present our tools on real experimental data about flocks, in order to make the discussion as practical as possible. Once this will be done, we will study static correlations in flocks of starlings and see what they tell us about the interaction ruling the systems (Section 2.2). The main experimental result here will be the scale-free nature of such correlations, namely the fact that different parts of the systems are correlated in a long-range way, which seems an interesting phenomenon. After some further technical details on the correlation functions (Section 2.3), Section 2.4 deals with the problem of building models compatible with the experimental information on correlations. However, all these discussions are static, hence it is unclear how and why scale-free correlations should be relevant in determining how information propagates within a biological system (Section 2.5).

This question will lead us to the second part of the work, where we will present experimental results about how information travels across a flock (Section 3.1). We will see that the change of direction of one bird spreads very efficiently to the rest of the group giving rise to a collective turn. We will demonstrate that scale-free correlations are indeed a necessary condition to achieve this result, but *not* a sufficient one: previous theories of collective motion had scale-free correlations, and yet were unable to describe information propagation of the sort we observe in real flocks (Section 3.2). We will therefore present a new theory of flocking, whose main player will be behavioral inertia in Section 3.3.

Measuring how a signal propagates across a system is certainly the best and most direct way to infer what is the equation of motion ruling a system; however, it is not always practical to do so experimentally. Taking inspiration from statistical physics, we will therefore show that one can get information about signal propagation from the way spontaneous dynamical fluctuations relax, without having to measure a signal actually propagating through the system. This will naturally lead us to the third and final part of the work, dedicated to dynamical correlation functions (Section 4). Here we will show that static correlation, information propagation and dynamical correlation are all manifestations of the same thing, namely of the dynamical equation ruling the behavior of the system, which is ultimately what we are after.

2. Scale-free correlations

Correlation functions play a central role in the description of collective phenomena, both at the theoretical and experimental level. In this Section we will introduce several correlation functions, briefly recall their meaning, illustrate experimental results and discuss their consequences from a theoretical perspective. Many biological systems, and animal groups in particular, exhibit ordered patterns or coordinated behavior in their movement. Our main interest will therefore be in those degrees of freedom that mostly characterize such patterns: velocities, flight directions and speeds. We will introduce global order parameters quantifying the degree of collective ordering in the system (global polarization, group velocity and average speed) and investigate individual fluctuations and their correlations. In this respect, magnetic models represent an important archetype of long-range directional order and will recurrently provide a benchmark and a guideline to describe and understand these systems. Due to the active nature of their components, biological assemblies also present non-trivial structural properties: individuals in a group exchange positions and change their mutual arrangement in space, determining an interaction network that continuously evolves in time. Directional and positional degrees of freedom are in general coupled, making the description of active matter extremely complex [7]. Many two-dimensional active systems, for example, display giant density fluctuations and traveling ordered bands, symptomatic of strong coupling between local density and local order [3,4]. On the contrary, many natural systems moving in three dimensions live in a regime where directional fluctuations dominate over positional ones [8]. This is the case which mostly interests us in this section.

2.1. The connected correlation function

2.1.1. Order parameter

We consider a group of N 'particles' (birds, but also insects, fish, etc.), placed at positions $\{\mathbf{r}_i\}$, $i = 1 \ldots N$, moving in space with individual velocities $\{\mathbf{v}_i\}$. We are interested in the collective order that arises in the velocity degrees of freedom. In some systems, positions may also develop a complex structure as a consequence of their coupling with velocities, but in our case of interest (bird flocks) the main source of order are the velocities (the structural order is indeed quite poor in flocks [9]). Thus we introduce the following order parameters to describe the degree of collective order:

$$\Phi = \frac{1}{N} \sum_i \frac{\mathbf{v}_i}{|\mathbf{v}_i|}, \qquad\qquad \text{polarization,} \qquad\qquad (1)$$

$$\mathbf{V} = \frac{1}{N} \sum_i \mathbf{v}_i, \qquad\qquad\qquad\qquad \text{average velocity,} \qquad\qquad (2)$$

$$S = \frac{1}{N} \sum_i |\mathbf{v}_i|, \qquad\qquad\qquad\qquad \text{average speed.} \qquad\qquad (3)$$

The polarization Φ measures the degree of global alignment; its modulus (scalar polarization, noted Φ) takes values between 0 and 1. Strongly polarized groups like flocks of birds typically have Φ near 1, large values of $|\mathbf{V}|$, and small speed fluctuations around the mean [10], while non-polar aggregates like swarms of insects have rather low polarization and a very small average velocity [11,12].

2.1.2. Fluctuations

When discussing response to external perturbations and propagation of information through the group, fluctuations, i.e. the departure of individual behavior from the group average, are the most relevant quantities. Most importantly, when considering correlations fluctuations will be the main players, as these are the quantities that can distinguish a truly interacting system with emergent behavior from a non-interacting one, in the presence of global order. Thus a very important quantity in what follows will be the instantaneous deviation of the velocity of i from its global mean,

$$\delta\mathbf{v}_i \equiv \mathbf{v}_i - \mathbf{V} = \mathbf{v}_i - \frac{1}{N} \sum_k \mathbf{v}_k. \qquad\qquad (4)$$

Similarly, we define the speed fluctuations as,

$$\delta s_i = |\mathbf{v}_i| - S, \qquad\qquad (5)$$

The fluctuations are the building blocks of the connected correlation function, which we now define.

2.1.3. Definition of the correlation function

To investigate the collective nature of the system's behavior we introduce the *connected* correlation functions, i.e. the space correlations of the velocity fluctuations (4). For a given pair i and j, the product $\delta\mathbf{v}_i \cdot \delta\mathbf{v}_j$ measures the degree of similarity between the two individual deviations; ideally, one would like to perform an average of this product over some statistical ensemble, $\langle \delta\mathbf{v}_i \cdot \delta\mathbf{v}_j \rangle$, in order to compute the correlation between i and j, as we do in ordinary statistical physics. Unfortunately, in collective animal behavior we do not have a well-defined statistical ensemble over which to perform such average. However, as long as interactions are local and distance dependent, the only property that affects the correlation between two individuals is their mutual distance. We can therefore group pairs of individuals according to their mutual distance and perform an average over all such pairs. We therefore define the distance dependent correlation function as,

$$C(r) = \frac{\sum_{i,j}^N \delta\mathbf{v}_i \cdot \delta\mathbf{v}_j \, \delta(r - r_{ij})}{\sum_{k,l}^N \delta(r - r_{kl})}, \qquad\qquad (6)$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$. Definition (6) quite naturally encapsulates the idea of a *spatial* average: we sum all the products $\delta\mathbf{v}_i \cdot \delta\mathbf{v}_j$ for those pairs i and j with a distance r_{ij} between r and $r+dr$, and then we divide by the number of such pairs (the denominator). If the system is large enough Eq. (6) involves an average over many different pairs and $C(r)$ gives a fair estimate of the statistical correlation at scale r at a certain instant of time. If several experimentally determined configurations are available under similar boundary conditions, a time average over the different configurations can also be performed, although before doing such average it is important to check that the system under investigation is at steady-state; if, on the other hand, there is a clear trend with time of the correlation function, it does not make sense to average over time. Connected correlation functions can be defined in a similar way also for the fluctuations of the speeds $|\mathbf{v}_i|$ and of the flight directions $\mathbf{v}_i/|\mathbf{v}_i|$. As we shall see, it is indeed useful to distinguish the directional degree of freedom from the modulus of the velocity.

The denominator of (6) is the number of pairs at mutual distance r; in a isotropic and homogeneous system, and *for large distances* r, this number can be approximated as the volume of the shell of radius r times the density ρ_0 of the system, that is,

$$\sum_{k,l}^N \delta(r - r_{kl}) \sim N\rho_0 4\pi r^2, \quad \text{isotropic – homogeneous – large distance – approx} \qquad\qquad (7)$$

where N counts the number of centers for each pair. For this reason, one may be tempted to use a different definition of the correlation function, namely,

$$\hat{C}(r) = \frac{1}{N\rho_0 4\pi r^2} \sum_{i,j}^N \delta\mathbf{v}_i \cdot \delta\mathbf{v}_j \, \delta(r - r_{ij}), \qquad\qquad (8)$$

which is more similar to the definition used in standard liquid theory [13]. However, using (8) would not be a good idea. Compared to (8), definition (6) conveniently separates and protects the correlation of the order parameter fluctuations

(in this case the orientation, $\delta\mathbf{v}_i$), from the (often nontrivial) correlations of the fluctuations in the *positions* of the particles, i.e. in the density. The point is the following: the sum over i and j of $\delta(r - r_{ij})$ is highly sensitive to the spatial distribution of the particles; however, in (6) this term appears both at the numerator *and* at the denominator, hence canceling out, while in (8) it appears only at the numerator, whereas the denominator assumes an homogeneous and isotropic mass distribution. As a result, definition (6) focuses *only* on the correlations of the order parameter (not of the positions), while (8) mixes the two kind of fluctuations (order parameter and positions) in a way that is difficult to disentangle. There are two crucial reasons why separating these two effects is very important:

(i) Discrete mass distribution. By far the most basic reason is that (8) is far too sensitive to the discrete nature of the system. Consider the simplest case, namely a system on a fixed regular lattice, for example the Ising model on cubic lattice. In this case, up to a few lattice spacings, definition (8) has very strong spikes due to the difference between the actual number of pairs on the cubic lattice, and its approximated expression, $N\rho_0 4\pi r^2$. Of course, for large r this effect vanishes, but it may still mask completely the correlation of the order parameter for small-intermediate r. Definition (6), on the other hand, emancipates completely $C(r)$ from the lattice structure.

(ii) Open boundary effects. In real biological systems we invariably have a finite size open border, hence we cannot work in the bulk, nor with period boundary conditions. This means that the number of pairs stops growing as r^2 beyond a certain distance, which is of the order of the system's size; hence, (8) is normalizing by the wrong factor, as the volume of the shell starts being empty when parts of the shell are out of the system. On the other hand, definition (6) always normalizes by the actual number of pairs, hence giving the correct average.

When we stress that the right definition of correlation function of the order parameter must separate its fluctuations from those of the mass distribution, we do *not* want to suggest that the latter are irrelevant. Of course, real biological systems do not sit on fixed regular lattices, and density fluctuations are indeed a crucial part of the physics active matter; for example, giant density fluctuations have been experimentally observed in different systems [3,4], and more in general we do not know a priori whether or not the system is homogeneous and isotropic. But exactly for this reason, namely because density fluctuations are interesting *per se*, we need to investigate them separately, and not entangled with the order parameter fluctuations, as (8) does. The correct way to separately investigate density fluctuations is of course to compute the density correlation function, namely the excess of particle pairs over the volume shell,

$$g(r) = \frac{1}{N\rho_0 4\pi r^2} \sum_{i,j}^{N} \delta(r - r_{ij}) \,, \tag{9}$$

which coincides with the radial distribution function defined for liquids [13] (except at $r = 0$ - see Appendix A). The function $g(r)$ is proportional to the probability to have a particle at distance r from any given particle and it therefore describes how particles (or individuals) are arranged in space. In a completely homogeneous system $g(r) = 1$, but in general $g(r)$ has a non-trivial space dependence (e.g. it has spikes on regular lattices/crystal, broad peaks in a liquid, or a more exotic structure in presence of anomalous density fluctuations [13]). Once defined $g(r)$, we finally realize that the connection between the two correlation functions, $C(r)$ and $\hat{C}(r)$, is simply,

$$C(r) = \frac{\sum_{i,j}^{N} \delta\mathbf{v}_i \cdot \delta\mathbf{v}_j \, \delta(r - r_{ij})}{\rho_0 N 4\pi r^2 g(r)} = \frac{\hat{C}(r)}{g(r)} \,. \tag{10}$$

For an homogeneous system $g(r) = 1$ and in that case dividing by number of pairs is equivalent to dividing by the shell volume; however, we have seen that this is not at all the generic case. We therefore understand that by dividing by $g(r)$ in Eq. (10) we are disentangling the correlations of the order parameter (the velocities) from those of the density (see Appendix A for an expanded discussion of this point, and for a definition of correlations starting from velocity fields rather than individual velocities).

The bottom line is that order parameter and density fluctuations must be studied separately, hence one should use (6) for the former and (9) for the latter. For natural flocks, experimental measurements have shown that the structure of $g(r)$ is quite bland [9]; its only nontrivial feature is a (rather expected) drop of probability at very small r, clear signal of a hard core: birds have a minimum nearest neighbor distance roughly equal to their wingspan [14]. Apart form this, no other interesting feature emerges in the density correlation of flocks. Hence, from now on, we will solely focus on the velocity fluctuations correlation function, (6). However, other organisms can have nontrivial correlations of the density fluctuations, hence in general it is important to study separately $g(r)$ and $C(r)$.

2.1.4. Non-connected vs connected correlation

Connected correlations are the primary tool of theoretical analysis and experimental work for several reasons. As we noticed in the Introduction, a crucial feature of connected correlations is that they measure the amount of similarity between the *fluctuations* $\delta\mathbf{v}_i$ of the velocities (or of any other quantity of interest), rather than the velocities themselves. If the system displays global order all the individual $\mathbf{v}_i$ are necessarily similar to one another to the extent that $\mathbf{V} = 1/N\sum_i \mathbf{v}_i \neq 0$: everybody is pointing on average in the same direction, defined by the collective group motion. The simple non-connected correlations between velocities are therefore dominated by this common contribution. On the other hand, connected correlations – defined in terms of the deviations of the velocities from their average value – capture correlations beyond the effect of global order. An important consequence of this is that connected correlations can to discriminate between systems
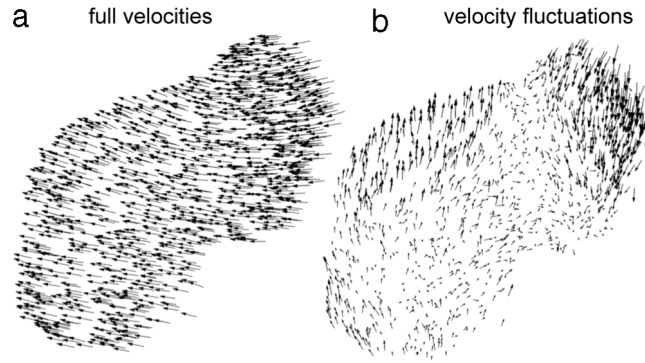
Fig. 1. Individual velocities in natural flocks of birds. (a) 2-*d* projection of the individual velocities within a starling flock at a fixed instant in time (1246 birds, linear size $L = 36.5$ m). Vectors are scaled for clarity. The flock is strongly ordered and the velocities are aligned. (b) 2-*d* projection of the individual velocity fluctuations of the same flock at the same instant (vectors scaled for clarity).
Source: From [10].

where order is due to some external common cause, in absence of mutual interactions, and systems where collective order is truly an emergent phenomenon (due to interactions between individuals). A simple example comes from ferromagnetic systems. If a lattice of non-interacting spins is subject to an external magnetic field, each spin will independently align with the field and the system will display a non-zero global magnetization, i.e. collective order. The non-connected correlation between spins will also be different from zero but for a trivial reason: all spins are aligned with the field and therefore also with one another. However, since spins do not interact they respond independently to thermal noise: their fluctuations from the average are therefore uncorrelated, leading to a zero connected correlation. If we consider instead a lattice of spins with local interactions and no external field, something different happens: at low temperature the system develops a non-zero global magnetization, but the spins fluctuate in a correlated way due to mutual interactions and $C(r)$ is therefore different from zero. In the context of collective animal behavior, the above examples would correspond to cases where collective motion is due to the presence of a leader individual, to an external global disturbance or to the movement of the ambient medium ($C(r) \sim 0$), or cases where it derives from the 'social' forces between group members ($C(r) \neq 0$).

We note that for an interacting system connected correlations can also be different from zero (and strong) even in absence of order. This is for example what happens close to critical points separating an ordered phase from a disordered one in condensed matter physics [15,16]. Swarms of insects are an example of animal groups where there is no collective motion (swarms are directionally disordered) but correlated patterns (described by a non-trivial $C(r)$) occur at the group scale [11,12]. Connected correlations therefore represent a better signature of collective behavior than order itself. More broadly, in statistical physics it is the connected correlations that are tuned by the control parameter of the system, display scaling behavior, and obey fluctuation–dissipation relationships. If, as we hope, collective behavior in biological systems can be described within a similar approach, it is from connected correlations that we must start our investigation. To illustrate the relevance of correlations, in the next sections we immediately show the correlation function (6) as computed from experimental data in natural flocks of birds, and discuss the implications of such findings.

2.2. Experimental evidence

The correlation function $C(r)$ has been computed in natural flocks of starlings in [10]. In that work, stereo-experiments were performed in the field with a high resolution video-acquisition system [17]. Using computer vision techniques the instantaneous three-dimensional positions and velocities of the individual birds in large groups of starlings have been retrieved for many flocking events. Fig. 1 shows the individual velocities $\mathbf{v}_i$ (panel A) and the corresponding fluctuations $\delta \mathbf{v}_i = \mathbf{v}_i - \mathbf{V}$ (panel B) at a given instant in time for a flock of approximately 1000 individuals. Although the figure displays the two-dimensional projections of the velocity and of the fluctuation vectors, it clearly shows the large degree of global ordering (polarization is above 0.9) and the presence of two large correlated domains. The extension of such domains can be quantified by looking at the connected velocity correlation function $C(r)$, Eq. (6). In general, the correlation function decays from large values at short distances, to smaller values when individuals do not belong to the same domain: the decay range of $C(r)$ – the correlation length ξ – is therefore a measure of how large these domains are.

The $C(r)$ obtained from the data is shown in Fig. 2a. As can be seen from this figure, the correlation does not display any exponential decay (which would provide a natural decay length-scale); rather it decreases almost linearly up to very large mutual distances. As discussed more in detail in Section 2.3.3, the correlation length ξ can be estimated in this case as the point where the correlation crosses the zero axis. This point indeed marks the distance beyond which fluctuations become anti-correlated, corresponding to pairs of individuals that deviate from the mean group velocity in an opposite way. It therefore measures the extension of the correlated domains observed in Fig. 1b. For a given flocking event, this procedure can be performed at different times to obtain the average correlation length for that flock. Fig. 2c shows the
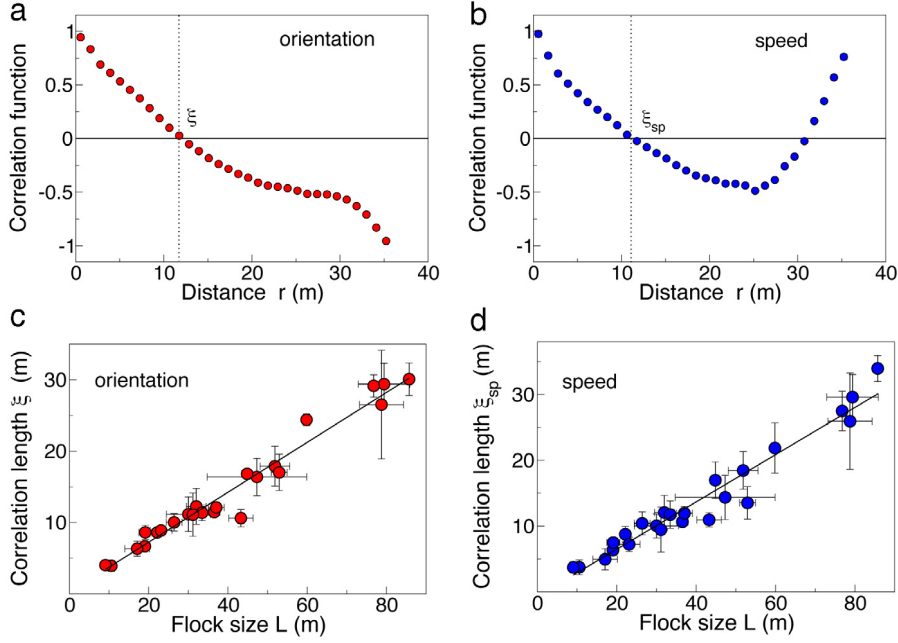
Fig. 2. Velocity correlation functions in natural flocks of birds. (a) Space correlation of the velocity fluctuations at a given instant of time, for the same flocking event as in Fig. 1. (b) Space correlation of the speed fluctuations for the same flocking event and for the same time as in (a). Both velocity and speed connected correlations are normalized to 1 for $r = 0$. (c) Correlation length of velocity fluctuations vs. system size for all the flocking events analyzed in [10]. (d) Correlation length of speed fluctuations vs. system size, for the same events as in (c) [10]. The density is different in different flocks, with a mean nearest-neighbor distance r_1 varying in the interval [0.6 m : 1.5 m]; the topological interaction range in units of birds is $n_c \sim 8$ neighbors [18,19], corresponding to a metric range $r_c \sim r_1 n_c^{1/3}$ varying in the interval [1m : 3.5m]. The correlation function therefore extends much beyond the interaction range, both in units of meters and in units of birds.

average correlation length of several flocks as a function of the flock's linear dimension, for sizes up to 4000 individuals. The plot shows that ξ increases linearly with the size L of the group. This means that there is no typical correlation length for the decay of correlations; rather, the only scale present in the system is its size.

2.2.1. Scale-free correlations

The fact that the correlation length is proportional to the system's size means that the correlations are scale-free in the thermodynamic limit: since the only characteristic length scale is the system size itself, when it becomes infinite there remains no characteristic length scale. Usually scale-free behavior is associated to a power law decay. Let us therefore show how the result $\xi \sim L$ implies a power-law decay in the thermodynamic limit. The leading contribution to the correlation function $C(r)$ can in general be written, when ξ is much larger than the microscopic scales such as the interparticle separation, a, as [20]

$$C(r) = \frac{1}{r^\gamma} f\left(\frac{r}{\xi}\right) \tag{11}$$

where $f(x)$ is a scaling function.[1] By multiplying and dividing by ξ^γ, one gets an equivalent form,

$$C(r) = \frac{1}{\xi^\gamma} g\left(\frac{r}{\xi}\right), \tag{12}$$

where $g(x)$ is another scaling function. The experimental finding that ξ grows proportional to the flock's size L can be formalized as,

$$\xi(L) = s\, \xi(L/s), \tag{13}$$

where $s > 1$ is a dimensionless scaling factor; that (13) makes sense can be seen by choosing the arbitrary scaling factor as $s = L/a$, where a is the natural microscopic length scale (as the interparticle separation), in which case we get the scale-free result, $\xi(L) \sim L$, up to irrelevant microscopic constants. By substituting (13) into (12) we get,

$$C(r; L) = \frac{1}{s^\gamma} C(r/s; L/s), \tag{14}$$

---

[1] In Eq. (11) one considers r much larger than the microscopic scale a; in fact, $C(r)$ has a finite limit for $r \to 0$.
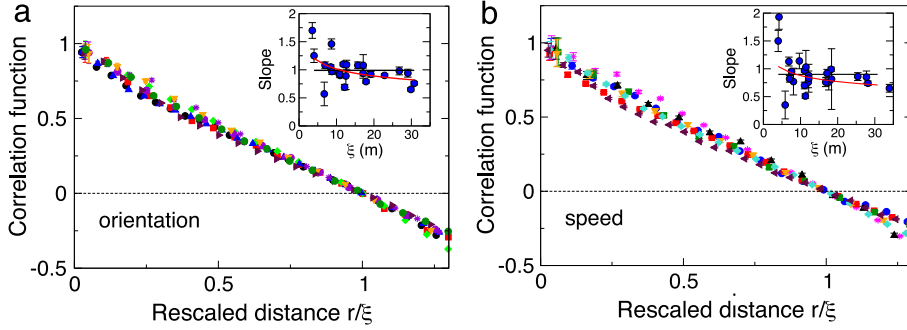
Fig. 3. Rescaling of the correlation functions. (a) The correlation function of the velocity fluctuations $C(r)$ is plotted as a function of the rescaled variable $x = r/\xi$. Different colors correspond to different flocking events. Inset: slope of the function $C'(x = 1)$ vs. the correlation length. Each point corresponds to a different flocking event; error bars are standard deviations across multiple times in the same flocking sequence. (b) Same as in (a) for the speed correlation function.
Source: From [10].

which tells us that the correlation function scales homogeneously under a rescaling of space, $r \rightarrow r/s$. We can now chose the arbitrary scaling factor as $s = r/a$, which is equivalent to measuring all distances in units of the microscopic sale a, hence integrating over all short wavelengths details (see Sec. 3.2 of [21] for a discussion of this point); in this way we obtain,

$$C(r; L) = \frac{1}{r^\gamma} h\left(\frac{r}{L}\right) , \qquad (15)$$

where $h(x)$ is yet another scaling function. This relation tells us that the correlation function becomes a pure power law for infinitely large systems, $L \rightarrow \infty$; on the other hand, at finite values of L, $C(r)$ is modulated by the scaling function, whose only scale is L itself. This is precisely what is found in flocks. We note that the behavior of the scaling function $h(x)$ also depends on the way one defines fluctuations: in our case, where we subtract to the fields the instantaneous space average, h has a zero at some value of its argument, say x_0 (see discussion in Section 2.3.2). As a function of r, the zero will then be at $r_0 = Lx_0$, consistent with the fact that for $L \rightarrow \infty$ the correlation becomes more similar to a power law. This is also in agreement with the statement that the estimate of ξ (the point where the correlation becomes negative) is proportional to L.

2.2.2. Scaling of the correlations

From Eq. (15) we see that the product $r^\gamma C(r; L)$ is a function of the scaling variable r/L, so that curves for systems at different sizes must collapse one on top of each other. We do not know the value of the exponent γ a priori, but we can extract it from the data. We can try different values of γ and search for that which optimizes the collapse. Notice that, because of the scale-free relation $\xi \sim L$, using r/L or r/ξ as a scaling variable is exactly the same thing. A word of caution about this method: it is important to notice that it is the product $r^\gamma C(r; L)$ and *not* simply the correlation $C(r, L)$, that collapses when plotted against the scaling variable r/L; if, on the other hand, we plot directly $C(r)$ vs $x = r/L$, we find $C(x) = 1/L^\gamma \hat{h}(x)$, hence, if $\gamma \neq 0$, one sees a flattening of the function for larger systems.

If we want to estimate γ more constructively, without using the collapse, we can plot $C(r; L)$ as a function of $x = r/\xi$ (equivalent to r/L in our scale-free case) and evaluate its derivative in $x = 1$. Indeed, from Eq. (15) one has $(dC(x)/dx)|_{x=1} = g(1)/\xi^\gamma$: one can therefore retrieve γ by looking at the slope of $C(x)$ at the crossing point. This procedure was followed in [10], where the exponent has been estimated as $\gamma \sim 0.19$ (see Fig. 3a). This very small value implies that correlations collapse quite well also when plotting them as a function of r/ξ without the multiplicative factor r^γ, as it is clearly seen from the figure. It also indicates that velocity correlations are not only scale-free, but also decay very slowly.

2.2.3. Speed correlations

Interestingly, a similar scale-free behavior is observed also for the fluctuations of the individual speeds. Analogously to the velocity correlation (6), we define the speed connected correlation function

$$C_{sp}(r) = \frac{\sum_{i,j}^N \delta s_i \, \delta s_j \, \delta(r - r_{ij})}{\sum_{i,j}^N \delta(r - r_{ij})}, \qquad (16)$$

where, we remind, $\delta s_i = |\mathbf{v}_i| - S$. The speed correlation function for a flock at a given instant of time is displayed in Fig. 2b. Again, as in the velocity the decay is not exponential, and the correlation length ξ_s is estimated as the distance where the correlation crosses the x-axis. The average correlation length as a function of the flock's size L is plotted in Fig. 2d: in this case too the correlation length scales linearly with the size, indicating that speed correlations are also scale-free. A scaling analysis of the same kind of the one discussed above for the velocity correlation functions shows that also for the speed correlations decay very slowly, with an exponent of the same order as the velocity one (see Fig. 3b).

2.2.4. Origins of the scale-free behavior

The presence of scale-free correlations in different degrees of freedom indicate that individuals in a flock are able to influence one another even at large distances, whatever the size of the group is. In this respect, they quantify the intuitive idea that these groups act and respond coherently. However, when it comes to understanding the microscopic mechanism giving rise to such correlations, we realize that scale-free correlations in the directional degrees of freedom and in the speed might require quite different explanations.

All models of flocking and polar active motion invoke interactions between individuals based on short-range mutual alignment [5,22,23,2,24], much as spins in a Heisenberg ferromagnet. We will see that analysis of real flock data do indeed support this assumption. Flocks are therefore systems endowed with a natural continuous symmetry – the rotation of the individual velocities – which happen to live in the polarized phase. We know from the physics of equilibrium systems that whenever a continuous symmetry is spontaneously broken, giving rise to global ordering, there are some fluctuation modes (those perpendicular to the global order parameter) that are scale-free (Goldstone's theorem [16,15]). The presence of such soft modes due to a symmetry breaking also holds in off-equilibrium models and has been verified numerically and analytically for a variety of models of active matter [7,2,25,26]. Thus, the scale-free correlations of the velocity fluctuations found in natural flocks seem a natural manifestation of the spontaneous breaking of the rotational symmetry. From this point of view they are generic. They are the consequence of the nature of interactions, and of the good accuracy with which individuals follow their behavioral rules (i.e. low noise). They occur contextually to global order: as soon as the group develops collective motion, there are some 'easy' fluctuations that cost very little (they do not affect the degree of order) and extend on the group scale. No specific values of the parameters regulating the system's behavior are required, just that the system is in an ordered state.

The situation is completely different if we consider the speed correlations. Speed is a scalar degree of freedom, and there is no continuous symmetry associated to it. Contrary to flight directions, speeds are stiff quantities, costly to change. In this case, therefore, the mechanism leading to the observed long-range correlations must be different. In equilibrium statistical mechanics, the only other way to produce scale-free static correlations is to bring the system close to a critical point. Pushing the analogy to flocks, the experimental findings therefore suggest that natural flocks might exhibit some kind of critical behavior. In Section 2.4 we will discuss the possible origin of such criticality. For now, we observe that such behavior – contrary to the velocity case – is highly non-generic and requires the fine-tuning of some relevant parameter close to a specific 'critical' value.

2.3. Crucial caveats (this is not an appendix)

We now need to discuss a few subtle and yet crucial issues related to the definition and interpretation of correlation functions in real biological systems; in doing so we will introduce some other quantities that play an important role in characterizing – both experimentally and theoretically – collective behavior.

2.3.1. Space averages vs phase averages

The connected correlation defined in Eq. (6) refers to orientational degrees of freedom, the velocities. In this respect it is similar in spirit to the correlation usually defined in the theory of ferromagnetic systems. There are however some interesting differences in the definition, which reflect the more complex nature of the system we had like to describe. Had the $\{\mathbf{v}_i\}$ represented the spin vectors of a ferromagnet, we would have defined the connected correlation as $C_{ij} = \langle \delta\mathbf{v}_i \cdot \delta\mathbf{v}_j \rangle$, where the fluctuation is measured from the phase average, $\delta\mathbf{v}_i = \mathbf{v}_i - \langle\mathbf{v}_i\rangle$, and the phase averages $\langle \cdots \rangle$ are taken over the stationary distribution of the $\{\mathbf{v}_i\}$. While this procedure is perfectly well-defined for an equilibrium system on a lattice, it can become practically and conceptually problematic for an active system. As we already noticed, when we are confronted with data from flocks or other living systems we do not know a priori what the phase average is. If we want to perform averages and get statistically robust quantities, the only reasonable choice is to perform space averages, i.e. averages over all the group. This is indeed what we did both at the level of fluctuations (defined as deviations from the group mean — see Eq. (4)) and at the level of the correlation function (defined as an average over the group of correlations pairs — see Eq. (6)). Additionally, active systems do not live on a fixed lattice, individuals exchange positions and gradually diffuse far away from one another. In this respect, averaging the C_{ij} for a given pair is in general not meaningful: sooner or later the correlation becomes zero and it does not have, strictly speaking, a stationary distribution. The correlation function $C(r)$ however overcomes this problem: it is an average over space of a quantity that depends on mutual distance (rather than the individuals' identities). It is reasonable to expect that this quantity has a well-defined stationary distribution and that, if the system is large enough, Eq. (6) gives a good estimate of the statistical behavior of this quantity.

2.3.2. The spatial constraint

The definition of $C(r)$ given in Eq. (6) involves fluctuations with respect to the instantaneous space average, rather than a phase (or ensemble) average, which would be the reasonable choice in equilibrium statistical mechanics, but that is here unavailable. As a consequence, the correlation function obeys some specific constraints and behaves slightly differently from what we are used to in equilibrium condensed matter systems. From the definition of Eq. (4), it follows that if we sum over all the individuals we get,

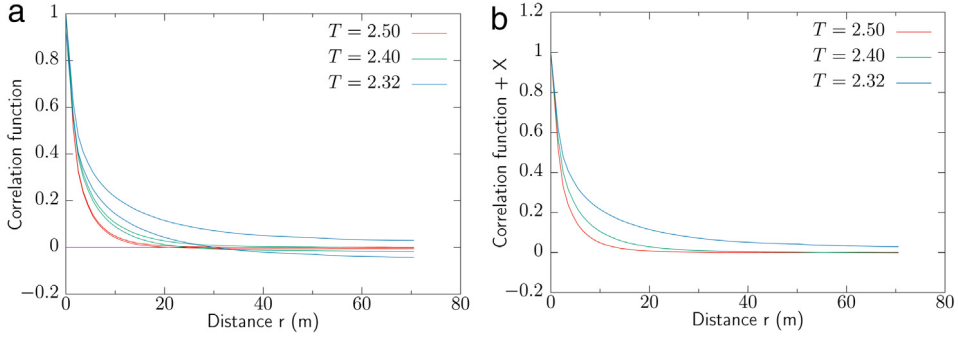$$\sum_i \delta\mathbf{v}_i = 0. \tag{17}$$

Fig. 4. Static correlation function. Numerical simulation of the Ising model in $d = 2$ with periodic boundary conditions in the disordered phase. Curves of the same color correspond to the same temperature. (a) The phase averaged correlation function $C^{\mathrm{ph}}(r)$ is computed by using phase averages, while the correlation $C(r)$ defined in (6) is instead computed by using the instantaneous space average. $C(r)$ always lies below $C^{\mathrm{ph}}(r)$. (b) Same curves as in panel (a), but we now plot $C(r) + \langle[\mathbf{V} - \langle\mathbf{V}\rangle]^2\rangle$ instead of $C(r)$: we see that the curves computed at the same temperature now lie one on top of each other. The critical temperature is $T_c = 2.27$ [27]. The system's linear size is $L = 100$. Correlations are normalized in zero for a better comparison.

This spatial constraint has a key consequences on the behavior of $C(r)$; from Eq. (6), (10), we have,

$$0 = \frac{1}{N}\sum_{i,j}\delta\mathbf{v}_i \cdot \delta\mathbf{v}_j = \rho_0 \int d\mathbf{r}\, g(r)C(r). \tag{18}$$

Since $g(r)$ is positive, this relation implies that $C(r)$ must have a zero. Hence the correlation function will not asymptotically decay to zero from above at large distances, as it does when defined in the standard way (i.e. subtracting the phase average), but must cross the axis and become negative. Since the space average $\mathbf{V}$ approaches the phase average $\langle\mathbf{V}\rangle$ in the thermodynamic limit, we expect $C(r)$ to correctly describe the typical decay of correlations for large systems. At a given size, if the decay rate of the correlation is much smaller than the system's size, the effect of the constraint is barely visible. On the contrary, if correlations extend over the whole system the constraint forces the correlation function to become negative in a region where correlations are still strong (we will see in the next section how to exploit this fact). To illustrate this point let us consider a case where we know how to compute phase averages and can therefore compare the correlation $C(r)$ defined in (6) (where one subtracts the space average) with the standard one where one uses the phase average, $C^{\mathrm{ph}}(r)$: this case is the ferromagnet. In Fig. 4 we display the behavior of the two correlation functions in the disordered phase of the $2d$ Ising model. We can see that for temperatures well above the critical temperature, the exponential decay behavior is the same and is clearly visible in both correlations; when T approaches T_c the standard connected correlation function approaches a power law behavior, while $C(r)$ has a zero due to the constraint. We note that, since $\delta\mathbf{v}_i = \mathbf{v}_i - \mathbf{V} = \mathbf{v}_i - \langle\mathbf{V}\rangle + \langle\mathbf{V}\rangle - \mathbf{V}$, we easily get

$$C(r) = C^{\mathrm{ph}}(r) - \langle[\mathbf{V} - \langle\mathbf{V}\rangle]^2\rangle . \tag{19}$$

This expression shows that the two correlation functions differ by a constant, which measures the average fluctuations of the global order parameter (we shall see in the next Section that this constant is essentially the susceptibility). At equilibrium such fluctuations decrease with increasing the system size as $1/N$ (unless we are close to a critical point), showing that in the thermodynamic limit the two correlations come close one to each other. In the simple example we are considering ($2d$ Ising model), we can numerically verify Eq. (19) (see Fig. 4b).

2.3.3. Subtleties about the correlation length

The role of the correlation length is to measure the extension of the correlated domains, i.e. the distance below which the individual deviations from the mean fluctuate in a correlated way. As usual, we expect the correlation to be strong at small distances and then fade away, which should be captured by the decay of $C(r)$ as r increases. However, $C(r)$ must cross the zero axis at some point r_0 due to the constraint. How can we then define ξ in an unambiguous way? Let us imagine that $C(r)$ exhibits an exponential decay well before the crossing point (as in the red curve of Fig. 4): in this case the most natural thing to do is to define ξ as the range of the exponential decay. Correlations become indeed negligible beyond that distance; the zero crossing happens due to a mathematical constraint, but correlations are already vanishing at that scale. The situation is different, however, when no exponential behavior is present: in this case correlations are non-zero for all distances and the only obvious scale that the correlation exhibits is the crossing point itself (as is the case for natural flocks of birds discussed in Section 2.2). We shall therefore adopt the following definition:

$$\text{exponential decay} \qquad \longrightarrow \qquad \xi : \quad C(\xi) \sim \frac{1}{e}, \tag{20}$$

$$\text{no exponential decay} \qquad \longrightarrow \qquad \xi : \quad C(\xi) = 0 \quad (\text{i.e. } \xi = r_0) \tag{21}$$

When correlated domains are small with respect to the size of the system, at any given instant of time there are many correlated regions within the system each one with coherent fluctuations along some direction. When considering pairs at distances larger than the domains' extension we count correlations between different regions: since such regions are many and uncorrelated their overall contribution is very small and one gets a correlation function decaying exponentially and being negligible beyond the domains typical size. However, when the size of the domains starts becoming comparable to the system's size this is no longer true. There will be few domains within the system: when considering pairs with distances larger than the domain size ξ, the correlation that we get describes the mutual orientation of these few domains, and it does not average to zero. In particular, the largest possible correlated domains must be two and have opposite orientations to satisfy the constraint $\sum_i \delta \mathbf{v}_i = 0$, in which case the correlation would be positive for $r < \xi$ and negative for $r > \xi$; this scenario can be quantitatively verified using random synthetic velocities chosen to be correlated on a given scale, and computing the correlation function (6) [10]. We note in this respect that the presence of anti-correlations is very different from the absence of correlations

To fully appreciate the difference between the crossing point r_0 and the correlation length ξ, and understand when the first can be used as a proxy for the second, it is convenient once again to consider the reference case of ferromagnetic systems. Looking back at Eq. (19), we notice that,

$$N\langle[\mathbf{V} - \langle\mathbf{V}\rangle]^2\rangle = (1/N)\sum_{i,j}\langle(\mathbf{v}_i - \langle\mathbf{V}\rangle)(\mathbf{v}_j - \langle\mathbf{V}\rangle)\rangle = \rho_0 \int d\mathbf{r}\, g(r) C^{\mathrm{ph}}(r)\,, \tag{22}$$

where $C^{\mathrm{ph}}(r)$ is the correlation function defined with the phase average; from this we get,

$$C(r) = C^{\mathrm{ph}}(r) - \frac{\rho_0}{N}\int d\mathbf{r}'\, g(r') C^{\mathrm{ph}}(r')\,. \tag{23}$$

From this equation we can obtain an explicit expression for r_0, the crossing point where $C = 0$, i.e.

$$C^{\mathrm{ph}}(r_0) = \frac{\rho_0}{N}\int d\mathbf{r}\, g(r) C^{\mathrm{ph}}(r)\,, \tag{24}$$

where the integral is extended up to the system's size, L. Quite in general, we can assume that the correlation function $C^{\mathrm{ph}}(r)$ is the product of a power-law (scale-free) part and of an exponential part (or any other short-range function) [15,16],

$$C^{\mathrm{ph}}(r) = \left(\frac{a}{r}\right)^\gamma \exp(-r/\xi_{\mathrm{exp}})\,, \tag{25}$$

where a is the lattice spacing, γ the degree of the power law, and ξ_{exp} is the correlation length defined by the exponential decay. If we measure the correlation over distances much larger than a, then the system is homogeneous and $g(r) \sim 1$. Putting all this back into Eq. (24) we can solve for r_0. The results depend on how large ξ_{exp} compared to the system's size L. There are two cases:

(i) $\xi_{\mathrm{exp}} \ll L$. In this case we get, to leading order,

$$r_0 \sim \xi_{\mathrm{exp}} \log\left(\frac{L}{\xi_{\mathrm{exp}}}\right)\,, \tag{26}$$

where L is the linear size of the system. We therefore see that, if ξ_{exp} is finite and small compared to the system's size, r_0 increases with ξ_{exp} but has a logarithmic dependence on the system's size L too. In this case r_0 is not such a good proxy of ξ_{exp} and one is better off by fitting the exponential decay of the correlation.

(ii) $\xi_{\mathrm{exp}} \gg L$. In this case $C^{\mathrm{ph}}(r)$ effectively decays as a power law,

$$C^{\mathrm{ph}}(r) \sim \left(\frac{a}{r}\right)^\gamma\,. \tag{27}$$

In this case there is no well-defined length scale in the systems apart from L itself, which is why we call this case scale-free. When this happens we cannot strictly speak of a 'correlation length' because there is no such thing. Indeed, given L and with $\xi_{\mathrm{exp}} \gg L$, we cannot distinguish between a correlation length finite but much larger than L and a truly infinite correlation length. In this case Eq. (24) gives,

$$r_0 \sim L\,, \tag{28}$$

which is quite faithful to what is actually happening: the system has no intrinsic scale apart from L, hence the only scale that we were able to practically compute, r_0, grows with L itself. This is why in real flocks we used r_0 as a marker of scale-free correlation. Note that sometimes one says that in the scale-free case the correlation length grows with the system size; there is a slight abuse of language in that, which can be forgiven if by 'correlation length' one does not mean the decay rate of an underlying exponential correlation function, but simply the only length scale experimentally available.

2.3.4. Susceptibility

The correlation function $C(r)$ measures the degree of pair correlations as a function of distance. Another quantity that is very useful to characterize collective behavior of the group is the *total* degree of correlation present in the system. In equilibrium statistical physics this quantity is called susceptibility, i.e. the statistical average of the sum of all the fluctuating

pairs [28],

$$\chi^{\mathrm{ph}} = \frac{1}{N} \sum_{i,j} \langle \delta \mathbf{v}_i \cdot \delta \mathbf{v}_j \rangle \,, \qquad (29)$$

where again the superscript stresses that this is a phase average. In equilibrium systems, χ^{ph} is related through the fluctuation–dissipation theorem to the linear response of the system to an external field conjugate to the $\mathbf{v}_i$ (Appendix C). When correlations are strong and long-range, for example at a critical point, fluctuations are correlated over all scales, the degree of correlation is maximal, and the system responds collectively to any external field, corresponding to a susceptibility that grows with the system's size.

Although we do not have the luxury of the fluctuation–dissipation theorem in biological systems (at least not yet), we would like to introduce a quantity, χ, as close as possible to the phase susceptibility, χ^{ph}, also for flocks, swarms and other living systems on the move, where phase averages are not available and we can only use fluctuations with respect to the instantaneous group mean and compute spatial averages. To do this, however, we have to be careful, because the spatial constraint we have discussed in the previous Section prevents us from merely replicating relation (29), given that,

$$\sum_{i,j} \delta \mathbf{v}_i \cdot \delta \mathbf{v}_j = \sum_i \delta \mathbf{v}_i \cdot \sum_j \delta \mathbf{v}_j = 0 \,. \qquad (30)$$

We therefore need to define χ in a smarter way. The aim of the susceptibility is to quantify the total amount of correlation present in the system by integrating (i.e. summing) fluctuations over space. We have seen that the correlation length marks the typical size of the correlated domains; therefore, if we sum pairs of fluctuations up to distances $r < \xi$ we expect the integrated correlation to increase. On the other hand, when we consider pairs with $r > \xi$, we are no longer adding *bona fide* correlation; moreover, due to the spatial constraint, the integrated correlation starts decreasing, eventually becoming zero when we integrate over the entire system, Eq. (30). This argument suggests defining χ as the maximum of the integrated correlation,

$$\chi = \frac{1}{N} \sum_{ij:r_{ij}<\xi} \delta \mathbf{v}_i \cdot \delta \mathbf{v}_j \,, \qquad (31)$$

where we sum all pairs up to the correlation length. This definition of χ captures the degree of correlation in the system when fluctuations are defined with respect to the instantaneous average. Besides, one can check that in standard equilibrium systems (e.g. in the same model of Fig. 4) it behaves asymptotically in the same way as the standard definition of susceptibility with phase average, Eq. (29). We also note that Eq. (29) implies,

$$\chi^{\mathrm{ph}} = N \left[\langle \mathbf{V}^2 \rangle - \langle \mathbf{V} \rangle^2 \right] \,, \qquad (32)$$

where, $\mathbf{V} = (1/N)\sum_i \mathbf{v}_i$. Hence the susceptibility is also equal to the (phase) fluctuation of the global order parameter [28]. Interestingly, this shows that the constant giving the difference between phase-averaged and space-averaged correlation functions in (19) is indeed the susceptibility. In principle, one could compute the susceptibility by using (32) substituting the phase average with the time average of the variance of the order parameter. In practice, this may work in numerical simulations, in which one has time sequences as large as one wants; in experiments, on the other hand, one is very often limited in the time duration of the acquired sequences, hence using (32) is not convenient in most cases. The best thing to do is to compute the susceptibility χ at each time frame by using (31) and then, if the experiment allows for such bonanza, to further average χ over the available time frames.

In some cases it might be convenient to define the susceptibility by using normalized dimensionless fluctuations,

$$\delta \hat{\mathbf{v}}_i \equiv \frac{\delta \mathbf{v}_i}{\sqrt{\frac{1}{N} \sum_k \delta \mathbf{v}_k \cdot \delta \mathbf{v}_k}} \qquad (33)$$

rather than the standard ones, Eq. (4). This might be useful for two reasons; first, dimensionless fluctuations allow to compare biological systems to numerical simulations, which have arbitrary physical dimensions. secondly, even in absence of a comparison with numerical simulations, sometimes one needs to analyze at the same time biological systems with the same physical units, but different scales (typically of velocity); by using dimensionless fluctuations one can do this in a homogeneous way, as it was done for insect swarms in [12] and [11]. Note, however, that using dimensionless fluctuations to compute the susceptibility has a drawback: the normalizing factor in (33) depends on the polarization of the system. For example, in a system where each velocity has constant modulus, $|\mathbf{v}_i| = v_0$, the susceptibility $\hat{\chi}$ calculated by using the dimensionless fluctuations $\delta \hat{\mathbf{v}}_i$ is connected to the standard susceptibility χ by the relation,

$$\hat{\chi} = \frac{\chi}{v_0^2 (1 - \Phi^2)} \,. \qquad (34)$$

If the system is in the unpolarized phase (as the swarms analyzed in [12,11]), then $\Phi \sim 0$ and the two susceptibilities are indeed very similar; if, on the other hand, the system is polarized, as it is clearly the case for flocks, then both the numerator

and the denominator at the r.h.s. of (34) decrease by entering in the ordered phase, thus giving an awkward susceptibility that may remain constant even for $\Phi \to 1$. For systems where the speed is not constant the relation is more complicated than (34), but the physics is the same, namely the normalization depends on the global degree of order of the system, which is in general very inconvenient. We therefore urge the reader *not* to use the dimensionless normalized fluctuations (33) to compute the susceptibility, unless one is sure that the system in unpolarized.

2.3.5. Correlation function in Fourier space

Sometimes it is useful to look at the correlation function in Fourier space, rather than in real space. As we shall discuss at the end of this work, this is indeed the case when extending the definition of correlation to its time-dependent version. We define the Fourier-space equal-time correlation as,

$$C(\mathbf{k}) = \frac{1}{N} \sum_{i,j}^{N} \delta\mathbf{v}_i \cdot \delta\mathbf{v}_j e^{i\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)} . \tag{35}$$

If the system is isotropic, after averaging over the polar angles of $\mathbf{k}$ we get,

$$C(k) = \frac{1}{N} \sum_{i,j}^{N} \int_{-1}^{+1} d(\cos\theta) e^{ikr_{ij}\cos(\theta)} \delta\mathbf{v}_i \cdot \delta\mathbf{v}_j$$

$$= \frac{1}{N} \sum_{i,j}^{N} \frac{\sin kr_{ij}}{kr_{ij}} \delta\mathbf{v}_i \cdot \delta\mathbf{v}_j. \tag{36}$$

If we consider the definition of the space-dependent correlation given in Eq. (6), we see that the relationship between $C(k)$ and $C(r)$ is

$$C(k) = \rho_0 \int d\mathbf{r}\, g(r)\, C(r)\, e^{-i\mathbf{k}\cdot\mathbf{r}} . \tag{37}$$

The presence of the radial distribution function $g(r)$ (defined in (9)) in the integral may seem strange and one could wonder why we did not define the $C(k)$ directly as the Fourier transform of the $C(r)$. The reason is that in defining $C(k)$ we want to preserve an important equation of statistical field theory, namely the relation between susceptibility and correlation at $k = 0$. To see this, let us momentarily work again with phase averages, so to emancipate from the usual sum-rule of space averages. If we set $k = 0$ in Eq. (37) and use (10), or simply from (36), we obtain,

$$C(k = 0) = \frac{1}{N} \sum_{i,j} \langle \delta\mathbf{v}_i \cdot \delta\mathbf{v}_j \rangle = \chi^{\text{ph}} , \tag{38}$$

which is the proper relationship between correlation in Fourier space and susceptibility [28]; (38) is also quite convenient when it comes to actually compute the susceptibility, because one single function, $C(k)$, contains all the information about both the spatial correlation and its total integral. Another way to understand Eq. (37) is to remember that we want to focus exclusively on the correlation and susceptibility of the order parameter, not of the density fluctuations. If instead of using (37), we performed the Fourier transform of $C(r)$ without the factor $g(r)$ in the integral, we would get,

$$\hat{C}(k = 0) = \frac{1}{N} \sum_{i,j} \langle \delta\mathbf{v}_i \cdot \delta\mathbf{v}_j \rangle \frac{1}{g(r_{ij})} , \tag{39}$$

which is certainly *not* the order parameter susceptibility, as each fluctuating pair is now weighted by the inverse of the density correlation, $g(r_{ij})$. In this respect, the term $d\mathbf{r}\, g(r)$ in (37) can be interpreted as an integration measure $d\mu(\mathbf{r})$ that makes the integration insensitive to the density fluctuations of the system (see also Appendix A). We notice that this prescription has nothing to do with the off-equilibrium active nature of biological systems; it is something we would sensibly do even for a standard statistical mechanics model on a regular fixed lattice in order to fulfill (38).

In the active biological case, where we have space rather than phase averages, we need as usual to be careful, lest we obtain a trivial result. Indeed, because of the spatial constraint, if we evaluate $C(k)$ at $k = 0$ we get a trivial result,

$$C(k = 0) = \frac{1}{N} \sum_{i,j} \delta\mathbf{v}_i \cdot \delta\mathbf{v}_j = 0 . \tag{40}$$

There is however a natural way to link the correlation in Fourier space with the susceptibility χ defined in Eq. (31). To see this, we notice that $C(k)$ provides an alternative route to computing ξ. If we compute $C(k)$ at large k and then slowly decrease it, we will be averaging over progressively larger length scales, therefore adding to (36) more correlated pairs, causing $C(k)$ to increase. When the momentum arrives at $k_{\max} \sim 1/\xi$, we start adding uncorrelated pairs, hence, $C(k)$ must level. Further decreasing k down to $1/L$ (where L is the system's size) we start to be affected by the spatial constraint, $C(k = 0) = 0$, hence $C(k)$ decreases until it eventually vanishes for $k = 0$ [29]. In a system where $\xi \ll L$ the static correlation therefore has – in log scale – a broad plateau between $k_{\max} \sim 1/\xi$ and $k \sim 1/L$. Hence, to compute the susceptibility χ from the $C(k)$ we must

not look at too small values of k, which would include correlated and non-correlated pairs, but rather add up only the pairs in the correlated domains. This is what we get by setting $k = k_{\max}$, because then the oscillating term in Eq. (35) will weigh only pairs with $r_{ij} < \xi$, hence giving,

$$C(k = k_{\max}) \sim \frac{1}{N} \sum_{ij:r_{ij}<\xi} \delta\mathbf{v}_i \cdot \delta\mathbf{v}_j = \chi \ . \tag{41}$$

Therefore, to preserve the natural relation between correlation function in Fourier space and susceptibility, even in the presence of the spatial constraint dictated by space averages, we simply evaluate $C(k)$ at its maximum; this is quite reasonable and in a way it is also what we do in a standard system with phase averages, where the maximum is achieved at $k = 0$. The effect of the space average is simply to shift this maximum at a larger value of k, but we still evaluate the Fourier correlation at its maximum.

2.4. The maximum entropy approach to flocks

In the previous sections we introduced the correlation functions of velocities and speeds, explained their definitions and described their behavior in natural flocks of birds. In doing so, we often resorted to the analogy between flocks and standard systems of statistical mechanics, or mentioned results obtained in models of active matter. Given the experimental data at hand, however, one would like to do better: can we exploit the knowledge of the empirical correlation functions to say something about the microscopic inter-individual mechanisms producing them? This question boils down to a problem of statistical inference. An approach that has been successfully applied to a variety of biological networks is the *maximum entropy* (ME) method. ME provides a systematic framework to infer statistical models from experimental data. Given a set of measured quantities $\{O_\mu(X)\}$, function of the microscopic degrees of freedom X of the system, the *maximum entropy* model for the statistical distribution $P(X)$ of the X, is the least structured model consistent with the experimental measurements. The prescription for "least structured" is that the entropy associated to the model, $S[P(X)] = -\sum_X P(X) \ln P(X)$, be the maximum with the constraint that the expected values of the $\{O_\mu(X)\}$ must be equal to the experimental results. Mathematically, one introduces a generalized entropy function

$$\mathcal{S}[P(X), \{\lambda_\mu\}] = -\sum_X P(X) \ln P(X) - \sum_\mu \lambda_\mu \left[\langle O_\mu(X) \rangle - \langle O_\mu(X) \rangle_{\exp} \right], \tag{42}$$

where $\langle \cdots \rangle = \sum_X P(X)(\cdots)$, $\langle \cdots \rangle_{\exp}$ are the experimental values, and $\{\lambda_\mu\}$ are a set of Lagrange multipliers. Maximization of $\mathcal{S}$ with respect to $P(X)$ ensures maximal entropy, while optimization with respect to the $\{\lambda_\mu\}$ ensures that the model's predictions for the $\{O_\mu(X)\}$ are equal to the experimental values. The resulting distribution has a Boltzmann form [30,31]

$$P(X) = \frac{1}{Z(\{\lambda_\mu\})} \exp\{-\sum_\mu \lambda_\mu O_\mu(X)\}, \tag{43}$$

where the values of the $\{\lambda_\mu\}$ are fixed by the equations

$$\langle O_\mu(X) \rangle = \langle O_\mu(X) \rangle_{\exp}, \tag{44}$$

or, equivalently, by substituting expression (43) into (42) and optimizing the generalized entropy $\mathcal{S}(\{\lambda_\mu\}) = \ln Z(\{\lambda_\mu\}) + \sum_\mu \lambda_\mu \langle O_\mu(X) \rangle_{\exp}$ with respect to the Lagrange parameters.

The ME model given by Eqs. (43) and (44) is an effective model: it gives the minimal description of the system that correctly reproduces the statistics of a given set of observables. Clearly, the larger the number of the input observables, the more detailed the resulting ME model. Still, using a large set of $\{O_\mu(X)\}$ is both unrealistic and counterproductive. On the one hand, it is very difficult to obtain experimentally robust measurements of many independent quantities (otherwise one could try to directly build the $P(X)$ from the data). On the other hand, a model with too many parameters is not very informative and might be redundant. A viable strategy is to start by considering ME models built on a small number of well-measured quantities. If this minimal ME model also provides good predictions for other quantities not included in the constraints, then it captures some important ingredient of the real behavior and can be considered as a relevant – though approximate – effective description of the system. If it does not, then one needs to progressively include further experimental information and develop more complex ME models.

2.4.1. The simplest ME model for flocks

Applied to empirical data of natural flocks, the aim of the ME approach is to build an effective model able to explain the statistics of the velocities and the emergence of scale-free correlations. To proceed we need a non-trivial quantity that can be robustly measured in an experiment. Since we suspect that scale-free correlations in the orientations are the consequence of symmetry breaking while some kind of critical point might be related to the speed behavior, we start by focusing on the

directional degrees of freedom. With this spirit, in [32] a ME model was built using as input experimental observable the directional correlation between a bird and its first n_c neighbors,

$$C_{\text{int}} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{n_c} \sum_{j \in n_c^i} \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j. \tag{45}$$

where $\boldsymbol{\sigma}_i = \mathbf{v}_i / |\mathbf{v}_i|$ is the individual flight direction and n_c^i is the set of the first n_c nearest neighbors of bird i. C_{int} is a local scalar quantity, averaged over all the group; it can therefore be measured in a robust way even at a single instant of time. The resulting ME model has a very simple form [32]

$$P(\{\boldsymbol{\sigma}_i\}) = \frac{1}{Z(J, n_c)} \exp\left[\frac{J}{2} \sum_{i,j=1}^{N} n_{ij} \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j\right], \tag{46}$$

where the adjacency matrix n_{ij} is 1 if j is one of the first n_c neighbors of i or vice-versa, and zero otherwise. J is the Lagrange multiplier associated to C_{int}, and its value is fixed by the condition

$$\langle C_{\text{int}} \rangle = \langle C_{\text{int}} \rangle_{\text{exp}} \tag{47}$$

The parameter n_c can also be fixed following a maximum likelihood criterion and optimizing the generalized entropy (42) with respect to n_c. The distribution (46) corresponds to a Heisenberg model with short range interactions over the scale n_c, and alignment strength J. In general, inferring the parameters of the model can be a hard problem. In many cases one resorts to numerical simulations to do this, but in the case of flocks, the computation can be done analytically in the spin-wave approximation (see Appendix B).

We can write each velocity vector as the sum of a longitudinal and a perpendicular component: $\mathbf{v}_i = v_i^l \mathbf{n} + \boldsymbol{\pi}_i$, where $\mathbf{n}$ represents the direction of collective motion. Disregarding speed fluctuations, we have $|\mathbf{v}_i| = v_0$ and $\boldsymbol{\sigma}_i = \sigma_i^L \mathbf{n} + \boldsymbol{\pi}_i / v_0$. In the highly polarized phase the $|\boldsymbol{\pi}_i|$ are very small, then $\sigma_i^L \sim 1 - (1/2)\pi_i^2/v_0^2$ and $\boldsymbol{\sigma}_i = \mathbf{n}(1 - (1/2)(\pi_i^2/v_0^2)) + \boldsymbol{\pi}_i / v_0$. With this approximation, the partition function (46) can be computed by integrating out the longitudinal degrees of freedom [32] to find at leading order in $\boldsymbol{\pi}$,

$$Z(J, n_c) = \int d^N \pi \, \delta\left(\sum_i \boldsymbol{\pi}_i\right) \exp\left[-\frac{J}{2v_0^2} \sum_{i,j=1}^{N} \Lambda_{ij} \boldsymbol{\pi}_i \cdot \boldsymbol{\pi}_j + \frac{JNn_c}{2}\right], \tag{48}$$

where $\Lambda_{ij} = n_c \delta_{ij} - n_{ij}$ is the discrete Laplacian. Notice that, within the spin wave expansion followed so far, the probability distribution of the $\boldsymbol{\pi}_i$ is therefore Gaussian.

In a flock the matrix n_{ij} depends on time because individuals can exchange positions, but we can disregard this mutual motion on timescales where this positional rearrangement is small (we will come back to this assumption later on). In this case the partition function can be computed by expressing the integral in terms of the eigenvectors of Λ_{ij} (which would be the Fourier modes, or plane waves, in a regular lattice, and are called spin waves in the theory of magnetism). The discrete Laplacian is a positive semi-definite matrix and has a zero mode – the Goldstone mode – which corresponds to the translational invariance of the $\boldsymbol{\pi}_i$, a remnant of the original rotational invariance of the flight directions. Calling a_k and w_k the eigenvalues and the eigenvectors of Λ, we get

$$\ln Z(J, n_c) = -\sum_{k>1} \ln(a_k) + \frac{JNn_c}{2}, \tag{49}$$

$$\langle \boldsymbol{\pi}_i \cdot \boldsymbol{\pi}_j \rangle = \frac{(d-1)v_0^2}{J} \sum_{k>1} \frac{w_i^k w_j^k}{a_k}. \tag{50}$$

These expressions allow to find analytically the values of J and n_c (see Fig. 5). They also show a very important property of the model: according to Eq. (50), the largest contribution to the pair correlation comes from the modes with the smallest eigenvalues, which are of order $1/L^2$ for a system roughly homogeneous in space. On a lattice this corresponds to a $1/k^2$ behavior in Fourier space and a power-law decay with distance. We therefore see that the scale-free nature of the correlation is strictly connected to the presence of the zero mode and to a spectrum that reaches out to zero. Scale-free correlations are therefore a consequence of the symmetry breaking of the rotational symmetry.

But does this model have any predictive power? It was shown in [32] that it does: with the appropriate boundary conditions, it reproduces not only the local correlation C_{int} (as it must, by construction), but also instantaneous correlation function $C(r)$ at large distances, as well as fourth-order correlations. It therefore appears that this very simple model truly captures some important property of the real system.

The ME model tells us that if we want to describe the statistical behavior of the flight directions in a flock, the most effective minimal model is one with pairwise short-range alignment. This conclusion translates within the ME approach the intuition of many flocking models that birds move driven by a social imitative force due to neighbors. The ME analysis also tells us for each flock the range and strength of such interaction: using the inferred parameters one sees that indeed these effective interactions are short-ranged.
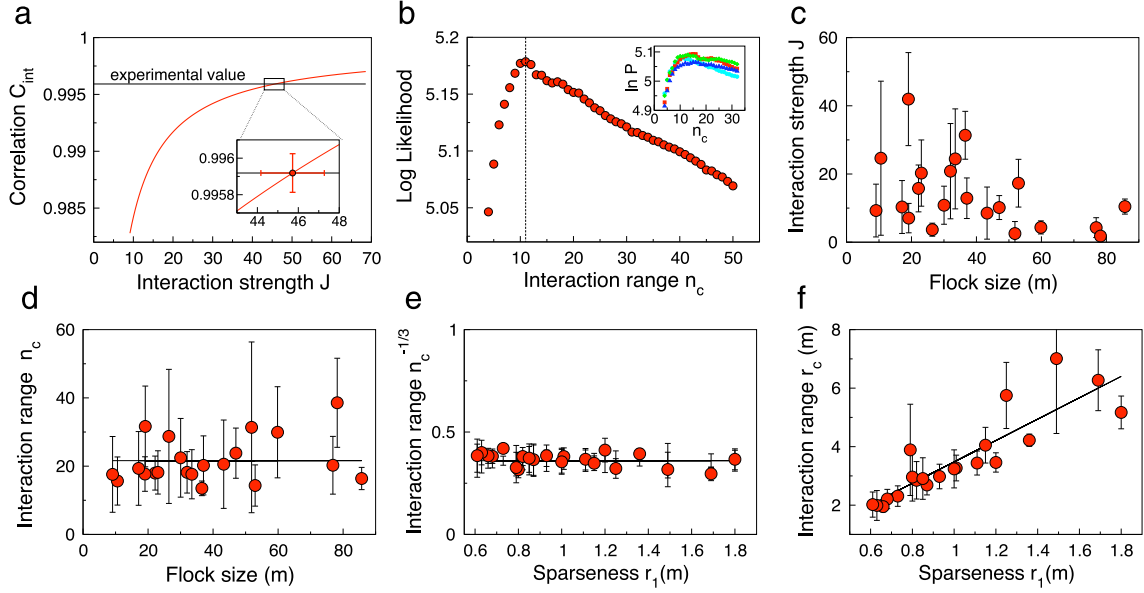
Fig. 5. A simple ME model for the flight directions. (a) The value of the parameter J is found by requiring that the prediction of the model for C_{int} is equal to the experimental value. (b) The value of n_c is found by maximizing the generalized entropy (log-likelihood) with respect to n_c. (c) and (d) Inferred values of J and n_c as a function of the flock size. (e) Inferred value of n_c as a function of the flock sparseness (i.e. the average nearest-neighbor distance r_1). (f) Inferred value of the metric interaction range as a function of sparseness.
Source: From [32].

2.4.2. Topological interactions

The inferred values of J and n_c [32] (Fig. 5) do not show any increase with the flock's size, showing that interactions are indeed short range. More interesting is the behavior of n_c with the flock's density, measured in terms of the nearest-neighbor distance r_1. Many flocking models assume that interactions have a well-defined metric range: an individual keeps track of neighbors closer than a certain interaction distance r_c. If this were the case, the number of interacting neighbors should decrease with increasing r_1, because the number of neighbors within distance r_c from a focal individual is (calling ρ the average density)

$$n_c = \frac{4}{3}\pi\rho r_c^3 \quad \longrightarrow \quad n_c^{1/3} \sim \rho^{1/3} r_c \quad \longrightarrow \quad n_c^{-1/3} \sim \frac{r_1}{r_c}. \tag{51}$$

This is not what is found in natural flocks. On the contrary, Fig. 5e shows that the number of interacting neighbors does not depend on the group's sparseness. This indicates that interactions in a flock are not based on metric proximity: each bird coordinates its motion with a given number n_c of neighbors independently of the flock's density. We call this kind of density-invariant interaction a "topological interaction" [18,32,19].

2.4.3. Short-range vs. long-range correlations

A further step in the ME approach consists in using more complex input observables. One can consider, for example, the topological pair correlation function

$$C(n) = \frac{1}{N}\sum_{ij}\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j\delta(k_{ij} - n), \tag{52}$$

where k_{ij} is the order of neighborhood of j with respect to i (i.e. $k_{ij} = n$ if j is i's nth-nearest neighbor [19]). $C(n)$ is analogous to the $C(r)$ defined in the previous sections but expressed in terms of the topological distance. It can also be measured quite robustly from empirical data and it is therefore an appropriate quantity for the ME procedure. The resulting ME model has again a Heisenberg form,

$$P(\{\boldsymbol{\sigma}_i\}) = \frac{1}{Z(J)}\exp\left[N\sum_n J(n)C(n)\right] = \frac{1}{Z(J)}\exp\left[\sum_{i,j=1}^N J(k_{ij})\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j\right], \tag{53}$$

where $J(n)$ are the Lagrange multipliers and $J(k_{ij}) = \sum_n J(n)\delta(k_{ij} - n)$. Here the discrete function $J(n)$ represents the strength of the effective alignment interaction between pairs at topological distance n. Solving this model therefore allows to infer the whole (topological) distance dependence of interactions, rather than just average range and strength. Again most of the
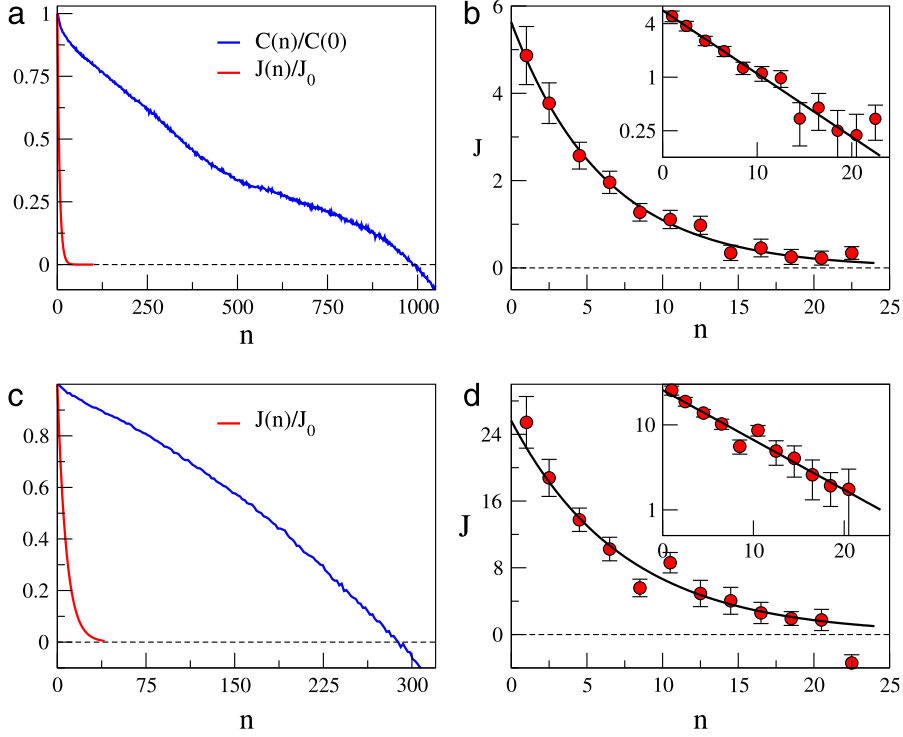
Fig. 6. Correlations vs interactions. (a) and (c) Experimental connected correlation functions (blue) compared to the inferred interaction function $J(n)$ for two flocking events. (b) and (d) Close-up of the interaction for the same events as in (a) and (c); insets: semilog plots of the same quantity. Event in (a) and (b): $N = 2126$, $J_0 = 5.63$, $n_c = 6.11$; event in (b) and (d): $N = 717$, $J_0 = 25.63$, $n_c = 7.41$.
Source: From [19].

computations can be performed analytically in the spin wave approximation [19]. Interestingly, one can show that only a small subset of the experimental $C(n)$ is necessary to perform the full inference (i.e. considering $C(n)$ for $n > n_{\max}$, with $n_{\max} \ll N$, does not change the results). The retrieved interaction strength is an exponentially decaying function $J(n) = J_0 \exp(-n/n_c)$ (Fig. 6) with range $n_c \sim 8$, consistent with previous estimates of the interaction range [18] and with the results of the simple ME model of Section 2.4.1. Since the ME procedure allows to choose among all possible functions $J(n)$, including long-range ones, that match the empirical correlations, the fact that the retrieved $J(n)$ is an exponential clearly shows that the long-range correlations observed in experimental data are generated by (effective) interactions that are in fact short-range.

2.4.4. ME for the speed

The ME method can also be applied to analyze the speed correlations [33]. The procedure is analogous to the one outlined in Section 2.4.1, but now including the speed in the input experimental observables. One considers a measure of local correlation of the full velocities that provides information of the mutual similarity of orientation *and* speed,

$$Q_{\text{int}} = \frac{1}{2n_c N} \sum_i \sum_{j \in n_c^i} (\mathbf{v}_i - \mathbf{v}_j)^2, \tag{54}$$

plus the scale of individual speeds,

$$S = \frac{1}{N} \sum_i |\mathbf{v}_i|, \qquad S_2 = \frac{1}{N} \sum_i |\mathbf{v}_i|^2. \tag{55}$$

The three quantities are scalar, local and averaged over all the flock. The corresponding ME distribution is

$$P(\{\mathbf{v}_i\}) = \frac{1}{Z(J, n_c, g)} \exp\left\{ -\frac{J}{4} \sum_{ij} n_{ij}(\mathbf{v}_i - \mathbf{v}_j)^2 - \frac{g}{2} \sum_i (|\mathbf{v}_i| - v_0)^2 \right\}, \tag{56}$$

where $v_0 = \langle S \rangle_{\exp}$ is the experimental value of the mean speed. There are now three Lagrange multipliers: J and n_c, as before, plus g, that controls the fluctuations of the speed around the mean value. This model therefore describes two distinct
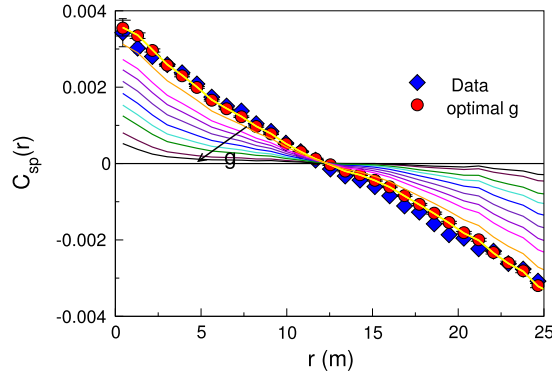
Fig. 7. Correlation function of the speed fluctuations, for different values of the control parameter g (color lines). The arrow indicates the direction of increasing g. The red points correspond to the prediction of the ME model for the optimal value of g, with $g/(Jn_c) \sim 10^{-2}$ close to the critical point; the blue points correspond to the experimental correlation.
Source: From [33].

competing mechanisms: the mutual adjustment of velocities driven by the social imitative interaction, and the individual control of the speed towards a preferred cruising value. In the highly polarized phase the spin-wave approximation gives [33]

$$\langle \boldsymbol{\pi}_i \cdot \boldsymbol{\pi}_j \rangle = (d-1)\frac{v_0^2}{J} \sum_{k>1} \frac{w_i^k w_j^k}{a_k}, \tag{57}$$

$$\langle (|\mathbf{v}_i| - v_0)(|\mathbf{v}_j| - v_0) \rangle = \frac{v_0^2}{J} \sum_{k>1} \frac{w_i^k w_j^k}{(a_k + g/J)}. \tag{58}$$

For the orientations, the result is the same as before, with long range correlations arising from the continuous symmetry breaking that implies the eigenvalues reach down to zero. In contrast, for the speed correlations we find a finite correlation length for $g \neq 0$ due to the g/J term in the denominator of (58). The correlation length is approximately [33]

$$\xi_{\text{speed}} \sim r_c \sqrt{\frac{Jn_c}{g}} \tag{59}$$

where r_c is the average distance of the first n_c neighbors. Thus the speed correlations are scale free only when the adimensional ratio $g/(Jn_c) = 0$: this value therefore corresponds to a critical point for the model.

Interestingly, the inferred value (i.e. that which satisfies the ME equations) of g/Jn_c is very low ($\sim 10^{-2}$) indicating that natural flocks are described by a ME model very close to its critical point (see Fig. 7). In summary, the ME model is telling us that in natural flocks the social imitative pressure (i.e. mutual adjustment interactions) dominates over speed control poising the system close to criticality.

2.4.5. Concluding remarks on the ME approach: what do we get and what do we miss

There are many models, which have been used in the last years to describe the collective behavior of biological assemblies and animal groups [34,1,2,35]. Models are usually based upon *a priori* reasonable assumptions, and eventually justified on the basis of their predictions, or fitted to experimental data when available. In many cases, though, it is difficult to understand the realm of validity of the model, its predictive content, and its limits. The ME approach follows a different strategy: it starts from the data, and from some variables one decides to describe, and produces an optimal model for those variables. What we get are effective models, whose degree of predictability can be systematically quantified and improved.

In the case of flocks, the method allowed us to infer some important properties of the interactions between individuals in a group. Some of these properties are consistent with intuitive ideas of the coordination mechanisms in a flock (and with typical assumptions of flocking models); some others, on the contrary, are much less trivial. Let us summarize them here:

(i) Interactions between birds in a flock are pairwise short-range alignment interactions.
(ii) Interactions are 'topological' rather than 'metric' (contrary to what assumed in typical flocking models).
(iii) Social imitation dominates over speed control, setting flocks in a very special point in parameters space, where the system behaves quasi-critically and scale free correlations arise.

What is, on the other hand, that we miss from the ME approach? All the ME models that we derived in this section, are built on static equal-time quantities. As a consequence, the models consist in static probability distributions. Flocks, however, are off-equilibrium systems where self-propelled organisms move in space with non-trivial spatio-temporal properties. If we want to capture and describe more exhaustively their dynamics we need to look at dynamical equations, and at multi-time experimental quantities. This is what we will do in the next sections.

2.5. Why are scale-free correlations relevant?

As we have seen, velocity fluctuations in starling flocks are correlated over long distances, and this applies to fluctuations of speed as well as direction. While long-range-correlated fluctuations of direction are an inescapable consequence of the breaking of a continuous symmetry and are found all throughout the ordered phase, long-range-correlated fluctuations of speed occur only for values of the control parameter very near a specific value (i.e. in the close vicinity of a critical point). It is thus natural to ask whether there is some "usefulness" to long range correlations that acts as evolutionary pressure in fine-tuning the system's parameters near criticality, which would otherwise result from an unlikely chance tuning. We will argue below that long-range correlations are related to information propagation, which is essential to collective decision making and maintaining group cohesion, which is in turn advantageous in the face of predators. Thus the argument implies that orientation fluctuations are biologically useful as well as physically necessary, and suggests that speed fluctuations are tuned to criticality to achieve information propagation of speed variations, although up to now only transmission of orientation variations has been unambiguously observed.

At first sight, it seems that information propagation does not require long range correlations: sound propagation in air is an example where meaningful information can propagate (as in human speech) in a medium where fluctuations are only correlated at very short distances. However, this depends on how information is encoded and decoded. Here we are seeking for a mechanism that can effect a *permanent* (or long-lived at least) change on the recipient: it is not enough that a distant bird changes briefly its orientation and then goes back to its original direction, but the whole flock must turn. To accomplish this with a passing wave one needs *memory:* a way (such as a change in the internal cognitive state of the recipient) that ensures the information that a wave has passed by is retained, and the other variables are adjusted accordingly. The other solution is to transmit what we call *directly useful* information, that is, information that does not require decoding (e.g. the orientation waves of Section 3.1). This simpler solution (in the sense that it requires less capabilities from the recipient of the information) needs, as we proceed to argue, long range correlations.

2.5.1. Transfer of information in linear response theory

Consider a field $\phi(\mathbf{x}, t)$ (say, the local orientation). We wish to achieve a change of the field by applying a perturbation at the origin. The field should change everywhere within the volume occupied by the system; in particular we will use the fact that $\mathbf{x}$ can be a point very far from the origin. We can write the localized perturbation as $h(\mathbf{x}, t) = \delta(\mathbf{x})f(t)$ and express its effect on the field at linear order as (see Appendix C)

$$\phi(\mathbf{x}, t) = \int_0^t dt' \, G(\mathbf{x}, t - t')f(t') = \int_0^t du \, G(\mathbf{x}, u)f(t - u), \tag{60}$$

where $G(\mathbf{x}, t)$ is the system's linear response $G(\mathbf{x}, \mathbf{x}'; t, t') = \delta\phi(\mathbf{x}, t)/\delta h(\mathbf{x}', t')$. The function $G(\mathbf{x}, t)$ encodes quite generally the response of ϕ to the perturbation, but $h(\mathbf{x}, t)$ and $\phi(\mathbf{x}, t)$ need not represent fields that interact directly at the physical level: ϕ could be the direction of an aircraft and $h(\mathbf{x}, t)$ a radio signal generated by an air controller. Or, in a simpler case, ϕ could be the local air pressure and h a pressure disturbance, so that G would be the Green's function of some differential equation. We impose the condition of *permanent change* on ϕ,

$$\phi(\mathbf{x}, t \to \infty) > \epsilon, \tag{61}$$

where ϵ is a sensitivity threshold.

Consider first $h(\mathbf{x}, t)$ as a control or trigger signal that lives for a short time T (the air controller radio signal). For large times then

$$\phi(\mathbf{x}, t) = \int_0^T dt' \, G(\mathbf{x}, t - t')f(t'), \qquad t > T, \tag{62}$$

and permanent change (61) requires that $G(\mathbf{x}, t \to \infty) > 0$, i.e. *long term memory*. This excludes situations where G represents propagation in a simple physical medium (which eventually returns to its equilibrium state and thus has no long term memory[2]), and requires instead a complex (e.g. cognitive) mechanism at work. It corresponds for example to the case of two human walkers coordinating the direction of motion through verbal communication ("wolf coming from the right!"). This case does not imply long-range-correlated fluctuations of $\phi(\mathbf{x}, t)$, but it requires instead a complex *decoding* mechanism.

The alternative to decoding is *directly useful information.* If we can alter the field at the origin and find a mechanism that propagates this change across space, we do not need decoding: the message is carried by ϕ itself, so that when the signal arrives, the intended effect is already achieved. A permanent change at $\mathbf{x}$ is thus in this case accompanied by a similar permanent change at the origin, so assuming that $G(\mathbf{x}, t)$ decays to zero with some characteristic time τ (i.e. there is no long-term memory) we have

$$\phi(\mathbf{x}, t \to \infty) = \lim_{t \to \infty} \int_0^t du \, G(\mathbf{x}, u)f(t - u) \approx \int_0^\tau G(\mathbf{x}, u)f(\infty). \tag{63}$$

---

[2] Actually there are a few cases of physical systems with infinite memory, linked to conservation laws, like the one-dimensional wave equation. However these cases do not arise in practice in the systems we are considering (three dimensional).

Introducing the static susceptibility $\chi(\mathbf{x} - \mathbf{x}') = \delta\phi(\mathbf{x})/\delta h(\mathbf{x}')$ (see Appendix C),

$$\chi(\mathbf{x}) = \int_0^\infty du \, G(\mathbf{x}, u) \approx \int_0^\tau du \, G(\mathbf{x}, u) \approx \phi(\mathbf{x}, t \to \infty) > \epsilon, \tag{64}$$

where we used the permanent change condition. Since $G(\mathbf{x}, t)$ is here the response of a physical system, we invoke the static fluctuation–dissipation theorem (Appendix C), which links the static susceptibility $\chi(\mathbf{x})$ to the connected correlation function to find,

$$C(\mathbf{x}) = \langle \phi(\mathbf{x}, t)\phi(0, t) \rangle - \langle \phi(\mathbf{x}, t) \rangle \langle \phi(0, t) \rangle \propto \chi(\mathbf{x}) > \epsilon. \tag{65}$$

Thus, under these assumptions, in the absence of long term memory, to achieve a permanent change of the field at $\mathbf{x}$ *requires that the fluctuations of the field be correlated over a distance* $|\mathbf{x}|$.

The argument is subtler than saying "since a change at the origin implies a change at $\mathbf{x}$, the field at the origin is obviously correlated with the field at $\mathbf{x}$", for this simply implies *nonconnected* correlations. The argument above, going through the response function, concludes that the *connected* correlation (i.e. the correlation of fluctuations) is long-ranged.

The weak part of the argument when applied to active matter systems is the use of the fluctuation–dissipation relation, which is a result that applies to equilibrium systems. Under what conditions can it be applied to active systems remains to be clarified, but we accept it as valid, at least approximately, on the grounds that Onsager's principle (that relaxation from a spontaneous fluctuation or relaxation from an external-field-induced value must proceed in the same way) should hold.

2.5.2. The need for dynamics

The argument above thus suggests that long-range correlations of fluctuations are a feature that must accompany transfer of information in the absence of a decoding mechanism. But this alone does not suffice to guarantee an efficient communication. In particular, we have not taken into account the fact that the information must arrive *timely:* if the signal that triggers a direction change takes too long to arrive, different parts of the group will have been moving in different directions for a long time, and the group's cohesion will be lost. One can guess that diffusive propagation (where the signal speed vanishes for long distances) or wavelike propagation in a medium with high damping are situations where information will not be efficiently transferred. Indeed, transfer of turning information in flocks occurs through wave propagation, in which inertia plays a crucial role, as we discuss in Section 3. Although this seems the most efficient attainable way of propagating information, and perhaps the only that guarantees timely propagation, the issue of the role played by inertia in information transfer has yet to be fully explored.

3. Information propagation

In the previous sections we discussed one of the distinctive features of collective behavior in flocks: the presence of scale free correlations between the velocities of individual birds. As we argued, such long range correlations are linked to the ability of the system to transfer information throughout the group, a minimal requirement for achieving a collective response. To further investigate the mechanism of information propagation, we will now consider phenomena that fully take into account the temporal evolution of the directional degrees of freedom. In the following sections we will describe some experimental findings that very clearly exhibit transfer of directional signals: collective turns in flocks of birds. Next, we will discuss the predictions of current models of collective motion, show that they are inconsistent with experimental data, and develop a new theoretical framework able to explain what is observed in natural groups.

3.1. Experimental evidence

Orientation waves (i.e. wave-like propagation of turn information) have been observed in starling flocks in the field [36] by studying the full 3-d reconstruction of the trajectories of individual birds within flocks that visually perform a turn while in the field of view of the experiment.

The trajectories were obtained with a trifocal method [37] which employs video sequences from three digital cameras. The recorded groups were European starlings (*Sturnus vulgaris*) that roost at the Piazza dei Cinquecento site in Rome, and 12 distinct flocking events including one collective turn were recorded.

The acceleration $\mathbf{a}_i(t)$ of each bird can be used to locate the time of the individual's turn, as its modulus has a noticeable peak during the turn (see Fig. 8). A robust way of establishing the time differences between individual turns in the presence of experimental errors and noise is through the acceleration overlap function

$$Q_{ij}(t) = \frac{1}{\sigma_i \sigma_j} \left[\left\langle \mathbf{a}_i(t_0) \cdot \mathbf{a}_j(t_0 - t) \right\rangle_{t0} - \left\langle \mathbf{a}_i(t_0) \right\rangle_{t0} \cdot \left\langle \mathbf{a}_j(t_0 - t) \right\rangle_{t0} \right], \tag{66}$$

where the average is over t_0: $\langle \ldots \rangle_{t_0} = \int dt_0 \, (\ldots)$ and $\sigma_i = \sqrt{\langle \mathbf{a}_i(t_0)^2 \rangle - \langle \mathbf{a}_i(t_0) \rangle^2}$. The time shift τ_{ij} is defined as the time that maximizes $Q(t)$, and a positive shift $\tau_{ij} > 0$ means that j turns before i. In the absence of noise, one would have $\tau_{ij} = \tau_{ik} + \tau_{kj}$ for every triplet i, j, k, but this time ordering is sometimes violated due to experimental errors. Thus to establish the order of the turns one can resort to a ranking procedure employed in sports, where the relation "wins over" is not necessarily transitive. In Ref. [36] the ranking was assigned according the score $\Omega_i = \sum_{j \neq i} \text{sign} \, \tau_{ij}$, as in a round robin tournament [38].
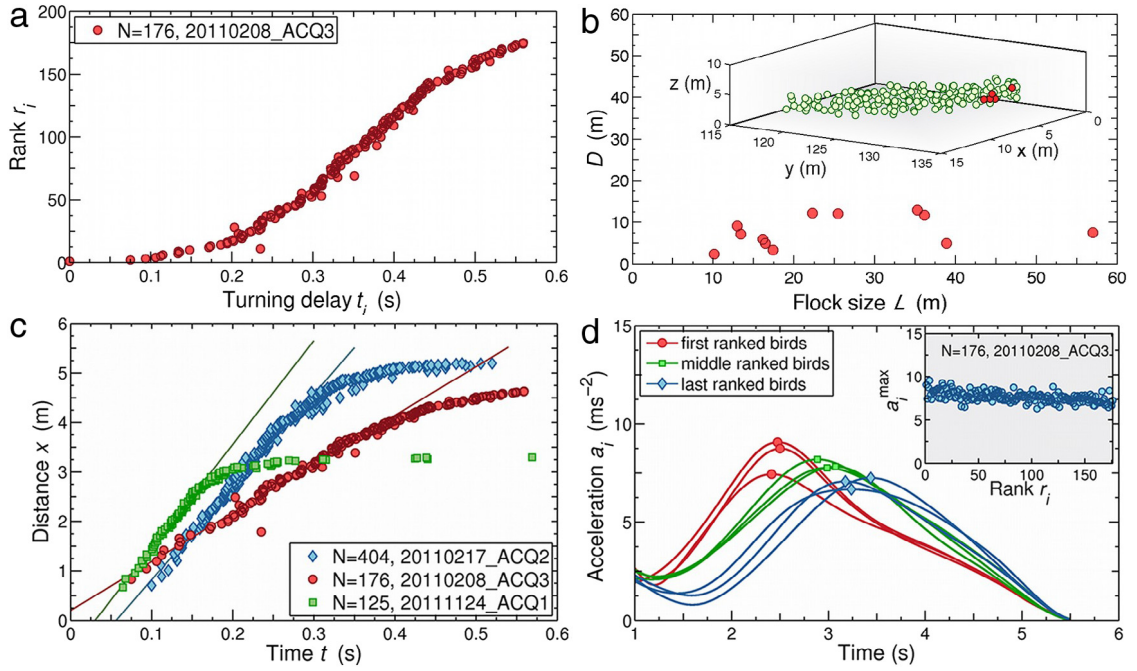
Fig. 8. Propagation of the turn across the flock. (a) Rank of each bird vs. absolute turning time (see text) for a single turning event. (b) Maximum mutual distance between the first 5 birds in the turning sequence. Inset: actual position of the first 5 birds (red) within the flock. (c) Distance x traveled by the turning information as a function of the absolute turning time, for three different turning events. The speed of propagation c_2 is given by the slope of the curve in the linear regime (solid lines are linear fits). (d) Radial acceleration as a function of the absolute turning time for several birds with low, intermediate and high rank in the turning sequence. The intensity of the peak (filled symbols) decreases very weakly in passing from the first to the last turning birds. Inset: value of the acceleration at the peak as a function of the rank of the bird.
Source: From Ref. [36].

In this way each individual is a assigned a *rank* r_i, so that $r_i = n$ means bird i was the nth to turn, and an absolute turning time t_i, i.e. the delay with respect to the top-ranked bird (the first to turn, $r_{\text{top}} = 1$, $t_{\text{top}} = 0$), computed as

$$t_i = \frac{1}{r_i - 1} \sum_{r_j < r_i} t_j + \tau_{ij}. \tag{67}$$

In this way the ranking curve $r(t)$ is obtained (Fig. 8a), which is the starting point of the information propagation analysis.

3.1.1. A propagation phenomenon
The first finding from the ranking curves is that there is a propagation phenomenon behind the dynamics of turns. The top-ranked birds (i.e. the first to turn) are always found spatially close to each other: in Fig. 8b, the maximum mutual distance between the top 5 birds is seen to be independent of flock size, indicating a spatially localized origin of the turn. Hence, the decision to turn is taken at some point in space, and then propagates across the group through a bird-to-bird ("social") transfer of information [39]. The alternative view, namely that the turn is caused by an external stimulus hitting all birds at the same time, would imply an independent response of each bird and thus a spatially unstructured distribution of the delay times, whereas the delays have a clear spatial modulation (Figs. 8c and 9). The propagation mechanism of the turn information has to be efficient enough that it reaches the whole group before the transient differences in velocity can split the group apart. Our aim is to determine the mechanism through which information propagates from its local origin to the rest of the flock.

3.1.2. Linear dispersion law
The second crucial finding is that the turn information propagates with a constant speed. This is seen when we compute the distance $x(t)$ the information has traveled in time t, which can be done from the ranking curve. Since we are in three dimensions and the turn has a localized origin, $x(t)$ is equal to the radius of the sphere containing the first $r(t)$ birds in the rank, namely

$$x(t) = \left(\frac{3r(t)}{4\pi \rho} \right)^{1/3}, \tag{68}$$
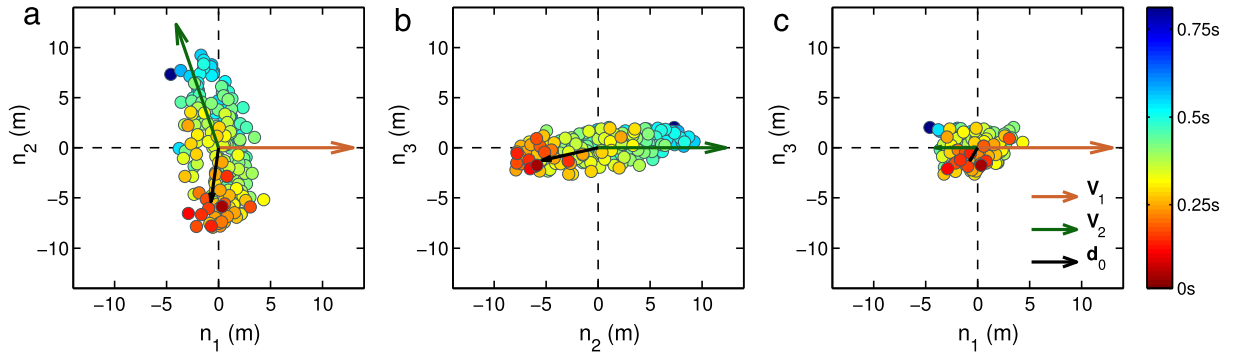
where ρ the density of the flock.

Fig. 9. Turn propagation for an event in a flock of 176 birds. Shown are projections of the birds' positions in an orthogonal system ($\mathbf{n}_1$, $\mathbf{n}_2$, $\mathbf{n}_3$) centered at the flock's center of mass (the view is from the top in panel (a), from the front in panel (b) and from a side in panel (c)). Before the turn, the velocity is $\mathbf{V}_1$ (orange vector), and after the turn it is $\mathbf{V}_2$ (green vector). The vector $\mathbf{d}_0$ is the mean position of the first 10 birds to turn. Colors indicate the turning time delays t_i as indicated by the color bar on the right.
Source: From Ref. [40].

The most important feature of the curve $x(t)$ (Fig. 8c) is that there is a clear linear regime for early and intermediate times. For long times border effects appear: in the bulk the rank grows as $\sim x^3$, but when the perturbation is near the border the number grows with a smaller exponent (for example, for a disc-shaped flock one would have $r \sim x^2$ when the perturbation has traveled the distance of the shortest linear dimension, for a tube-shaped flock it would be $r \sim x$), so that for late times $r_{\text{late}} \sim x^\alpha$ with $\alpha < 3$, and then $x \sim r^{1/3} \sim x_{\text{late}}^{\alpha/3}$, i.e. $x(t)$ shows a saturation for long times as is seen in the figure.

We conclude that the distance traveled by the information grows linearly with time,

$$x(t) = c_2 t. \tag{69}$$

The parameter c_2 is the speed of propagation of the directional information; it was measured in the range 20–40 m/s *in the center-of-mass reference frame*, for all the turning events. These values are high: this makes it possible that the decision to turn can sweep through a flock of 400 birds in little more than half a second. We stress that $x(t)$ is the distance traveled by the information in the flock's reference frame, and that it is also uncorrelated to the birds' speed v_0 (which is in the range 7–12 m/s). Thus this transport is not the mere effect of the flock's absolute motion, and the information is not being carried through the displacement of individuals themselves (which in the center-of-mass frame can be at most $2v_0$ for the case of a bird that makes a 180° turn). This conclusion is confirmed by the measured stability of the network during the turn [40], i.e. the fact that the local neighborhood of a bird does not change significantly during a turn.

3.1.3. Negligible attenuation

Another interesting piece of information comes from the acceleration curves: the information to turn propagates across the flock with negligible attenuation (Fig. 8d). Flocks are large, and the information to turn reaches all birds through many intermediate steps, so that substantial damping might be expected. Yet it is not so. Both sublinear propagation and attenuation would result into a physical spread of the flock, and eventually into its disruption, but on the contrary, the fact the propagation is linear and fast, together with the low damping of the signal, are key factors in preserving the flock's cohesion.

3.1.4. Equal-radius paths

During the process of turning, each bird performs its own individual turn following a specific trajectory in space, and as a consequence the flock as a whole performs a collective turn. These two dynamics are strictly interconnected, and the way individuals move relative to each other during the global turn was studied in detail in Ref. [40]. From this analysis it emerges that starlings turn following paths of approximately the same radius, a feature first observed in a seminal experimental work on flocks of rock doves [41].

Equal-radius turning is very different from how a rigid assembly would turn (Fig. 10): in the rigid case, all particles turn around the same center point (fixed in the center-of-mass reference frame), following parallel paths with different radius of curvature and different speeds. To an external observer (reference frame fixed to the ground), the relative positions change (after a 180 degree turn, bird i, which was to the left of bird j, is now to the right of j), but the relative positions and orientations in the center-of-mass frame remain unaltered. In contrast, in equal-radius trajectories, the relative positions as seen by an external observer remain unchanged, but the internal orientational topology and organization changes (a bird that was near the front of the group finishes near its back, the orientation of the axes giving the shape of the group can change with respect to the direction of polarization, see Figs. 10 and 12). Equal-radius turning is advantageous in the sense that a very quick collective turn can be accomplished without significant changes in the individuals' speeds. The resulting change in the
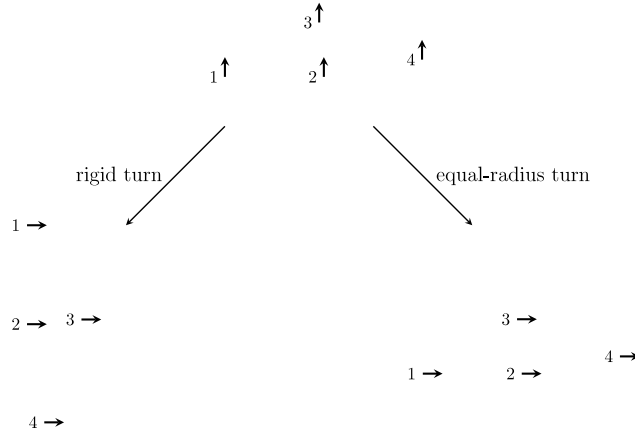
Fig. 10. Schematic depiction of the difference between a rigid and an equal-radius turn. For a perfectly ordered flock of four birds, the final state is shown after a 90° turn in the case of a rigid turn (left) and an equal-radius turn (right). In the latter case the orientational topology is altered: for example bird 1, which was at the left of the flock, ends up at the rear. See also Fig. 12.

relative local orientations causes as by-product a redistribution of boundary locations and consequently of risk among the individuals [40].

Evidence of equal-radius turning is shown in Fig. 11. A quantity that can be computed to check for equal-radius turning is the change in the angle of the vector giving the relative position of two birds. We define

$$\cos\theta_{ij}(t) = \frac{\mathbf{r}_{ij}(0)}{|\mathbf{r}_{ij}(0)|} \cdot \frac{\mathbf{r}_{ij}(t)}{|\mathbf{r}_{ij}(t)|}, \qquad \mathbf{r}_{ij}(t) = \mathbf{r}_i(t) - \mathbf{r}_j(t). \tag{70}$$

The angle $\theta_{ij}(t)$ measures how much the relative positions have rotated at time t, as seen from an external reference frame. This angle should be zero for a perfect equal-radius turn, which preserves the relative positions as seen from the outside. The average of θ_{ij} over all pairs, $\theta_r(t)$ is plotted in Fig. 11, and is seen to remain very small during a turn of about 180°. The internal reorientation is also clearly observed (Fig. 12).

3.1.5. Wavelike propagation

The experimental findings we have described point to a wavelike propagation of turn information: a perturbation with a localized origin reaches the whole flock propagating with a well-defined front that moves with constant speed. In addition this speed is independent from the speed of the source, and is instead a property of the propagation medium (the flock).

In Fig. 8c one sees that c_2 is different for different flocks. However, it is not immediately obvious what properties of the flock determine the speed. The familiar sound waves suggest examining a possible relationship between c_2 and density, but Fig. 13 makes it clear that the propagation speed is unrelated to the density. One might perhaps have guessed that c_2 should be independent of density given the finding that the interactions determining static correlations are topological (Sections 2.4.2 and 2.4.5). But anyway Fig. 13 shows that propagation speed varies significantly from flock to flock.

It is a remarkable prediction, verified by experimental data, of the second sound theory developed below (Section 3.3) that the speed of propagation of turning waves c_2 depends on the *polarization* of the flock (see Fig. 14). The relationship between speed and polarization derives from the former's dependence on the velocity–velocity coupling, and is explained in Section 3.3.4; it turns out that in the high polarization phase (the regime relevant to the data of Fig. 14) one has

$$c_2 \propto \frac{1}{\sqrt{1-\Phi}}. \tag{71}$$

The link between c_2 and behavioral polarization Φ is the mathematical consequence of a symmetry. However, the specific level of polarization of a given flock is not fixed by mathematics or by symmetry, but by adaptive factors. In many social species, polarization is very large [42,35,32]: global order is indeed the most conspicuous trait of collective behavior. However, an especially large polarization is not required if the only concern of a bird were just to avoid colliding with its neighbors, since flocks are rather dilute systems, with packing fractions lower than 0.01 [14]. Yet these same flocks are very ordered, with polarization Φ close to 1. Why is that?

Relation (71) may be one of the reasons behind the large value of polarization. In collective decision-making, swift transfer of information is beneficial to the cohesion of the group (this is quite clear in the case of turns, since during the turn the wavefront divides the flock into two groups with different directions of motion, and such misalignment causes spatial spread of the flock and loss of cohesion). The slower the speed c_2 of the wavefront, the more severe this loss. It is reasonable to believe that this is a general mechanism for moving biological groups. Every collective decision causes momentary lapse of cohesion in the group, due to the transient coexistence of different behavioral states. The link (71) between high behavioral
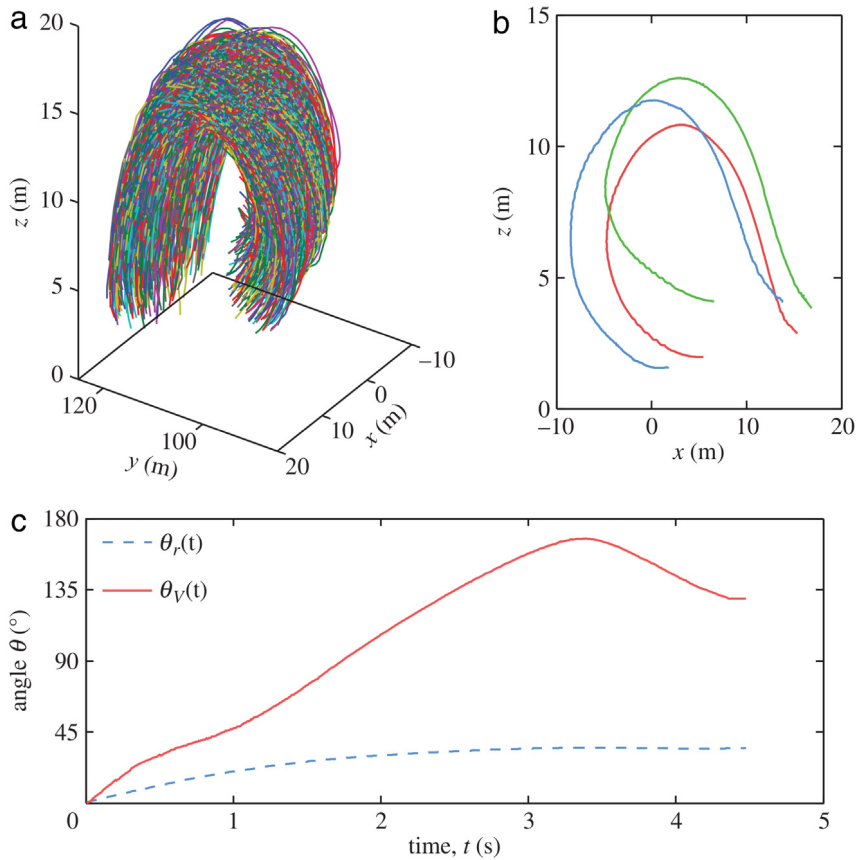
Fig. 11. Equal-radius turning of starling flocks. (a) Three-dimensional trajectories of the whole flock during a turning event. (b) Two-dimensional projection of three individual trajectories of nearby birds, showing the equal-radius path with each bird turning around a different center but with the same radius of curvature. (c) Evolution of the angles $\theta_r(t)$ measuring the change in relative positions as seen from an external observer during the turn (see text and Eq. (70)), and the angle $\theta_V(t)$ the instantaneous average velocity $\mathbf{V}(t)$ makes with the initial orientation. *Source:* From Ref. [40].
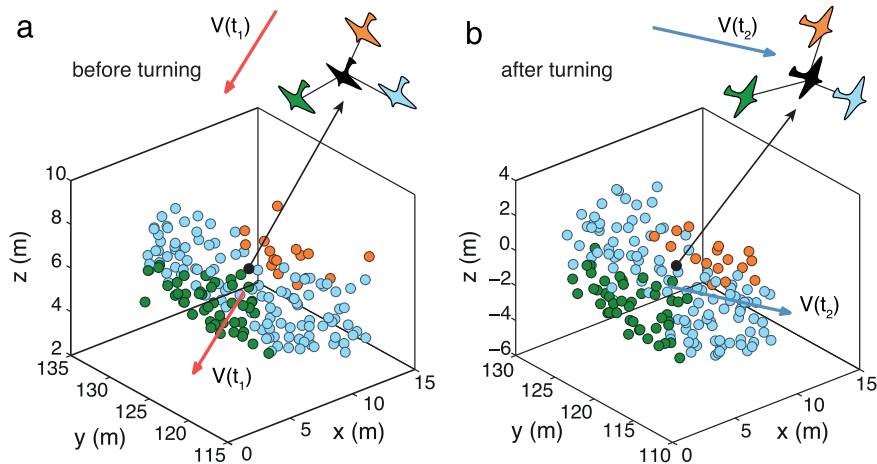
Fig. 12. Experimental observation of internal reorientation of a flock during a turn. A reference bird (painted black) is chosen near the center of the flock, and the rest are painted according to their orientation with respect of the reference (front: green, sides: light blue, behind: orange). After the turn, the light blue birds are in front and behind the reference, while green (orange) birds finish to its right (left). *Source:* From Ref. [40].

Fig. 13. Speed of propagation vs. nearest-neighbor distance (panel a) and flock size (panel b). Each point corresponds to a different turning event. There is no correlation between the propagation speed and the density or the size of the flock.

Fig. 14. Propagation speed vs. polarization. (a) Propagation speed (indicated as c_s in the original paper), normalized by the average nearest neighbor distance, plotted as a function of $(1 - \Phi)^{-1/2}$. The new theory discussed in Section 3.3 predicts a l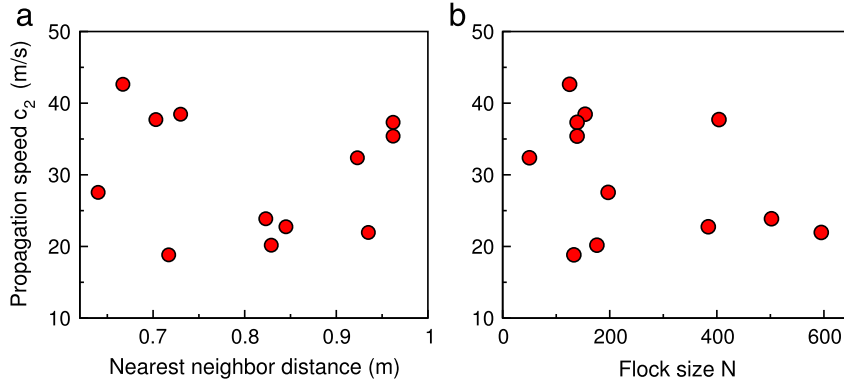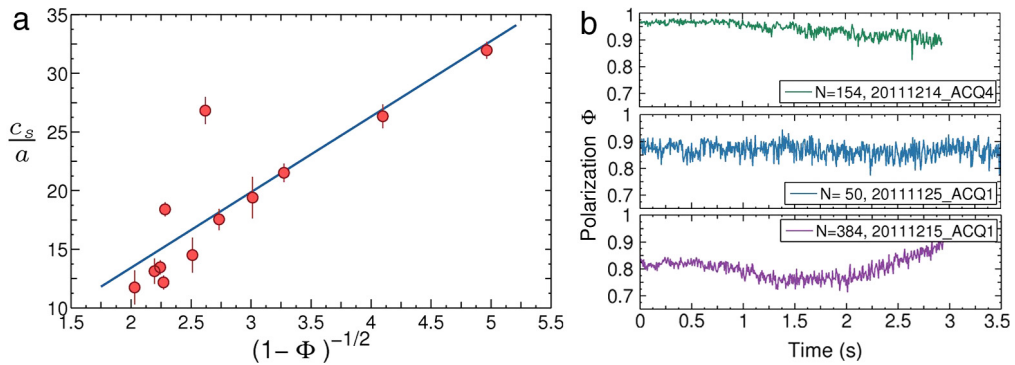inear relationship between these two quantities. This highly non-trivial functional relationship between propagation speed and polarization is very well obeyed by the empirical data (p value: $p = 3.1 \times 10^{-4}$; correlation coefficient $R^2 = 0.74$). (b) polarization as a function of time for three turning events. The value of Φ used in (a) corresponds to a time average over the entire turning sequence.
Source: From Ref. [36].

polarization and fast propagation of information suggests that keeping this lapse to a minimum, therefore achieving a fast collective decision, may be the adaptive drive for the high degree of order observed in many living groups.

3.2. First sound

We have seen that bird flocks are characterized by propagating phenomena. In this Section we will provide a discussion of the various theoretical frameworks that can be introduced to describe linear signal propagation in the context of collective motion.

3.2.1. The Vicsek model of collective motion

The Vicsek model [5] (VM) is a prototypical model of flocking, perhaps the simplest active matter model that displays a transition from a disordered state with no polarization to a symmetry-broken phase with finite polarization (see e.g. [43] for a review). In three dimensions, the Vicsek model is defined by the equations,

$$\mathbf{v_i}(t + 1) = v_0 \mathcal{R}_\eta \Theta \left[\mathbf{v}_i + \sum_j n_{ij} \mathbf{v}_j(t) \right], \tag{72}$$

$$\mathbf{r_i}(t + 1) = \mathbf{r}_i(t) + \mathbf{v}_i(t + 1), \tag{73}$$

where $\Theta(\mathbf{x}) = \mathbf{x}/|\mathbf{x}|$ is the normalization operator and R_η rotates its argument randomly within a spherical cone centered at it and spanning a solid angle $4\pi\eta$. The connectivity matrix n_{ij} defines as usual which pairs interact. In the original VM the connectivity matrix elements take only the value 0 or 1. The matrix can be *metric* (i.e. $n_{ij} \neq 0$ if and only if $r_{ij} < r_c$) or

topological (i.e. $n_{ij} \neq 0$ if j is one if i's first n_c neighbors[3]). The active nature of the model is encapsulated in the constraint on the speed of the particles, which is fixed to v_0,

$$|\mathbf{v_i}| = v_0, \qquad \forall i. \tag{74}$$

Except in the fully connected case where $n_{ij} = 1$, the $\{n_{ij}\}$ depend on the positions $\{\mathbf{r}_i\}$, which in turn change with the velocities, which are determined by the connectivity matrix itself (and the random noise). This dependence introduces a coupling between local order and local density fluctuations characteristic of active systems and responsible for the peculiar properties of the VM, such as the spontaneous breaking of a continuous symmetry even in $d = 2$ and the presence of giant density fluctuations [25,43]. Several variants of the model have been introduced in the literature [2,24], with vectorial rather than scalar noise, slightly different update rules, or cohesion forces. All the variants describe a phenomenology where a transition occurs between a disordered phase (high noise, small density) and a polarized phase of collective motion (small noise, large density).

The VM incorporates within a dynamical framework the alignment interactions that we inferred for natural flocks in the context of the Maximum Entropy (ME) approach. In the same way as the ME models Eqs. (46) and (53) for the statistics of the flight directions, it is endowed with a continuous rotational symmetry, which is broken in the polarized phase. Indeed it also exhibits scale-free correlations of the velocity fluctuations (only of the orientations, of course, as the speed is fixed), as observed in natural groups (even though with a different exponent). Contrary to the simple static ME distributions that we derived in the previous sections, however, it is a dynamical model and explicitly includes the rearrangement of the network in its description.

3.2.2. Continuous time Vicsek model

The VM has been extensively investigated using numerical simulations, and its original formulation (discrete update equations, minimal number of parameters) has been important to obtain a deep understanding of the phase diagram, and of the control parameters regulating the transition. Some of the features of Eq. (72) are however unpractical for an analytic investigation: (1) they are defined for discrete steps, while it would be helpful to have a model defined in the continuous time limit; (2) the normalization operator Θ is difficult to handle; (3) there is no parameter explicitly regulating the strength of alignment. For these reasons, we now define a slightly different version of the original model, which can be more easily studied in the continuous time limit and within the spin wave expansion.

As a first step, we notice that in the social force at the r.h.s. of Eq. (72), the term $i = j$ has the same weight as the other neighbors. If we want to consider the limit for small time increments (rather than $\Delta t = 1$ as in (72)), we need the self contribution to remain finite at the r.h.s. of the update equation in order to build a time derivative. Besides, we expect the force exerted by neighbors to become smaller if we consider smaller time increments. Thus instead of $\mathbf{v}_i + \sum_{j \neq i} n_{ij} \mathbf{v}_j(t)$, for small increments dt we will consider a term of the kind $\mathbf{v}_i + J dt \sum_{j \neq i} n_{ij} \mathbf{v}_j(t)$, where J is the strength of the alignment interaction, also known as stiffness. Besides, instead of the Θ operator, we can use a Lagrange multiplier to enforce the fixed speed constraint. In this way we can take the limit for $dt \to 0$ and get

$$\eta \frac{d\mathbf{v_i}}{dt} = J \sum_j n_{ij} \mathbf{v}_j(t) + \lambda_i \mathbf{v}_i + \boldsymbol{\zeta}_i, \tag{75}$$

$$\frac{d\mathbf{r_i}}{dt} = \mathbf{v}_i, \tag{76}$$

where we added a parameter η fixing the timescale of the dynamical update. Here $\boldsymbol{\zeta}_i$ is a Gaussian noise with variance,

$$\langle \boldsymbol{\zeta}_i(t) \cdot \boldsymbol{\zeta}_j(t') \rangle = 2d\eta T \delta_{ij} \delta(t - t'), \tag{77}$$

and the Lagrange multiplier λ_i is fixed by the condition $|\mathbf{v}_i| = v_0$. We note that, in this formulation, Eq. (75) is very similar to a Langevin spin dynamics in presence of ferromagnetic interactions [44]

$$\eta \frac{d\mathbf{v_i}}{dt} = -\frac{\delta H}{\delta \mathbf{v}_i} + \lambda_i \mathbf{v}_i + \boldsymbol{\zeta}_i, \tag{78}$$

with the pseudo Hamiltonian,

$$H = -J \sum_{ij} n_{ij} \mathbf{v}_i \cdot \mathbf{v}_j. \tag{79}$$

Of course, compared to the lattice ferromagnetic case, in the VM the network is not fixed, n_{ij} changes with time, and the system is in general out of equilibrium. However, it is useful to keep in mind this analogy. On timescales where the network's movement is not relevant, Eq. (75) describes a standard relaxational dynamics for the orientational degrees of freedom. It also illustrates how dynamical update equations of the VM kind are related to the ME analysis of Section 2.4: the pseudo Hamiltonian appearing in Eq. (78) is of the same kind as the ME one (see Eq. (46)). Indeed, continuous time Vicsek models can

---

[3] This is the case relevant to starling flocks, which have $n_c \approx 6 - 7$, as discussed in the previous sections.
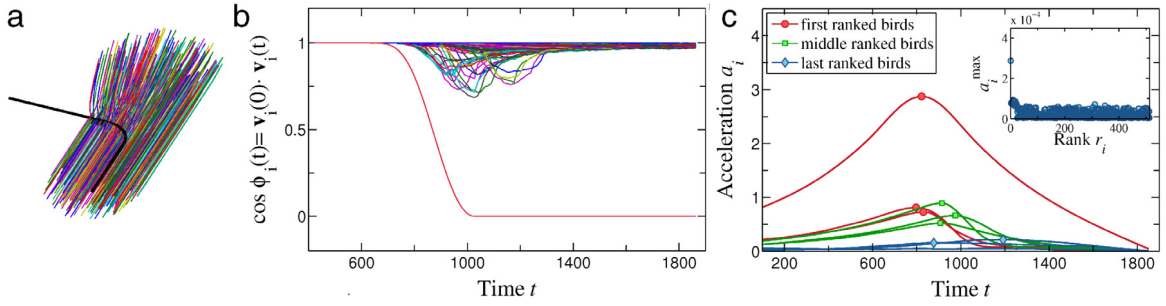
Fig. 15. Attempt to propagate turn information in the 3-*d* Vicsek model, showing that the turn cannot propagate and results in an individual leaving the flock. A direction change is imposed on one individual (black trajectory). Top panel: 2-*d* projection of trajectories. Middle panel: cosine of the individual velocities with respect to the original direction (before forcing the turn). Lower panel: individual acceleration profiles.
Source: From Ref. [46].

be derived directly from the data using a dynamical ME approach: one can show that the ME model consistent with velocity data at consecutive time intervals (rather than at the same time), is a Markov model of the VM kind similar to (75) [45,8].

To conclude, we note that if we explicitly compute λ_i (e.g. by multiplying both sides of Eq. (75) by $\mathbf{v}_i$, and imposing $d(\mathbf{v_i}^2)/dt = 0$) we get an equivalent expression for the velocity update equation

$$\eta \frac{d\mathbf{v_i}}{dt} = \left[J \sum_j n_{ij} \mathbf{v}_j(t) \right]^\perp + \boldsymbol{\zeta}_i^\perp, \tag{80}$$

where $\mathbf{w}^\perp = \mathbf{w} - (\mathbf{w} \cdot \mathbf{v}_i)\mathbf{v}_i/v_0^2$ indicates the component of the vector $\mathbf{w}$ perpendicular to $\mathbf{v}_i$. In the rest of the paper we will often use this continuous time version of the VM as reference benchmark of models of the Vicsek kind.

3.2.3. Collective turns in the Vicsek model

The VM is successful at modeling the static properties of flocking behavior, explaining how a short-range velocity-aligning interaction can lead to globally ordered phase where all members of the group share a common velocity, and producing scale free correlations. Besides, it offers the minimal Markovian description of collective motion. Given this success and the simplicity of the VM, it is natural to ask whether it can also describe time-dependent quantities, the kind of information propagation observed in real flocks, and the occurrence of collective turns. The answer to this question is negative; we will thoroughly discuss the theoretical reasons for this in the next sections, but we can anticipate that the problem is that the VM describes an *overdamped* update equation for the velocities of the active particles (see Eq. (75)). This issue was first studied numerically in Ref. [46], where one of the particles of an otherwise aligned and highly polarized flock was forcefully assigned a different direction. It was found the over-damped dynamics in a Vicsek-like model cannot effectively propagate this perturbation across the system, with the result that the initiator of the turn leaves the group, while the rest of the particles display at most a slight and transient change in their orientation, and the group sticks to the initial direction (Fig. 15).

Thus the dynamics of flocks is not captured by the VM and requires the introduction of a model that, while preserving the statics of the VM, allows fast (wave-like) propagation of turning information. This is true even if the VM *can* sustain waves: but the waves in the VM are *density* waves, not orientation waves. They are related to the coupling of local order and density fluctuations mentioned above, and have different properties from the orientation waves that have been observed to propagate the turn information. A detailed explanation of the VM waves, their inadequacy, and how to formulate a model that can sustain orientation waves is rather challenging, and is the subject of the next sections.

3.2.4. Spin-wave expansion of the Vicsek equation: fixed network case

Numerical simulations show that the Vicsek model (VM) in the polarized phase does not exhibit collective turns as real flocks. How can we understand this behavior from a theoretical perspective? Clearly the VM – despite its simplicity – is very complex to treat analytically as it involves interactions between velocities (orientational degrees of freedom) and movement in space, within an off-equilibrium context. We will therefore address some limiting situations where computations can be performed: the fixed network case and the hydrodynamic regime. In the first case, we assume that the particles have fixed mutual positions so that the interaction network n_{ij} does not depend on time. This approximation is reasonable to describe processes that occur on timescales smaller than the rearrangement scale of the network (collective turns in natural flocks seem to be close to this limit). The hydrodynamic regime, on the other hand, describes the very large scales, where network movements cannot be disregarded and the system behaves as an active fluid fully mixing orientational and density modes. Let us start by discussing the fixed-network approximation.

We consider the VM in the deeply polarized phase. In this regime we can use the spin-wave approximation discussed in Appendix B. We expand each velocity vector with respect to the mean velocity direction (that we consider as the *x*

direction): $\mathbf{v}_i = v_i^x \mathbf{n}_x + \boldsymbol{\pi}_i \sim \mathbf{n}_x v_0 \left(1 - \frac{1}{2}\pi_i^2/v_0^2\right) + \boldsymbol{\pi}_i$, with $\boldsymbol{\pi} = v_0(0, \varphi^y, \varphi^z)$. We can now plug these expressions into the full dynamical equations describing the model (e.g. in Eq. (80)) and expand them up to the first order in the phase, so to obtain equations directly for the φ_i. We get the same equation for both φ^y and φ^z, namely,

$$\eta \frac{d\varphi_i}{dt} = -J \sum_j \Lambda_{ij}\varphi_j + \zeta_i \, , \tag{81}$$

where Λ_{ij} is the discrete Laplacian matrix defined in Section 2.4, $\Lambda_{ij} = -n_{ij} + \delta_{ij}\sum_k n_{ik}$.

The Laplacian matrix depends in general on time, since n_{ij} is updated every time individuals exchange positions and leave each other's neighborhood. Let us assume that this process is slow on the scales of interest so that the network can be considered as fixed. If we look at spatial scales larger than the mean inter-particle distance – and we must do that, lest all our scaling relations lose their validity – we can then approximate the discrete Laplacian with its continuous counterpart (we call a the mean inter-particle distance and n_c the number of interacting neighbors),

$$J \sum_j \Lambda_{ij} \to -Jn_c a^2 \nabla^2 \, . \tag{82}$$

Similarly, we can substitute the discrete-space phases with continuous fields,

$$\varphi_i(t) \to \varphi(\mathbf{x}, t) \, . \tag{83}$$

In this way we can rewrite the VM as,

$$\eta \partial_t \varphi(\mathbf{x}, t) = Jn_c a^2 \, \nabla^2 \varphi(\mathbf{x}, t) + \zeta(\mathbf{x}, t) \, . \tag{84}$$

where ζ is a Gaussian white noise,

$$\langle \zeta(\mathbf{x}, t)\zeta(\mathbf{x}', t') \rangle = 2\eta \, T \, a^3 \delta^{(3)}(\mathbf{x} - \mathbf{x}')\delta(t - t') \tag{85}$$

and where the factor a^3 is necessary to keep the original physical dimensions once we introduce the spatial Dirac's delta. From Eq. (84), going to Fourier space, we immediately get the dispersion relation of the VM in the fixed network approximation (Appendix E):

$$\omega = i\frac{Jn_c a^2}{\eta}k^2 = iDk^2, \tag{86}$$

where we introduced the 'diffusion' coefficient

$$D = Jn_c a^2/\eta. \tag{87}$$

The structure of the homogeneous equation is therefore that of a diffusion equation and the dynamical propagator for the VM is indeed a diffusive propagator (see Appendix D). A disturbance of the phase at the origin spreads diffusively through the system, and arrives at distances r – damped – after a typical time $t \sim r^2/D$. This is clearly very different from what found in natural flocks where the disturbance, the deviation of the initiator of the turn from the group flight direction, travels distances linearly in time and spreads undamped through the whole flock. Hence, within a fixed network approximation the VM does not produce linear waves. However, the fixed network one is indeed an approximation, so let us ask: does the full Vicsek model, with a moving network, produce linear waves?

3.2.5. Toner–Tu theory and the coupling between density and phase: first sound

The fixed-network approximation of the previous Section may seem unreasonable given that collective motion is all about *motion*, so that particles move following their own velocities and in so doing they change the local density. Hence, it is clear that a continuum limit description of the VM must incorporate density fluctuations. This is what Toner and Tu (TT) did in their remarkable theoretical effort [6,47–49]. The approach developed in these works follows the hydrodynamic perspective: instead of looking at the microscopic dynamics and fields on the scale of the single particles, one defines coarse-grained velocity and density fields over small volumes of the 'active' fluid of particles. The dynamical equations for such coarse-grained fields can be written on the basis of symmetry arguments [6], or by performing in detail the coarse-graining procedure [50–52]. As usual, one expects the coarse-grained fields and the microscopic ones to display the same correlation and response functions at large scales, whereas differences can occur below some hydrodynamic crossover scale. Crucially, since hydrodynamics deals with the large scales, the movement of the network is fully included in the description and the dynamical equations are coupled equations for the orientational (velocity) and positional (density) fields.

The hydrodynamic approach allowed to demonstrate theoretically the occurrence of a transition to the ordered phase, underlying the crucial role of the non-linear terms related to the activity in stabilizing order even in $d = 2$. Here we are interested in understanding the structure of the normal modes predicted by the theory in the low-noise region. To do so, we can simplify the TT equations to only those ingredients that are essential to capture the form of the dispersion law. We therefore disregard the non-linear terms and the terms explicitly breaking the Galilean invariance, as well as the anisotropy

present in the modes in the polarized phase. All these aspects are certainly important in the general theory, but do not change the essence of the TT arguments. We refer the interested reader to [7,3,48] for an exhaustive explanation of the TT hydrodynamic equations and their analysis.

The hydrodynamic equations of Toner and Tu in their simplest form [3] are given by

$$D_t \mathbf{v} = \mathcal{J} \nabla^2 \mathbf{v} - \nabla P - \frac{\partial V}{\partial \mathbf{v}}, \tag{88a}$$

$$\partial_t \rho = -\nabla \cdot (\rho \mathbf{v}). \tag{88b}$$

In these equations, $\mathbf{v}$ and ρ are the velocity and density hydrodynamic fields (i.e. they represent a coarse-grained version of the microscopic fields defined in Appendix A; we use the same symbols in an abuse of notation). $D_t = \partial_t + \lambda \mathbf{v} \cdot \nabla$ is the covariant derivative; the parameter λ breaks the Galilean invariance if $\lambda \neq 1$, which is the case in active fluids; however, since this difference is not relevant to our analysis we will set $\lambda = 1$ in the following. The pressure $P(\rho)$ is a function of the density, ρ_0 is the average density, and the confining potential $V(\mathbf{v}) = \int d\mathbf{r}[-(\alpha/2)v^2 + (\beta/2)(\mathbf{v} \cdot \mathbf{v})^2]$ sets the value of the average fluid velocity in the polarized phase, $\hat{v}_0 = \alpha/\beta > 0$. The parameter $\mathcal{J}$ plays the role of the kinematic viscosity or stiffness depending on whether one views $\mathbf{v}$ as a velocity or an orientation, a difference of interpretation that will play a major role in the following sections. To write the spin-wave expansion of Toner–Tu equations we consider fluctuations around the equilibrium values, $\rho = \rho_0 + \delta\rho$ and $P = P_0 + \sigma\delta\rho$; in this way we obtain [7,3]

$$\partial_t \varphi = \mathcal{J}\nabla^2\varphi - \frac{\sigma}{\hat{v}_0}\partial_\perp \delta\rho, \tag{89}$$

$$\partial_t \delta\rho = -\rho_0 \hat{v}_0 \partial_\perp \varphi, \tag{90}$$

where $\delta\rho$ is the density fluctuation and φ is either the y or z component of the velocity fluctuation, i.e. $\delta\mathbf{v}^\perp = \hat{v}_0(0, \phi_y, \phi_z)$. It is crucial to compare this system of equations with the fixed-network Vicsek equation (84): whereas in (84) there is clearly no propagation, as that is simply the diffusion equation with purely imaginary frequency $\omega = iDk^2$, Eqs. (89) and (90) are different: having taken into account the *motion* of the particles through the introduction of the density fluctuation field, $\delta\rho$, has changed radically the dispersion relation. In the Toner–Tu equations velocity fluctuations are coupled to density fluctuations, so that if we take the second derivative of the phase φ with respect to time we obtain a closed *second-order* dynamic equation for the velocity fluctuations,

$$\partial_t^2 \varphi - \mathcal{J}\nabla^2 \partial_t \varphi - \rho_0 \sigma \nabla_\perp^2 \varphi = 0. \tag{91}$$

If we forget for a moment the strange form of the first-order dissipative term and the anisotropic Laplacian, we see that (91) has the form of a linear wave equation, so that propagating phenomena are to be expected. Indeed, the dispersion relation associated to (91) is,

$$\omega_\pm = -i/\tau_1 \pm c_1 k\sqrt{\sin^2\theta - k^2/k_1^2}, \tag{92}$$

where c_1 is the sound speed, $\tau_1 = 2/k^2\mathcal{J}$ is the damping time, $k_1 = c_1\tau_1 k^2$, and θ is the angle between the wave vector $\mathbf{k}$ and the mean direction of motion. The frequency Eq. (92) has a nonzero real part, and thus propagating waves, in some region of the parameter space; in particular, for $k \to 0$, and $\theta \neq 0$ there is always propagation. The bare speed of propagation is given by

$$c_1^2 = \rho_0 \sigma. \tag{93}$$

Because these modes are carried by density fluctuations, as in standard fluids, we call this propagating mode *first sound*, to distinguish it from the spin-wave modes that we will introduce later. However, it is important to note that these density fluctuations are coupled to velocity fluctuations, hence a propagating perturbation in the density is inevitably associated to a propagating wave in the orientations of the particles.

Hence, the hydrodynamic translation of the VM, namely the theory of Toner–Tu admits propagating phenomena. This seems good news, as linear propagation of velocity fluctuations, with finite sound speed, is exactly what we observe in real flocks (Section 3.1). As we shall see in the next Section, the situation is in fact more complex.

3.2.6. The problems with first sound

The first odd thing we notice in the first sound equations is that $\mathcal{J}$ acts as a damping factor: the ϕ-dependent part of Eq. (89) has the same structure as the diffusion equation, hence $\mathcal{J}$ acts as a diffusion constant, which is why we find it as the prefactor of the first-order dissipative term in (91), $\mathcal{J}\nabla^2\partial_t\varphi$. In the deeply polarized phase we can expect the (coarse-grained) parameter $\mathcal{J}$ to be mainly determined by the microscopic alignment interaction coefficient J. Hence, the first sound theory of Toner and Tu implies that signal propagation is more damped (and thus less effective) the stronger the tendency to get ordered is in the flock. This seems unnatural: at a merely intuitive level we are inclined to think that 'closing ranks' (boosting the alignment order of the group) should not damp signal propagation, but rather strengthen it. One would naively expect that in the correct theory, a larger $\mathcal{J}$ should at least not impair signal propagation. We shall see that this is exactly what happens in the new theory that we will develop in the next sections.

At a more quantitative level, first sound has other problems when compared to empirical signal propagation in real flocks. From Eq. (92) we see that first sound is strongly *anisotropic*: it does not propagate in the direction of motion of the flock ($\theta = 0$, the longitudinal direction) but it is a transverse mode. In fact, a longitudinal mode exists in TT theory, but it is a rather trivial one as its speed of propagation in the laboratory reference frame is approximately the same as the mean speed of the group, meaning that the speed of this mode in the flock reference frame (i.e. the co-moving frame) is negligible [53]. Turning waves in flocks, on the other hand, have a speed that is totally unrelated to the speed of the flock and that, as we have seen, is quite large (up to 40 ms^{-1} *in the flock reference frame*). Such anisotropy of first sound is quite odd at the biological level, as it would imply that a change of direction started by an individual at the front of the flock could not propagate backward to the rest of the group. Although empirical evidence is scant about the direction of propagation of turns [40], a mechanism of propagation that does not work in the longitudinal direction seems unlikely in a real biological system.

Secondly, as we have seen, first sound is carried by density fluctuations, whose coupling with the velocity fluctuations is crucial to give a real frequency; however, the data about turning flocks shows no trace of propagating density fluctuations coupled to propagating velocity fluctuations [36]. In fact, the structure of the network of the individuals is quite stable during the (very fast) propagation of the signal [36]. Moreover, the speed of propagation of first sound, c_1, depends on density, while no such dependence is found in the data: as we have seen in the previous Section, the speed of signal propagation in flocks varies quite significantly from flock to flock, but we have excluded a dependence of this speed on density.

Finally, a crucial feature of signal propagation in the TT theory is its k dependence: we see from (92) that first sound is overdamped (imaginary frequency) at short wavelengths, namely at large k. First sound, and in fact the whole TT construction, is an eminently hydrodynamic theory, valid and universal only at long wavelengths. Because in any finite system of size L the smallest wave vector is of order $1/L$, we conclude that first sound only propagates in *asymptotically large* flocks. This has been confirmed by numerical simulations [47], which succeeded to observe first sound only in *very* large systems ($N = 320\,000$ particles). On the other hand, in small systems k is too large, and first sound does not propagate. Such feature of first sound clashes directly with experimental evidence, which shows that signal propagation in turning flocks occurs clearly also in medium–small flocks, suggesting that this cannot be purely a $k \to 0$ feature, but that it must be sustainable also at intermediate values of the wave number. In fact, even though we do not have empirical data on very large flocks ($N > 10^3$), our field experience while collecting data is that very large flocks are very inefficient in transporting waves, and rather show a 'wobbling' attitude, with very rare sharp changes of direction. At a more general level, we may accept *upper* limits in the maximum size a biological group can attain to sustain linear waves, while it seems very unnatural to accept a *lower* limit in the size: very small groups changing swiftly direction of motion, or propagating other kinds of signal, are a common experience that is very hard to ignore.

To summarize, in Section 3.1 we have presented clear evidence of propagating turning waves in flocks; these waves are linear, very weakly damped, with a variable speed that we could not account for. Simulations show that the simple VM is unable to reproduce these phenomena, at least on the scales and size of realistic flocks. The best theoretical treatment of Vicsek model, namely the Toner–Tu hydrodynamic theory of flocking, displays propagating linear phenomena, but of a kind that does not match the experimental evidence. This is consistent with the inability of simulations of the 3-d VM to reproduce orientational waves. We conclude that a new theory of propagating phenomena is needed to explain the empirical evidence of signal propagation in bird flocks.

3.3. Second sound

As we have seen, the Vicsek model has a diffusive character, Eqs. (81) and (84). A diffusion-like equation is clearly non-ideal for the description of propagating phenomena; however, the network over which this diffusive dynamics takes place moves in time, giving an extra degree of freedom (density fluctuation), which is ultimately responsible for the appearance of non-diffusive modes. It is precisely the strong coupling between velocity and density fluctuations that allows the emergence of propagating modes in the Vicsek model, as clearly described by the hydrodynamic theory of Toner and Tu.

Unfortunately, we have seen that such strong coupling between density and velocity is not observed in real flocks, possibly due to their finite size, whereas TT theory is only valid in the hydrodynamic limit of very large size and very long times: when Vicsek dynamics is observed on medium–short scales, its diffusive character takes over and all propagating TT modes are damped. On the other hand, we have seen in the sections about static correlations that the short-range alignment interaction of the Vicsek model, which translates into the Laplacian of Eq. (84) describes the data rather well, suggesting that this kind of interaction is probably not the main problem with the Vicsek equation. Hence, we need to formulate a new theory that reproduces linear propagating phenomena in *absence* of velocity–density coupling and that it does so even at short–medium scales, namely a theory whose frequency is real not only in the hydrodynamic limit, $k \to 0$.

3.3.1. The key point: the double role of the velocity

To make progress, we must reflect on the double role played by the velocity in the Vicsek model at the conceptual level. Let us go back to the Toner–Tu equation,

$$D_t \mathbf{v} = \mathcal{J} \, \nabla^2 \mathbf{v} - \nabla P + \frac{\partial V}{\partial \mathbf{v}} \,. \tag{94}$$

The last term is specific to active fluids and is responsible for fixing the average fluid speed in the polarized phase. Let us for the moment disregard this term, and focus instead on the others. On one hand, and quite obviously, **v** can be viewed as

a velocity (in fact, it *is* the velocity): according to this interpretation, the l.h.s. of (94) is an acceleration, and thus a *second-order* time derivative, while at the r.h.s. we have the diffusive term, $\mathcal{J} \nabla^2 \mathbf{v}$, which is morally a dissipation, and the force term embodied by the pressure gradient, $-\nabla P$. This is the standard fluid-dynamic view, where the hydrodynamic version of the Vicsek model, Eq. (94), is essentially interpreted as a Navier–Stokes-like equation, that is basically Newton's equation. It is important to note that in this context $\mathcal{J}$ plays the role of kinematic viscosity (or diffusion constant) and for this reason it ends up as a prefactor of the dissipative term in the dispersion relation of TT theory, Eq. (92). This is the ordinary role of $\mathcal{J}$, which is also present for normal (non active) fluids, where it arises from the shear stress divergence. For a fluid of active particles, however, as we have already noted, the dissipative role of $\mathcal{J}$ is odd. In these systems – where we have not considered cohesive forces among the particles – this term arises from the microscopic alignment interactions between the velocities: at low noise we expect $\mathcal{J}$ to resemble the microscopic interaction strength J of the Vicsek model, which we do not intuitively identify with a viscosity. In fact, the term $\mathcal{J} \nabla^2 \mathbf{v}$ is the continuous version of the discrete term $J \sum_j n_{ij} \mathbf{v}_j$ in the original Vicsek model,

$$\partial_t \mathbf{v}_i = J \sum_j n_{ij} \mathbf{v}_j + \cdots \tag{95}$$

But the r.h.s. of this equation obviously has the role of a social *force*, not of a viscosity! This force actually moves and modifies the birds' velocity in order to keep the system polarized. This leads us to a second, and radically different, interpretation of the velocity and of Eq. (94): $\mathbf{v}$ is an orientation vector whose dynamics is regulated primarily by the force $\mathcal{J} \nabla^2 \mathbf{v}$, so that $\mathcal{J}$ is not a viscosity, but a *stiffness* regulating the strength of this orientational force [3,54]. In this alternative view of active dynamics, the velocity is the fundamental degree of freedom of the theory, i.e. its generalized coordinate. So that the correct way to interpret the l.h.s. of (94) is that of a *first-order* time derivative of the principal degree of freedom, whereas at the r.h.s. we have the force, $\mathcal{J} \nabla^2 \mathbf{v}$. Hence, under this second interpretation of the velocity as an orientational degree of freedom, Eq. (94) becomes an *overdamped first-order Langevin equation*, and not an underdamped second-order Newton equation. This view is manifest in the microscopic model, when we remember (as we did on Section 3.2.2, Eq. (75)), that the social force is actually the derivative *with respect to the velocity* of an alignment pseudo-Hamiltonian,

$$\partial_t \mathbf{v}_i = \mathbf{F}_i + \boldsymbol{\zeta}_i \qquad \mathbf{F}_i = -\frac{\delta H}{\delta \mathbf{v}_i} \,. \tag{96}$$

Here the overdamped Langevin structure is evident: the r.h.s. is a social force and indeed it is the derivative of a cost function (the pseudo-Hamiltonian H) with respect to $\mathbf{v}$. This clearly indicates that $\mathbf{v}$ is the actual generalized degree of freedom of interest, but in this view the Vicsek equation describes an *overdamped* theory for the orientation. No surprise, then, that orientation waves do not exist within TT theory as independent modes, but only coupled to propagating density modes.

On the contrary, what we need is a new theory able to support propagating orientational modes irrespective of density fluctuations. To do this, let us (momentarily) brutally simplify the arena: given that we need a theory able to propagate information also in absence of density fluctuations, let us set to zero the pressure term coupling density to velocity, and concentrate solely on the dynamics of the velocity (we will reinstate such term later in Section 3.3.7). In this way, the Vicsek equation (and thus TT theory) becomes,

$$D_t \mathbf{v} = \mathbf{F}, \tag{97}$$

where the force is $\mathbf{F} = \mathcal{J} \nabla^2 \mathbf{v}$. As we have said, this seems a Newton equation if we assign to $\mathbf{v}$ the role of velocity (time derivative of the primary degree of freedom); however, along that path we meet a contradiction, because if $\mathbf{v}$ is a velocity, then $\mathcal{J} \nabla^2 \mathbf{v}$ is a viscosity, not a force. If indeed we want to interpret $\mathcal{J} \nabla^2 \mathbf{v}$ as a force, we must recognize that it is a derivative of a 'Hamiltonian' with respect to $\mathbf{v}$, and therefore $\mathbf{v}$ becomes the primary degree of freedom, not its time derivative.

3.3.2. How viscosity kills inertia and Newton's law becomes an overdamped Langevin equation

To try and fix ideas in this rather confusing situation, let us put aside the velocity for a moment and talk about a generic degree of freedom, i.e. a generalized coordinate, y. This could be anything relevant we choose to describe our system. We have a theory based on the equation

$$\dot{y} \sim F, \tag{98}$$

which is unable to sustain propagating waves for an obvious reason: the r.h.s. (which in our case is a Laplacian) is a *second-order* space derivative of y, while the l.h.s. is a *first-order* time derivative of y, i.e. the structure of the diffusion equation.

In fact, given a generic degree of freedom y and a force F acting on it, Eq. (98) is not in general the correct starting point. Instead, Newton's law states,

$$\ddot{y} \sim F, \tag{99}$$

because force controls the second-order time derivative, not the first. What is the general mechanism by which Eq. (99) reduces to (98)? This is a very basic piece of statistical mechanics and we apologize with the connoisseur, but it is essential to be clear on this point. In both Eqs. (98) and (99) we have disregarded all physical coefficients to convey the general idea, but if we want to understand what really happens we have to be more specific. First of all, we must introduce the coefficient

of the second-order time derivative, that is *inertia*. Because y is a generalized degree of freedom (not necessarily the position) we will call χ the generalized inertia (instead of the usual coefficient m for the mass, which is the mechanic inertia associated to the second-order time derivative of position). Hence we write

$$\chi \ddot{y} = F. \tag{100}$$

This is the correct equation for y in absence of friction. If, however, dissipation is present in the system (and it is always wise to assume so), we must add two terms to the r.h.s.,

$$\chi \ddot{y} = F - \eta \dot{y} + \zeta \tag{101}$$

where η is the viscosity and ζ is the noise. Viscosity and noise are just two sides of the same coin, representing all the interactions with the surrounding environment that we are unable to directly describe with explicit forces; for this reason the correlator of the noise contains η, according to Einstein relation [55]. But let us not linger on this point. Dimensional analysis of Eq. (101) shows that $\chi/t^2 \sim \eta/t$, hence a fundamental time scale immediately emerges:

$$\tau_d \sim \chi/\eta. \tag{102}$$

For $t \ll \tau_d$ or, equivalently, for fixed time but large inertia-to-viscosity ratio, inertial effects dominate; this is the *transient* short-time regime. On the other hand, for $t \gg \tau_d$ or, equivalently, for fixed time but low inertia-to-viscosity ratio, inertia becomes sub-dominant with respect to viscosity; this is the so-called *asymptotic* long-time regime, in which we can therefore disregard inertia and write,

$$\eta \dot{y} = F + \zeta. \tag{103}$$

This asymptotic regime is also called *overdamped*, because damping and friction completely wash out inertia. But it is important to understand that this is a regime defined by time, not simply by the mutual values of viscosity and inertia; in other words, we cannot say that the equation is overdamped because viscosity is large, as there is *always* a time regime (however short this may be) in which inertial effects matter. Hence, overdamping is a propriety that holds in a certain regime, i.e. for times much longer than the ratio χ/η.

We clearly understand now the origin of the Vicsek overdamped equation: the underlying assumption of it is that we are observing the system for asymptotically long times (and, once space is introduced, asymptotically long distances), in which all inertial effects are washed out by dissipation. At first sight, this seems reasonable: after all, who is interested in transients? In physics, we typically never are: we are only interested in very large times and very long distances. However, this is biology, not physics, and when we think about the phenomenon we want to describe, namely the propagation of fast information waves across finite-size groups, we realize that in fact we may very well be interested in transients. Everything about information transfer in biological systems is about transients. Bottom line: we must reinstate inertia in the theory.

3.3.3. Reinstating inertia: symmetry, generators, Poisson brackets and the emergence of spin

So we have to reinstate inertia in the theory. But what kind of inertia are we talking about? Since our primary degree of freedom is the velocity orientation vector, $\mathbf{v}$, not the position, inertia cannot be the standard mechanical mass m. This is a crucial point: we are *not* going to simply recover the mass, i.e. the standard Newtonian inertia of birds. In fact, that kind of inertia is *already* implicitly contained in the theory, because, as we have seen, if we had interpreted $\mathbf{v}$ as the velocity, and thus as the first-order derivative of the primary degree of freedom – space – then the term $\partial_t \mathbf{v}$ in Vicsek equation (and in TT theory) would be the inertial term, the acceleration. Yet we have argued that this is not the right interpretation of $\mathbf{v}$ and that the right role of this vector is that of an orientation subject to a ferromagnetic alignment force with its neighbors. So, again, what kind of inertia do we need in this context?

To make progress it is useful to turn to the Hamilton–Poisson formalism, according to which inertia is the crucial link between the parameter and the generator of a *symmetry*. Let us illustrate this with the simplest symmetry, namely space translations: the coordinate parametrizing translations is the space position q, while the generator of this symmetry is the linear momentum p. The mutual relationship of parameter and generator of the symmetry is formally defined by the Poisson relation,

$$\frac{df}{dq} = \{f, p\}, \tag{104}$$

i.e. the change of any observable f along the transformation parametrized by q is ruled by the Poisson bracket of f with the symmetry generator p. Within this general structure, the canonical pair of coordinate and momentum, (q, p), is regulated by Hamilton dynamics,

$$\dot{q} = p/m, \tag{105}$$
$$\dot{p} = F, \tag{106}$$

which *defines* the inertia, in this case the mechanical mass m. This formalism is useful because it indicates a possible path to build the new theory:

1. identify the fundamental degree of freedom (the generalized coordinate);
2. find the symmetry parametrized by this coordinate;
3. find the generator of this symmetry — this will be the conjugate momentum;
4. write the dynamical equations and define inertia.

We have repeatedly explained that what we need to describe is the dynamics of the *orientations* of the birds. Hence, it seems reasonable to assume that the fundamental degree of freedom is an angle, which clearly parametrizes a rotation. But what angle and what rotation is that? To answer this question we must carefully reflect on the different ways a set of points can turn. There are two ways to turn (to fix ideas we work in a planar context, in which a rotation is parametrized by one angle — we will generalize to 3-*d* rotations later on):

- *Parallel path turning.* Let us consider a set of points (the birds of the flock) on the plane and let us now operate a rotation by an angle θ. This rotation acts on the position of each point with respect to the common origin of coordinates, so the overall effect of the transformation is actually of rotating the whole flock, namely a *turn*. This symmetry is generated by angular momentum, l, and parametrized by the angle θ. However, if we look at the trajectories drawn by the particles during this kind of turn, we see that there is something wrong: this is a *parallel path* turn, also known as rigid assembly turn (Fig. 10) in which the particles' trajectories do not cross, but the radius of curvature (with respect to the origin) is different from bird to bird. This kind of turn is correct for a rigid body, but it is definitely not what happens in animal groups: during parallel path turning particles on the external side of the turn have a speed *larger* than particles on the internal side of the turn, a fact not biologically plausible, and which becomes completely absurd as the group grows in size. In all biological groups, individual speeds can fluctuate, but only within physiologically reasonable ranges, certainly not proportionally to the turning radius.
- *Constant radius turning.* In fact, we have seen in Section 3.1.4 that not only common sense, but real experiments show that flocks (both of pigeons [41] and starlings [14,40]) do not turn with parallel paths, but with *constant radius* turns: in this kind of turn, each particle has the same radius of curvature, and therefore the same speed; as a consequence paths cross (Fig. 11), a scheme impossible for a rigid body, but obviously ideal for a biological group. Hence, experiments themselves clearly indicate that the symmetry we need to describe is the rotation giving rise to equal radius turning. This kind of rotation is *not* described by the canonical pair (θ, l) (which generate parallel path turning). What is the correct generator–parameter pair in the case of constant radius turning? In order to attain equal radius rotations we must rotate the direction of motion of each bird i, which is what we call *phase*, φ_i. Unlike θ, a rotation of φ_i changes only the orientation of the velocity of i, not its position. The phase is thus our fundamental degree of freedom and it parametrizes the *internal* rotations, i.e. the rotation within the internal space of the order parameter $\mathbf{v}_i$, as opposed to the rotation in the external space of coordinates (or world sheet). We must now identify the generator of this symmetry; inasmuch φ is different from θ, the generator we are looking for must be different from the angular momentum l. Perhaps surprisingly, some help comes from quantum mechanics, which has a simple name for the generator of the rotation in the internal space of the order parameter: it is the *spin*. We will use this same name, and indicate it with the symbol s. We conclude that the canonical pair of coordinate and conjugated momentum, (φ, s).

Mathematically, the relation between phase and spin is as usual expressed by Poisson's bracket,

$$\frac{df}{d\varphi} = \{f, s\}. \tag{107}$$

The canonical dynamics of phase and spin defines the *inertia*, χ:

$$\dot{\varphi} = s/\chi, \tag{108}$$
$$\dot{s} = F \tag{109}$$

where the effective force F is actually an effective torque. It is important to understand that χ is a generalized inertia: it is not the mechanical mass, m, nor the mechanical moment of inertia, I; χ is effectively defined by (108). To understand its meaning we merely proceed as in textbooks and write,

$$\ddot{\varphi} = F/\chi \tag{110}$$

from which we clearly see that the generalized inertia χ embodies the resistance of the active particle to a change of $\dot{\varphi}$, namely to a change of its turning rate, caused by a social force (or social torque, actually), F. The ingredients of this resistance, both mechanical and neural, are certainly complicated and it is safe to assume that χ is an effective (or phenomenological) parameter. For this reason in [36] we called χ *behavioral inertia*. We invite the reader to see the Supplementary Information of [36] for a thorough discussion of the link between χ and the flight parameters of a bird.

We stress again that phase and spin is a *different* pair from angle and angular momentum. The Poisson relation between the latter is,

$$\frac{df}{d\theta} = \{f, l\} \tag{111}$$

and their dynamics is governed by the equations,

$$\dot{\theta} = l/I, \tag{112}$$

$$\dot{l} = F. \tag{113}$$

The inertia in this case is the standard moment of inertia I. As we have repeatedly explained above, Eqs. (112) and (113) describe parallel path turning, which is *not* the relevant case in biological systems.

Hence, we have finally reinstated inertia in the Vicsek model of mutual imitation. Let us see what are the practical consequences on the information transfer mechanism.

3.3.4. The conservation of the spin and the emergence of second sound

We can now recall that in the Vicsek model and in the TT equation the 'social force', F, (or social torque) is in fact the Laplacian of the velocity, $\mathcal{J}\nabla^2 \mathbf{v}$, which expresses the tendency of each bird to align with its neighbors; under the spin wave expansion this term becomes simply the Laplacian of the phase, $\mathcal{J}\nabla^2\varphi$ (see Eq. (89)). Hence, from (108) and (109) we obtain the new equations for the spin and phase fields,

$$\dot{\varphi}(\mathbf{x}, t) = s(\mathbf{x}, t)/\chi, \tag{114}$$

$$\dot{s}(\mathbf{x}, t) = \mathcal{J}\nabla^2\varphi(\mathbf{x}, t). \tag{115}$$

We notice that the second equation can be rewritten as a continuity equation for the spin: if we define the current $\mathbf{j} = \mathcal{J}\nabla\varphi$, (115) becomes,

$$\frac{\partial s(\mathbf{x}, t)}{\partial t} - \nabla \cdot \mathbf{j}(\mathbf{x}, t) = 0, \tag{116}$$

from which we conclude that the total spin is *conserved*:

$$\frac{d}{dt} \int d\mathbf{x}\, s(\mathbf{x}, t) = 0. \tag{117}$$

The conservation of the spin is a consequence of the symmetry of the problem under the rotations parametrized by the phase, φ. By taking a second derivative of the phase in (114) we get,

$$\frac{\partial^2\varphi}{\partial t^2} - \frac{\mathcal{J}}{\chi}\nabla^2\varphi = 0, \tag{118}$$

which is nothing else than D'Alembert equation of wave propagation, stating that a signal generated as a disturbance of the phase, i.e. of the direction of motion of one bird, travels linearly across the flock with a speed

$$c_2 = \sqrt{\frac{\mathcal{J}}{\chi}}. \tag{119}$$

Several remarks are in order here. First, we have assumed that the network is fixed, hence density fluctuations cannot participate in this signal propagation; we will reinstate density fluctuations later, but for now this proves that the propagation mechanism described by (118) is not due to the coupling between density and phase, but rather to the coupling between spin and phase, namely between the momentum and its canonically conjugated coordinate. Second, because density is out of the question, this linear sound mode is radically different from the first sound of Toner–Tu equations. We will discuss this in detail later, but for now it is sufficient to notice the completely different expression of c_2 with respect to first sound speed (93); in particular, it is crucial to notice that this new sound mode propagates faster in flocks with a stronger alignment interaction, $\mathcal{J}$, while first sound speed does not depend on $\mathcal{J}$ and in fact, as we have seen, the alignment acts as a damping of first sound. Third, the theory we have introduced is mathematically identical to a spin wave model introduced in the 1960s for the description of superfluid Helium, where φ is the phase of the complex wave function of the particles, which is the order parameter of the system [56–58]; in that context, the propagating mode of the phase is called *second sound*, to distinguish it from the standard (first) sound coupled to density waves. Given our need to distinguish these phase modes from the first sound of Toner and Tu, in [36] we have adopted the same nomenclature, hence c_2 is the second sound speed.

The new theory is very simplified and it needs of course some adjustments, chiefly the introduction of some dissipative term (we passed from a completely overdamped equation to a completely undamped one, which is unrealistic) and a generalization of it from the level of the phase (which is only meaningful in the context of the spin wave expansion, namely for polarized systems), to the more general level of the velocity $\mathbf{v}$. However, before we take care of these aspects, there is an immediate prediction the new theory makes, which is worth seeing. The second sound speed, Eq. (119), must match quantitatively the speed of propagation of the signals that have been experimentally observed in real flocks. We saw in Section 3.2.6 that previous theories are unable to do this: a purely diffusive dynamics does not even have a linear speed of propagation, while first sound, Eq. (93) predicts a speed that fundamentally depends on the density, which is not experimentally the case. In fact, we could not account in any way, not even phenomenologically, for the wide variations of propagation speed that we observe in real flocks; we can therefore say that this experimental fact was *not* one of the inputs

of developing the new theory. Now we have (119): does it work? The behavioral inertia, χ, is a phenomenological parameter and we do not know its value: calculating it on the basis of an *ab initio* mechanical calculation is out of question and the only way we could infer it is from c_2, which would of course beg the question. However, it seems reasonable to assume that χ is on average approximately the same from flock to flock, so that the different observed values of c_2 are not due to differences in the behavioral inertia. The stiffness $\mathcal{J}$, on the other hand determines how strongly birds tend to align to each other and therefore it must have an effect on the global polarization of the flock. We conclude that second sound should be faster in more polarized flocks: alignment's strength boosts signal propagation. This sounds reasonable, especially compared to the prediction of first sound, according to which alignment's strength depresses signal propagation. More importantly, this is a prediction we can check quantitatively.

In Appendix B, Eq. (B.9), we have derived an explicit relation between polarization and phase fluctuations within the spin wave expansion, namely

$$\Phi = 1 - \langle \varphi^2 \rangle. \tag{120}$$

On the other hand, we have seen in Section 2.4.1 that within the spin wave approximation the distribution of the phases is Gaussian, Eq. (48). In the present continuum space context we can write,

$$P(\varphi) \sim \exp\left(-\frac{1}{2}\beta \int \frac{d^3k}{a^3} \mathcal{J}(\nabla\varphi)^2\right), \tag{121}$$

from which we have

$$\langle \varphi^2 \rangle \sim \frac{1}{\beta \mathcal{J}}. \tag{122}$$

From (120) and (122) we finally obtain

$$\mathcal{J} = \frac{T}{1 - \Phi}, \tag{123}$$

where $T = 1/\beta$ is the temperature, in a static sense (see Section 2.4 on maximum entropy). Eq. (123) states the intuitive fact that, in the polarized phase, the only way to push the polarization to exactly 1 at finite temperature is to have infinite alignment strength $\mathcal{J}$. By plugging this relation into (119), we obtain a nontrivial dependence of second sound speed on polarization, first obtained in [36],

$$c_2 \sim \frac{1}{\sqrt{1 - \Phi}}. \tag{124}$$

This is the central prediction of the new theory and experimental data show that it is verified (Fig. 14 and [36]). Notice that (124) is a nonlinear, dimensionless function of Φ that was hard to guess *a priori* merely on the basis of the flock-to-flock variation of the speed of propagation across the experimental data. It is also a hard quantitative fact not reproduced by other theories. We conclude that inertia is indeed essential to account for the experimental evidence about information propagation in flocks.

Eq. (124) is important for a second reason: it provides the most compelling explanation about why network rearrangements are scarcely relevant for second sound. According to (124) the speed of propagation of second sound is larger the larger the polarization, Φ: more ordered flocks (in which, thus, the stiffness $\mathcal{J}$ is larger) sustain faster phase waves. Hence, the time scale needed for the signal to cross the system, and therefore the time scale of second sound propagation, τ_{prop}, *decreases* when the polarization increases. On the other hand, let us ask ourselves: how does the time scale over which some substantial rearrangements of the interaction network occur depend on polarization? For low Φ, everybody is going in a different direction, hence the rearrangement will be quick; conversely, if the polarization is very close to 1, the fluctuations of each individual with respect to the mean velocity will be very small; but these fluctuations are responsible for the mutual diffusion of the individual, which is the origin of network rearrangement. This means that the time scale for network rearrangement, τ_{netw}, *increases* when the polarization increases. We conclude that these two time scales have opposite limits for large polarization,

$$\lim_{\Phi \to 1} \tau_{\text{prop}} = 0, \qquad \lim_{\Phi \to 1} \tau_{\text{netw}} = \infty. \tag{125}$$

Given (125), it is natural to expect that, at least in polarized systems, the two time scales separate sharply, so that second sound propagation can be studied under the assumption of fixed interaction network. As we have seen, this is exactly what seems to be happening at the experimental level in real flocks, which do not significantly change their network wiring during turns.

In fact, this argument can be generalized beyond information transfer to the more general level of the correlation between fluctuations in flocks. The time scale of network rearrangement can be compared to a third time scale, namely the local relaxation time of the velocities, τ_{relax}. When this is done [8], one discovers that $\tau_{\text{prop}} \gg \tau_{\text{relax}}$, hence the local degrees of freedom are relaxed well before the network has dynamically changed. This is the fundamental reason behind the success of pseudo-equilibrium theories for these systems, as those we have described in the maximum entropy section. The interested reader can learn more about this topic in [8].

3.3.5. Dissipation, quasi-conservation of the spin and the cut

The new theory for the phase that we have built up to now lacks a crucial ingredient, namely dissipation. The dissipative term of the former theory, $\mathcal{J}\nabla^2\varphi$, now acts as a social force; accordingly, $\mathcal{J}$ is now a stiffness. A consequence of having reinstated inertia is that spin is conserved, meaning that it can only be transferred from bird to bird (this is exactly how a disturbance of the direction of motion propagates across the flock in the new theory), but cannot be dissipated locally. However, this means that a single noninteracting bird would maintain its turning state forever, which is not very reasonable. Clearly, to solve this paradox we need to introduce a weak dissipative term, slowly driving to zero the spin in absence of any perturbation, so that the unperturbed state of motion of a bird is to fly straight. At the same time, we need to recognize that this can only be true on average: small adjustments around the direction of motion are always likely. In fact, it is intuitive to expect that the many factors leading to a slow dissipation of the spin are also responsible for its random fluctuations. This is the classic duality between *dissipation* and *noise* typical of statistical mechanics: they are just two sides of the same coin, their common origin being the unknown interaction with the environment (i.e. the heat bath). Here, we follow an identical approach and write

$$\dot{\varphi} = s/\chi, \tag{126}$$

$$\dot{s} = \mathcal{J}\nabla^2\varphi - \frac{\eta}{\chi}s + \zeta, \tag{127}$$

where η is a generalized friction coefficient and ζ is a white noise satisfying

$$\langle \zeta(\mathbf{x}, t)\rangle = 0, \tag{128}$$

$$\langle \zeta(\mathbf{x}, t)\zeta(\mathbf{x}', t')\rangle = 2\eta T\delta(\mathbf{x} - \mathbf{x}')\delta(t - t'), \tag{129}$$

where T is the analogue of the temperature in physical systems, namely the parameter linking noise and viscosity. The spin is no longer conserved, but if the friction η is small enough there is still linear propagation in the system. Let us see this explicitly. The deterministic part of the equation for the phase now becomes

$$\chi\frac{\partial^2\varphi}{\partial t^2} - \mathcal{J}\nabla^2\varphi + \eta\frac{\partial\varphi}{\partial t} = 0. \tag{130}$$

The dispersion relation (Appendix D and Appendix E) of this equation is

$$\chi\omega^2 - i\eta\omega - \mathcal{J}k^2 = 0. \tag{131}$$

By dividing by χ and introducing the reduced friction

$$\gamma = \frac{\eta}{2\chi} \tag{132}$$

we can rewrite the dispersion relation as

$$\omega^2 - 2i\gamma\omega - c_2^2 k^2 = 0, \tag{133}$$

where c_2 is the second sound speed introduced before. From this we can solve for the frequency:

$$\omega = i\gamma \pm c_2 k\sqrt{1 - k_0^2/k^2}, \qquad k_0 = \gamma/c_2. \tag{134}$$

If $\eta = 0$, the dispersion relation is trivial: $\omega = \pm c_2 k$, the frequency is purely real and there is linear propagation. The crucial parameter to understand the effect of dissipation is k_0, which is an inverse length scale, and must be compared to the wavenumber k. For $k > k_0$ (which happens either for small wavelength, or small dissipation, or large stiffness), the square root is real, so that the frequency has a propagating part; in this case the imaginary part is constant, giving a damping time equal to $\tau = 1/\gamma$. On the other hand, for $k < k_0$ (large wavelength, large dissipation, low stiffness) the frequency has no real part: all modes are overdamped, and they do not propagate (these kind of modes are sometimes called evanescent waves). Hence, the dispersion relation has a cut at $k = k_0$: above this cut modes propagate and the equation is underdamped (in fact, for $k \gg k_0$ propagation is as good as undamped), below the cut modes are overdamped. So we see that the overdamped limit of the theory is not only a large friction limit, but is also a small k limit, i.e. the *hydrodynamic limit*: even a very small dissipation takes over in this limit. As we shall see in the next section, in this limit the new theory goes back to the old one, namely to the Vicsek model and Toner and Tu theory.

Before that, we notice that the cut k_0 immediately sets a limit on the maximum size of a flock across which a linear signal can propagate within this theory: in order to transfer immediately useful information (Section 2.5.1) across a system of finite size L we need to be able to propagate the mode with wave number $k = 1/L$, but this also has to be above the cut, lest it be overdamped. We conclude that we must have,

$$L < 1/k_0. \tag{135}$$

In this way we can talk about overdamped vs. underdamped case in relation to the system's size, rather than to the wavenumber k, which is perhaps more intuitive. When the system is such that $L < 1/k_0$ propagation is effectively linear,

with constant damping γ. Notice that this does not necessarily imply a very small system: if η is small and $\mathcal{J}$ large, the cut k_0 may be very very small, so that even large flocks may enjoy linear and weakly damped propagation of phase waves. This seems to be the case in real flocks, as the data presented in Section 3.1 show. On the other hand, unlike in first sound, the small L limit is clearly always a propagating regime for second sound; this seems to meet our intuition that coordinated changes of direction are *easier*, rather than more difficult, in small flocks.

It is interesting to obtain the same constraint on the size of the flocks by comparing time scales, rather than length scales. When there is propagation, the speed is c_2, hence the time the signal needs to cross the system is L/c_2. On the other hand, due to dissipation, the signal gets damped in a time $1/\gamma$. Clearly, we need the signal to reach the entire system *before* it gets damped, hence we impose the condition,

$$L/c_2 < 1/\gamma \tag{136}$$

By using the expressions for k_0 as a function of γ and c_2 we obtain exactly the same constraint as (135).

To summarize, the second sound theory we have introduced is able to sustain linear propagation of the disturbances in the direction of motion even in absence of density fluctuations, but not in the hydrodynamic limit, $k \rightarrow 0$. In the next sections we will reinstate the motion of the network and therefore density fluctuations; we shall see that in going from large to small k the theory will cross over from second to first sound.

3.3.6. The Inertial Spin Model (ISM)

Let us now proceed to build a complete theory, which also takes into account density fluctuations. As we have discussed at the beginning of this section, this can be done following two different strategies: writing equations for the individual particles (i.e. as in the Vicsek model), or looking at coarse-grained fields (as in the TT equations). In Section 3.3.4 we have been discussing the role of inertia in terms of fields (see Eqs. (114) (115)): in this context our aim is to write coupled equations for the velocity field, the spin field, and the density field. Before doing that, however, it is convenient to go back to a description in terms of individual particles and write a self-propelled particle model, generalization of the VM, which incorporates the inertial dynamics described above. To do so, we notice that we can write equations analogous to (114)(115) also for the phases and spins of the individual particles (remembering that $\nabla^2 \rightarrow -\sum_j \Lambda_{ij}$ - see Eq. (82))

$$\dot{\varphi}_i = s_i/\chi, \tag{137}$$

$$\dot{s}_i = -J \sum_j \Lambda_{ij}\varphi_j - \frac{\eta}{\chi}s_i + \zeta_i, \tag{138}$$

where now J is the microscopic interaction strength and we have a discrete rather than a continuous Laplacian. In [46] it is shown how to go from φ_i to $\mathbf{v}_i$ by using Poisson brackets. Here we will simply 'read' the equations for $\mathbf{v}_i$ directly from those for φ_i and s_i by appealing to the similarity with standard rotatory motion. In that case, Eqs. (112) and (113) give rise to the standard equations for the position $\mathbf{r}$,

$$\frac{d\theta}{dt} = \frac{\delta H}{\delta l} \qquad\qquad \longrightarrow \qquad\qquad \frac{d\mathbf{r}}{dt} = \mathbf{r} \times \frac{\delta H}{\delta \mathbf{l}}, \tag{139}$$

$$\frac{dl}{dt} = -\frac{\delta H}{\delta\theta} \qquad\qquad \longrightarrow \qquad\qquad \frac{d\mathbf{l}}{dt} = -\mathbf{r} \times \frac{\delta H}{\delta \mathbf{r}}. \tag{140}$$

In complete analogy with this, using the correspondence, $\mathbf{r} \rightarrow \mathbf{v}_i, I \rightarrow \chi, \mathbf{l} \rightarrow \mathbf{s}_i$, we obtain

$$\frac{d\mathbf{v}_i}{dt} = \frac{1}{\chi}\mathbf{s}_i \times \mathbf{v}_i, \tag{141}$$

$$\frac{d\mathbf{s}_i}{dt} = \frac{\mathbf{v}_i}{v_0} \times \left(\frac{J}{v_0} \sum_i n_{ij}\mathbf{v}_j - \frac{\eta}{v_0}\frac{d\mathbf{v}_i}{dt} + \zeta_i \right), \tag{142}$$

$$\frac{d\mathbf{r}_i}{dt} = \mathbf{v}_i, \tag{143}$$

where the third equation expresses the fact that the system is active; hence, the network of interaction now *does* depend on time, namely $n_{ij} = n_{ij}(t)$. This system of equations defines a new model of collective motion, which can be readily simulated numerically; it was introduced in [46] and, due to the crucial role played by the behavioral inertia χ and by the spin $\mathbf{s}$ it has been called *Inertial Spin Model* (ISM).

It is straightforward to check [46] that the spin wave expansion (large polarization limit) of (141) and (142) coincides with the equations for the phase and spin discussed above, Eqs. (137) (138); hence, the ISM contains all the linear propagation features that we need to describe waves in real flocks. This can be clearly seen with numerical simulation. We have seen in Section 3.2.3 that an attempt to propagate a local change of direction within a Vicsek flock is not successful: the pilot bird turns, but the rest of the flock fails to follow it. One can perform the same numerical experiment using the ISM: in the ordered phase of the flock, we select one particle (say, around the center of the flock, as we did for Vicsek) and turn it by 90 degrees. What happens in this case? The video provided in Ref. [46] speaks more than a thousand words, but the same quantitative message can be seen in Fig. 16: because propagation is now linear, the phase disturbance has a chance to travel *faster* than the
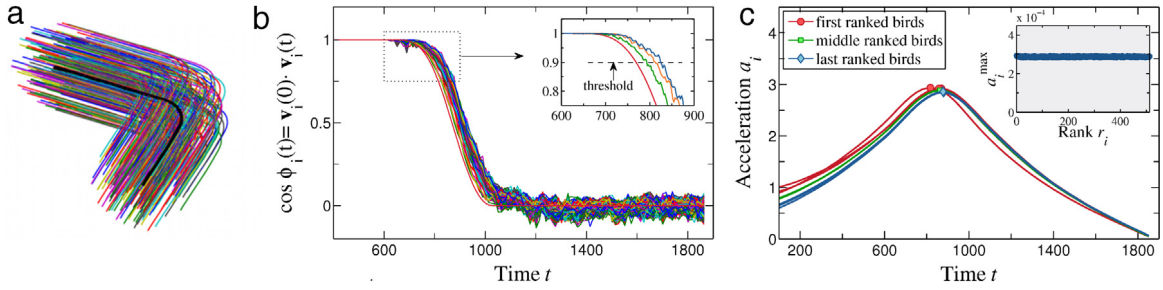
Fig. 16. Propagation of turn information in the 3-*d* ISM model, in the deeply underdamped regime. A direction change is imposed on one individual (black trajectory), following the same procedure as in Fig. 15. Top panel: 2-*d* projection of trajectories. Middle panel: cosine of the individual velocities with respect to the original direction (before forcing the turn). Lower panel: individual acceleration profiles.
Source: From Ref. [46].

pilot bird; hence, provided that $c_2 = \sqrt{J/\chi}$ is large enough (namely, larger than v_0, which happens if the stiffness is large), the turn propagates coherently to the entire flock. All birds change direction giving rise to a global collective turn. Notice that the time the signal takes to cross the flock is much shorter than the time needed to significantly rearrange the network, hence this is a phenomenon that effectively happens over a fixed network, even though the ISM equations (141)–(143) *do allow* for the movement of the network.

The linearly traveling waves we have just described occur when the ISM equations are underdamped, namely for large stiffness J and low friction η. On the other hand, in the overdamped limit, namely when the effective friction η is large compared to the behavioral inertia χ, the ISM equations reduce exactly to the Vicsek model [46]. This is consistent with what we have seen above: the inertial theory has a cut, which can be crossed either by increasing friction over inertia, or by increasing the wavelength; beyond this cut the theory enters into the overdamped regime, identical to the Vicsek model, in which second sound is completely damped and non-propagating. As we have seen in Section 3.2.3 numerical simulations of this case show that a local change of direction does not propagate at all across the flock.

3.3.7. A single field theory for both first and second sound

So the ISM has a linearly propagating second sound mode in its underdamped regime, while this mode does not propagate in the overdamped regime, where the ISM reduces to the Vicsek model. However, we have shown in Section 3.2.5 that the Vicsek model, in the continuous formulation of Toner–Tu, *does* have a propagating mode, first sound. Hence, we expect to recover first sound in the overdamped limit of a continuous version of the ISM, with the mild paradox that this regime is overdamped for what concerns second sound, not first. We have to understand how one crosses over from second to first sound propagation by increasing friction (or wavelength) in the inertial equations. To do this we need to write the field counterpart of Eq. (143), namely we need to write a set of field equations that include the density fluctuations.

Applying to (141)–(143) the same strategy as Toner–Tu for turning a discrete self-propelled dynamical system into a set of continuous field-theoretical equations, we can write [59],

$$D_t \mathbf{v} = \frac{1}{\chi}\mathbf{s} \times \mathbf{v} - \nabla P - \frac{\delta V}{\delta \mathbf{v}}, \tag{144}$$

$$D_t \mathbf{s} = \frac{\mathbf{v}}{v_0} \times \frac{\mathcal{J}}{v_0}\nabla^2 \mathbf{v} - \frac{\eta}{\chi}\mathbf{s}, \tag{145}$$

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{v}), \tag{146}$$

where we recall that P is the pressure and V is a nonquadratic potential for the velocity, working effectively as a soft constraint on its modulus v_0. This is a continuous dynamical field theory which includes both density fluctuations and inertial effects; it has an additional field (spin) apart from the velocity and density fields of the original Toner–Tu theory. In the large friction limit, though, the spin drops out of the equations and we recover Toner–Tu equations (88).

Hence, Toner–Tu theory is the overdamped limit of the inertial field equations (144)–(146). Calling this regime 'overdamped' may seem odd, since the Toner–Tu equations also have propagating modes (Section 3.2.5). To avoid this paradox we should call this limit the *spin-overdamped* limit, as this is what really happens in the large η/χ limit (i.e. second sound, namely propagating spin disturbances, are killed). However, in the regime where spin fluctuations are overdamped, density fluctuations are not, and because of their coupling with velocity fluctuations they give rise to first sound. We have seen this mechanism in Section 3.2.5, leading to the first sound frequency (92). On the other hand, in the limit of negligible density fluctuations, the field $\rho(x, t)$ drops out of the calculation and we get the 'spin-underdamped' dispersion relation of Section 3.3.5, Eq. (134). To bridge the gap between these two limits, one has to write the full dispersion relation of (144)–(146), which has been done in [59]. In this general case there are three different fields, hence (after linearization
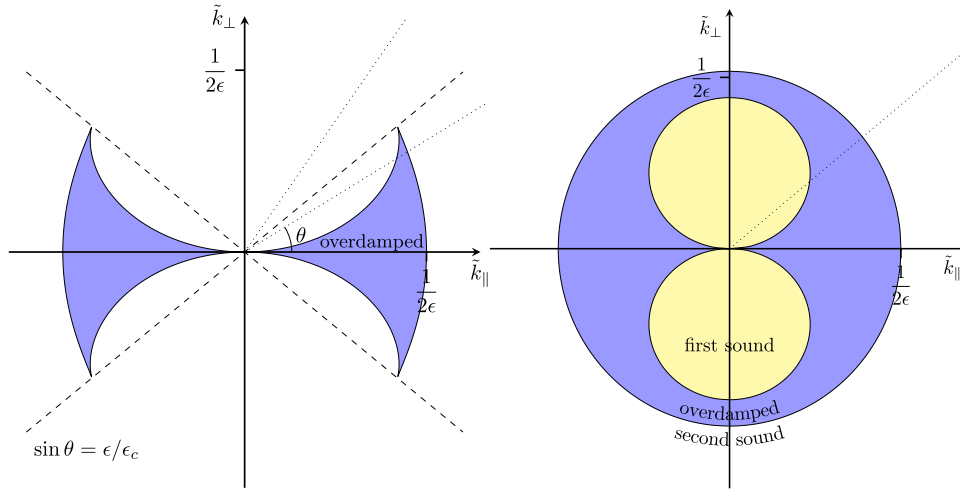
Fig. 17. Appearance of a gap in the dispersion relation (147) for the hydrodynamic theory including first and second sound. The axis represent the components $k_\parallel$ and $k_\perp$ of the wavevector **k**, respectively parallel and perpendicular to the direction of motion. The blue regions indicate the overdamped regions where the frequency is purely imaginary and there is no propagation. Left: the case $\epsilon < \epsilon_c$, where the overdamped region has the shape of two rounded wedges, and a gap appears only for **k** in directions near the parallel axis $k_\parallel$. First and second sound are not well separated, and hybridization occurs. Right: the case $\epsilon > \epsilon_c$. In this case the overdamped region separates the plane into regions where only first sound (yellow) or second sound (white) can propagate, separated by a gap of nonpropagating (imaginary frequency) wavevectors.
Source: From Ref. [59].

of the equations), three degrees of freedom. The dispersion relation is thus of the third order and it reads [59],

$$\tilde{\omega}^3 + i\tilde{\omega}^2 - \tilde{k}^2(\sin^2\theta + \epsilon^2)\tilde{\omega} - i\tilde{k}^2\sin^2\theta = 0 \tag{147}$$

where $\tilde{\omega}$ and $\tilde{k}$ are the dimensionless frequency and wave number respectively [59], whereas $\epsilon = c_2/c_1$ is the ratio between second and first sound, a key parameter many of the propagation properties of the theory depend on. The mathematical details of the general dispersion relation (147) are quite intricate and are discussed in [59]; here we simply summarize the final result. There is a critical value of the speed ratio, $\epsilon_c = \sqrt{8}$, separating two different regimes. For $\epsilon < \epsilon_c$ (first sound speed large compared to second sound speed), there is a hybridization of first and second sound, i.e. of the Toner–Tu and ISM propagating modes, in a large part of the **k** plane. On the other hand, for $\epsilon > \epsilon_c$ these two modes become completely separated: at small k we only find first sound propagating modes, at large k only second sound modes and between these two regions there is a *gap* where no propagating modes exist (purely imaginary frequency). These results are depicted in Fig. 17.

The situation for $\epsilon > \epsilon_c$ has two interesting consequences: first, the gap in k implies that no sound, be it first or second, can propagate in medium-sized flocks [59]. Second, the fact that second sound propagates faster than first sound means that inertial phase disturbances travel much faster than density fluctuations, and for this reason it seems reasonable to expect to have little or no density fluctuations during these second sound waves; the mathematical implication is that the fixed network assumption seems reasonable when we study second sound, while it is clearly not so when we deal with first sound, which is explicitly carried by density fluctuations.

4. Space–time correlations

It should by now be clear that efficient propagation of information across the group can be key to its survival, and that the most efficient realistic mechanism one can imagine is one that allows transfer of information with constant speed and minimal dissipation. Wave propagation is precisely one such mechanism, which is rendered possible due to the existence of behavioral inertia. This mechanism was formulated in the Inertial Spin Model (ISM) after direct experimental observation of wave fronts propagating in starling flocks. However, if behavioral inertia is present in the dynamics, it will play a role in situations where there are no visible propagating fronts. This may be because one observes a group where no external perturbation has been applied (say an undisturbed flock of birds) or a group which is naturally disordered so that even when a perturbation is applied its propagation is very hard to follow (such as an insect swarm). Hence it is important to find a way to assess, given some experimental observations of a group, whether its dynamics is ruled by inertia or not. This is what we discuss in this section.

The tool we use follows from an idea that is natural in the context of linear response theory of equilibrium systems: instead of studying how the system responds to an external perturbation, we focus on its unperturbed spontaneous *fluctuations:* we are then able to tell whether or not a system is inertial by sampling its dynamics, rather than manipulating it. The

mathematical tool is the space–time correlation function that measures how the behavioral change of individual i at time t_0 influences that of individual j at a later time $t_0 + t$.

We will first examine the behavior of the time correlation in the ordered phase of the Vicsek model (VM) and of the ISM and explore the differences introduced by the inertial dynamics, and then provide a definition of the space–time correlation suitable for the analysis of experiments and numerical simulation.

4.1. The space–time correlation of velocity fluctuations

The actual correlation functions suitable to study experimental or simulated systems are a direct generalization of those defined in Section 2.1.3. We define a correlation function in space and time that compares fluctuations at different places and times:

$$C(\mathbf{r}, t) = \left\langle \sum_{i,j}^{N} \frac{\delta\mathbf{v}_i(t_0) \cdot \delta\mathbf{v}_i(t_0 + t)\delta\left[\mathbf{r} + \mathbf{r}_i(t_0) - \mathbf{r}_j(t_0 + t)\right]}{\sum_{k,l}\delta\left[\mathbf{r} + \mathbf{r}_k(t_0) - \mathbf{r}_l(t_0 + t)\right]} \right\rangle_{t_0}, \tag{148}$$

where positions are computed in the center-of-mass reference frame, $\mathbf{r}_i(t) = \mathbf{R}_i(t) - \mathbf{R}_{\mathrm{CM}}(t)$, and the bracket indicates an average over the time origin,

$$\langle f(t; t_0)\rangle_{t_0} = \frac{1}{T_{\max} - t} \sum_{t_0=1}^{T_{\max}-t} f(t; t_0), \tag{149}$$

and $T_{\max}$ is the total time of the signal. To remain as general as possible, in Eq. (148) we consider the case where the correlation can depend on the full vector $\mathbf{r}$. In an isotropic system, $C(\mathbf{r}, t)$ will be independent of the orientation of $\mathbf{r}$; in this case it is more convenient to compute the isotropic version of (148),

$$C(r, t) = \left\langle \sum_{i,j}^{N} \frac{\delta\mathbf{v}_i(t_0) \cdot \delta\mathbf{v}_j(t_0 + t)\delta\left[r - r_{ij}(t_0, t)\right]}{\sum_{k,l}\delta\left[r - r_{kl}(t_0, t)\right]} \right\rangle_{t_0}, \tag{150}$$

where $r_{ij}(t_0, t) = |\mathbf{r}_i(t_0) - \mathbf{r}_j(t_0 + t)|$. For $t = 0$, this definition reduces to the static correlation discussed before (Eq. (6) in Section 2.1.3).

The sum rule (17) constrains the time correlation as well. The argument of Section 2.3.2 applies for all t (see Section 4.2), so that $C(\mathbf{r}, t)$ must have at least one zero as a function of $\mathbf{r}$ for all times.

The isotropic correlation (150) (or its Fourier space counterpart Eq. (152) introduced below) is suitable for systems where orientations are indistinguishable, such as swarms, where there is no net collective displacement. Flocks, on the other hand, are not isotropic: the existence of a nonzero polarization breaks rotational invariance and the average velocity defines a special direction (the *longitudinal* direction), different from the *transverse* directions (those lying in the plane orthogonal to the average direction of motion). Two kinds of anisotropies arise: One is that the transverse fluctuations of velocity are far stronger than the longitudinal ones, leading to different susceptibilities in the two directions, and implying that the correlations of $\delta\mathbf{v}_i$ are actually dominated by the transverse correlations [60,61]. This anisotropy is present also in equilibrium physical systems such as magnetic materials. The second kind is particular to active systems with spontaneously broken symmetry, and is caused by a feedback between velocity and position that produces an anisotropic decay of the correlation [47,48].

Thus in principle one should use (148) to study space–time correlations in flocks. However, this is true if one attempts a detailed characterization of fluctuations. In Section 4.4 we will use correlations to attempt to distinguish between different dynamics (inertial vs. non-inertial). As we will show, it turns out that for this purpose the isotropic definition (150) is sufficient. This is rather fortunate, as in most real biological data it is very difficult to obtain enough statistics to be able to separate the longitudinal and transverse components of the correlation.

4.2. Space–time correlations in Fourier space

We define the Fourier-space counterpart of (148) analogously to (35):

$$C(\mathbf{k}, t) = \frac{1}{N}\left\langle \sum_{i,j}^{N} \delta\mathbf{v}_i(t_0) \cdot \delta\mathbf{v}_j(t_0 + t)e^{-i\mathbf{k}\cdot[\mathbf{r}_i(t_0)-\mathbf{r}_j(t_0+t)]} \right\rangle_{t_0}. \tag{151}$$

In the isotropic case $C(\mathbf{k}, t)$ does not depend on the direction of $\mathbf{k}$, i.e. $C(\mathbf{k}, t) = C(k, t)$. Of course, the individual terms in the sum above can in general depend on the full vector $\mathbf{k}$, the isotropic function resulting only after averaging. To obtain an expression that explicitly depends only on the modulus k, one can average over the directions of $\mathbf{k}$ and then switch the

order of the averages. In 3-*d*, we have $C(k) = (1/4\pi) \int d\Omega\, C(k, t) = (1/4\pi) \int d\Omega\, C(\mathbf{k}, t)$, so that

$$
\begin{aligned}
C(k, t) &= \frac{1}{N} \left\langle \sum_{i,j}^{N} \delta\mathbf{v}_i(t_0) \cdot \delta\mathbf{v}_j(t_0 + t) \frac{1}{4\pi} \int d\Omega\, e^{-i\mathbf{k}\cdot[\mathbf{r}_i(t_0)-\mathbf{r}_j(t_0+t)]} \right\rangle_{t_0} \\
&= \frac{1}{N} \left\langle \sum_{i,j}^{N} \frac{\sin k r_{ij}(t_0, t)}{k r_{ij}(t_0, t)} \delta\mathbf{v}_i(t_0) \cdot \delta\mathbf{v}_j(t_0 + t) \right\rangle_{t_0}.
\end{aligned}
\tag{152}
$$

For $t = 0$ we recover the static correlation function in Fourier space, $C_0(k) \equiv C(k, t = 0)$. Generalizing (37), the relation between $C(\mathbf{k}, t)$ and $C(\mathbf{r}, t)$ is

$$
C(\mathbf{k}, t) = \rho_0 \int d\mathbf{r}\, g(\mathbf{r}, t) C(\mathbf{r}, t) e^{i\mathbf{k}\cdot\mathbf{r}},
\tag{153}
$$

where $g(\mathbf{r}, t) = (1/\rho_0 N) \sum_{ij} \delta(\mathbf{r} - \mathbf{r}_{ij}(t))$ is proportional[4] to van Hove's space time-dependent pair correlation function [13] (the generalization of the pair correlation function $g(r)$ introduced in Section 2.1.3). We note that both $C(\mathbf{r}, t)$ and $C(\mathbf{k}, t)$ can be defined starting from microscopic space–time dependent velocity fields, rather than the individual velocities − for a discussion of this derivation see Appendix A.

Let us consider the limits $k \to 0$ and $k \to \infty$. When $k \to 0$, both $e^{i\mathbf{k}\cdot[\mathbf{r}_i(t_0)-\mathbf{r}_j(t_0+t)]}$ and $\sin k r_{ij}/k r_{ij}$ tend to 1, so that due to the sum rule (17) we have

$$
\lim_{k \to 0} C(k, t) = \frac{1}{N} \sum_{i,j} \left\langle \delta\mathbf{v}_i(t_0) \cdot \delta\mathbf{v}_j(t_0 + t) \right\rangle_{t_0} = 0.
\tag{154}
$$

As in the static case, this result implies that $C(\mathbf{r}, t)$ must have a zero, for using (153) we have

$$
0 = C(k = 0, t) = \rho_0 \int d\mathbf{r}\, g(\mathbf{r}, t) C(\mathbf{r}, t),
\tag{155}
$$

and since $g(\mathbf{r}, t)$ is positive we conclude that $C(\mathbf{r}, t)$ must have a zero as a function of $\mathbf{r}$ for all times.

We can work out the opposite limit for $t = 0$. We note that the sum is dominated by pairs with distance such that $k r_{ij} < 1$, as larger distances are suppressed by the rapidly oscillating sine factor. For $k \to \infty$ all terms with $r_{ij} \neq 0$ are killed, and

$$
\lim_{k \to \infty} C(k, t = 0) = \frac{1}{N} \left\langle \sum_{i}^{N} \delta\mathbf{v}_i^2 \right\rangle_{t_0}.
\tag{156}
$$

When decreasing the momentum from $k = \infty$, the sum starts including neighbors separated by $r \sim 1/k$. If $k > 1/\xi$ (where ξ is the correlation length) these pairs will be correlated and thus $C_0(k)$ will increase. Further lowering k will include still more pairs, so we expect that $C_0(k)$ will continue increasing up to $k \sim 1/\xi$, where it should level off, and start decreasing again for $k \sim 1/L$ when the sum starts feeling the effects of the sum rule.

4.3. Space–time correlations in the ordered phase of the Vicsek and inertial spin models

In this section we compute, in an approximate case, the space–time correlations we have just defined for the two models (VM and ISM) we have discussed in Section 3. We wish to establish the main qualitative features peculiar to inertial dynamics that show up in the correlations. We start with the theoretical case because it is simpler to appreciate the differences between the inertial and overdamped dynamics when one can assume a system of infinite size and time signals of infinite duration. In actual experiments and simulations, finite-size effects (in time as well as in space) introduce a series of complications we postpone until Section 4.4.

The full analytic treatment of the VM or ISM is very difficult, but as before we will restrict ourselves to the case of the highly polarized regime, and for time scales smaller than the ones of network rearrangements. Then we can resort to two approximations that make the problem analytically tractable: the fixed network approximation, which decouples the connectivity matrix from the particles' positions (making it time-independent) and the spin-wave expansion, which allows linearization of the equations of motion. We have presented the linearized equations for this case in Section 3, which we recall here:

$$
\eta \partial_t \varphi(\mathbf{x}, t) - J n_c a^2 \nabla^2 \varphi(\mathbf{x}, t) = \zeta(\mathbf{x}, t), \qquad\qquad \text{VM}
\tag{157}
$$

$$
\chi \partial_t^2 \varphi(\mathbf{x}, t) + \eta \partial_t \varphi(\mathbf{x}, t) - J n_c a^2 \nabla^2 \varphi(\mathbf{x}, t) = \zeta(\mathbf{x}, t,) \qquad \text{ISM}
\tag{158}
$$

---

[4] The difference is just a factor $1/\rho_0$ we add to arrive at a formula as similar as possible to the static formula (37).

with the random force obeying $\langle \zeta(\mathbf{x}, t)\zeta(\mathbf{x}', t')\rangle = 2\eta T a^3 \delta(\mathbf{x} - \mathbf{x}')\delta(t - t')$. In this case we have a scalar field, so that the time correlation analogous to (148) is

$$C(\mathbf{x}, t) = \langle \varphi(\mathbf{x}, t)\varphi(0, 0)\rangle. \tag{159}$$

To compute (159) one solves Eqs. (157) and (158) by Green's function method (see Appendix D, the equations are of form (D.5)). Once obtained, the Green's function is used to write the full time evolution of the field under a given realization of the random force. Then one can write the product $\varphi(\mathbf{x}, t)\varphi(0, 0)$ and perform the average. The average is done over the realization of the random force $\zeta(\mathbf{x}, t)$, because, since we can assume that the time duration of the signal is infinite, the time average is equivalent to an average over the random force. Details are in Appendix F; here we just quote the results. The Green's function in Fourier space is

$$G(k, \omega) = \begin{cases} \left(-i\eta\omega + Jn_c a^2 k^2\right)^{-1}, & \text{VM} \\ \left(-\chi\omega^2 - i\eta\omega + Jn_c a^2 k^2\right)^{-1}. & \text{ISM} \end{cases} \tag{160}$$

From these, the Fourier transform of (159) is obtained as

$$C(\mathbf{k}, \omega) = 2\eta T a^3 G(\mathbf{k}, \omega)G(-\mathbf{k}, -\omega). \tag{161}$$

Finally the integral over ω to transform back into the time domain can be done Appendix F, yielding the analogous of (151) for the scalar case. With the definitions

$$\gamma \equiv \frac{\eta}{2\chi}, \quad c_2^2 = \frac{n_c a^2 J}{\chi}, \quad k_0^2 = \frac{\gamma^2}{c_2^2}, \quad \tau_0(k) \equiv \frac{\gamma^{-1} c_2^2 k^{-2}}{2} = \frac{n_c a^2 J}{\eta k^2}, \tag{162}$$

$$\omega_\pm \equiv -i\gamma \pm c_2\sqrt{k^2 - k_0^2} \equiv -i\gamma \pm \hat{\omega}(k), \tag{163}$$

the result is

$$C(\mathbf{k}, t) = C_0(k)e^{-t/\tau_0(k)}, \quad \text{VM} \tag{164}$$

$$C(\mathbf{k}, t) = C_0(k)e^{-\gamma t}\left[\cos\hat{\omega}(k)t + \frac{\gamma}{\hat{\omega}(k)}\sin\hat{\omega}(k)t\right], \quad \text{ISM.} \tag{165}$$

The static correlation $C_0(k)$ is the same for both models, since they share the interactions and only differ in the dynamic equation. It reads

$$C_0(k) = \frac{T\tau_0(k)}{\eta} = \frac{Tn_c a^2 J}{k^2}. \tag{166}$$

Three important observations can be made about these results. The first is that the presence of inertia drastically alters the shape of the correlation function. In the VM, the correlation decays as a pure exponential with $\tau \sim k^{-2}$ for all k. In contrast, the ISM correlation has damped *oscillations* at sufficiently high k. In fact, the frequency that enters the ISM correlation (165) is given by the dispersion relation (163), which we have already encountered in the discussion of second sound, Eq. (134). This is an explicit example that shows how the correlation functions can yield information on the response of the system, since the oscillations of the correlation, ruled by the same dispersion relations as the waves caused by the propagation of a perturbation induced by an external field, are present in the absence of such field, driven solely by the system noise (represented here as the random force ζ). The discussion of the dispersion relation in Section 3.3.5 applies here as well, and the inverse length scale k_0 marks the cut below which $\hat{\omega}$ is imaginary and the relaxation is exponential also in the ISM: for $k < k_0$, the correlation decays with a combination of two exponentials, with the dominant relaxation time for $k \to 0$ behaving as k^{-2}. On the other hand, when $k > k_0$, there are damped oscillations (with a k-independent damping constant γ), with a frequency that is almost linear in k for $k \gg k_0$.

The second observation is that the expansion to lowest order in time reads

$$\frac{C(\mathbf{k}, t)}{C_0(k)} \sim 1 - \frac{\eta k^2}{n_c a^2 J}t, \quad \text{VM,} \tag{167}$$

$$\frac{C(\mathbf{k}, t)}{C_0(k)} \sim 1 - \frac{c_2^2 k^2}{2}t^2, \quad \text{ISM,} \tag{168}$$

where the ISM expression is valid for all k (even below k_0). Thus the initial decay has different shape (linear for VM, quadratic for ISM), but also the dominant time scale at short times scales differently with k in the two models (as k^{-2} for VM, as k^{-1} for ISM).

Finally, the initial time derivative of the correlation is different: in the ISM the derivative at $t = 0$ vanishes, while it is finite (proportional to k^2) in the VM case. This is related to the previous comment (the initial derivatives can be directly read from the short-time expansion), but the vanishing of the derivative is a more general feature that can be seen to follow

quite from the pole structure of the Green's function: as we prove in Appendix G, the null initial derivative is caused by the existence of two or more poles in the Green's function, which cannot be the case in a diffusive (first-order in time) equation.

The last point deserves further comment. The derivative can be studied by considering the function

$$h(x) = -\frac{1}{x} \log \frac{C(k,t)}{C_0(k)}, \qquad x \equiv t/\tau_0(k), \tag{169}$$

in the interval $x \in [0,1]$, that is for times shorter than the characteristic relaxation time. If the correlation is a pure exponential, $h(x)$ will tend to one in the limit $x \to 0$, but if the correlation is flat in this limit it will approach zero:

$$\lim_{x \to 0} h(x) = \begin{cases} 1, & \text{VM}, \\ 0, & \text{ISM}. \end{cases} \tag{170}$$

This simple result is very useful to distinguish between the two types of behavior, especially when the correlation is obtained from experimental or simulation data, with discrete sampling in time that makes it very difficult to obtain clean numerical estimates of the derivative.

The result (170) may seem strange in view of the fact that a model with non-inertial dynamics (the VM in this case) can be viewed as the over-damped limit of model with inertial dynamics (here the ISM). As one takes this limit, for any finite values of η and χ the equation will be second order, and thus the correlation must be flat by the argument of Appendix G (as it can also be checked directly in (F.9)). So how does the correlation acquire a finite derivative? The answer is that the over-damped limit implies a rescaling of time as well, since otherwise a short period of inertial dynamics is always present. So what happens is that the flat region spans shorter and shorter times until it becomes much shorter than the relaxation time, and unobservable in the time scale required to observe the decay (in an experiment, below the experimental time resolution). This is discussed in detail for the case of a single harmonic oscillator in Appendix H.

The three qualitative differences between inertial and non-inertial dynamics recorded by $C(\mathbf{k},t)$ can be used to detect inertial effects in simulation data of the full VM and ISM models and in experimental field data. The detailed analytic results (164) and (165) are of course limited to the highly ordered phase where the spin-wave approximation is valid, but the qualitative observations we have made remain essentially valid, as will be seen below. However, to analyze the experimental and numerical data, and in absence of known analytic expressions, we need to formulate a different way to quantify these differences. This what we consider next.

4.4. Dynamic scaling of space–time correlations

When analyzing experimentally obtained time correlations to establish whether there are significant inertial effects in the dynamics, one would naively proceed to compute $C(\mathbf{k},t)$ from the trajectories using (152) and investigate the three qualitative indicators mentioned above (pure decay vs. oscillations, dominating time scale at short times behaving as k^{-1} vs k^{-2}, finite vs. vanishing initial time derivative). However, these features will not hold strictly as stated in experimental systems. This is in part due to the approximations involved in deriving them, but these should not be too dangerous if the system is highly ordered. More important are the effects of the system's finite size and of the finite available experimental times.

Although oscillations should still be present in a finite inertial system, trying to distinguish between them and a pure decay may be difficult if the experimental time is limited to one oscillation. Although this could in principle be cured by looking at higher values of k, in practice one is limited by the requirement that $1/L < k \ll 1/a$, so in the case of relatively small flocks the range of k is very limited. This also hampers efforts to study the behavior of the relaxation time, hoping to find $\tau \sim k^{-2}$ in the non-inertial case vs $\tau \sim$ const in the inertial case (read somehow from the envelope of oscillations). An additional difficulty is the fact that the finite size will change the behavior of τ at small k in both cases. Only the third indicator (the initial derivative) can be reasonably hoped to be observed as stated.

A way to obtain useful results on the face of these limitations has been proposed in Ref. [29], and is ultimately rooted in dynamical scaling [54]. In a nutshell, dynamical scaling consists in assuming that when the correlation length is large, the time correlation can be written (in the thermodynamic limit) as

$$C(k,t) = C_0(k) g\left(\frac{t}{\tau_0(k,\xi)}; k\xi\right), \tag{171}$$

where the characteristic time scale τ_0 is a homogeneous function of k,

$$\tau_0(k,\xi) = k^{-z} \Gamma(k\xi), \tag{172}$$

where z is an *a priori* unknown exponent called *dynamical critical exponent* [54].

The dynamic scaling hypothesis Eq. (171) thus assumes that the parameters controlling the order present in the system (such as temperature, noise, or interaction strength) affect the shape of $C(k,t)$ at fixed k only through the correlation length, ξ. In addition, changing the observation length scale k can change the shape of the relaxation (the function g above), but only through the adimensional quantity $k\xi$. Finally the time scale τ_0 is ruled by $k\xi$ but must also be homogeneous in k.

The general method to test dynamic scaling and determine z is to compute the correlation length, ξ, and then to fix the wave number k such that the product $k\xi$ is constant; in this way, thanks to (171) and (172), the quotient $C(k,t)/C_0(k)$

becomes purely a function of the scaling variable $k^z t$, so that correlation functions at different values of the control parameter (but with $k\xi$ kept constant) all collapse on the same master curve for an appropriate value of the dynamic critical exponent z. The shape of this master correlation only depends on the nature of the system (overdamped vs inertial) and on the value of the product $k\xi$ [54].

We have not yet defined the time scale $\tau_0(k, \xi)$. The scaling should not depend on the details of how τ_0 is defined, as long as it represents the characteristic time dominating for large ξ. To obtain a time scale from the simulated correlation functions quoted below we have followed [54], who define a characteristic k-dependent *frequency* $\omega_0(k) = 1/\tau_0(k)$ by the requirement that half of spectral contribution to the static correlation results from the interval $[-\omega_0, \omega_0]$. That is, the static correlation $C_0(k) = C(k, t = 0)$ contains dynamic contributions at all frequencies, $\int_{-\infty}^{\infty} C(k, \omega) \, d\omega/2\pi$; the characteristic frequency is defined by

$$\int_{-\omega_0}^{\omega_0} C(k, \omega) \frac{d\omega}{2\pi} = \frac{1}{2} C_0(k). \tag{173}$$

Expressing $C(k, \omega)$ in terms of $C(k, t)$, the time scale can be defined directly in the time domain as

$$\int_{-\infty}^{\infty} \frac{dt}{t} \frac{C(k, t)}{C_0(k)} \sin(t/\tau_0) = \frac{\pi}{2}. \tag{174}$$

The general scaling that we have just discussed has been applied with an important modification in [29] to the study of the ordered phase of a continuous system, i.e. of flocks. As we have mentioned, in the ordered phase of an infinite system with spontaneously broken continuous symmetry, the correlation length for fluctuations of the order parameter is infinite (Goldstone theorem [62]); hence, $\xi = \infty$ cannot be used in (171), because it does not reflect the changes of the control parameters. The quantity that enters in (171) in this case is Josephson's correlation length [63,54], which diverges at the critical point, but is otherwise finite across the ordered phase. The scaling relation should thus in principle be used near the critical region, where Josephson's length is large; this is the approach used in [64] to study continuous systems below but close to T_c. In [29], instead, Eq. (171) is applied in the highly ordered ($\Phi \sim 0.9$) region of *finite* systems of size L; in this case, we are far from the critical point, and indeed the Josephson's correlation length is too small to be used within dynamical scaling. However, we have hypothesized in [29] that one can still apply dynamic scaling by using the standard correlation length, ξ, because, owing to the scale-free nature of the ordered phase (Goldstone mechanism), one has $\xi \sim L$, which is finite. Hence, in [29] the dynamical scaling hypothesis (171) and (172) have been reformulated for a system with broken continuous symmetry in its deeply ordered phase as

$$C(k, t) = C_0(k) \, g\left(\frac{t}{\tau_0(k, L)}; kL\right), \tag{175}$$

and

$$\tau_0(k, L) = k^{-z} \Gamma(kL). \tag{176}$$

The dynamic scaling method employed in [29] consists then in determining the exponent z by attempting to scale the correlations *of systems of different sizes* holding $kL \sim 1$ fixed; if we do this, Eqs. (175) and (176) give

$$\frac{C(k, t)}{C_0(k)} = f(k^z t), \tag{177}$$

hence all correlation functions must collapse on the same master curve when plotted against the scaling variable $k^z t$, with $k = 1/L$. The method discussed above has been tested in simulations of the ISM and VM models at low temperature [29]. It is found that the presence of inertia indeed changes the dynamical exponent from $z \approx 2$ in the overdamped case (VM), to $z \approx 1$ in the underdamped case (ISM) (Fig. 18). The relaxation time is also found to scale as in (176), with the same exponents (Fig. 19). Hence, the dynamic scaling analysis of the deeply ordered phase is fully confirmed by numerical simulations in two different models. The different value of the exponent z is therefore in the ordered phase a very effective tool to distinguish systems where the inertia is relevant from systems where it is not. In general, though, namely in systems that are not deeply ordered, we must be careful.

The substitution of L for ξ in the dynamical scaling hypothesis can only be done in the deeply ordered phase of a system with broken continuous symmetry, where $\xi \sim L$. In the renormalization group jargon, in this way we are studying the properties of the $T = 0$ fixed point, and not those of the $T = T_c$ critical fixed point; to study this latter case, namely the critical dynamics of a system, one must employ the scaling hypothesis (171) either slightly above T_c with the standard correlation length, ξ, or slightly below T_c with Josephson's length. A similar caveat holds for the determination of the dynamical critical exponent z. The exponent z is in general highly non-trivial, and its value is not easily read from the dispersion relation, because the factors involved in the dispersion relation get renormalized near the $T = T_c$ critical point. For example, given the linear dispersion relation $\omega(k) = c \, k$, one would naively guess $z = 1$, but this is only true if c remains constant. However, near the critical point c is renormalized, and depends on the control parameters through the correlation length, so that $c \sim 1/\xi^\alpha$. The dispersion relation then becomes $\omega \sim \xi^{-\alpha} k = k^{1+\alpha} (k\xi)^{-\alpha}$, and it follows that $z = 1 + \alpha$ [54]. Hence, in general, we *cannot* read the dynamical critical exponent directly from the dispersion relation, and we must use the full dynamic scaling
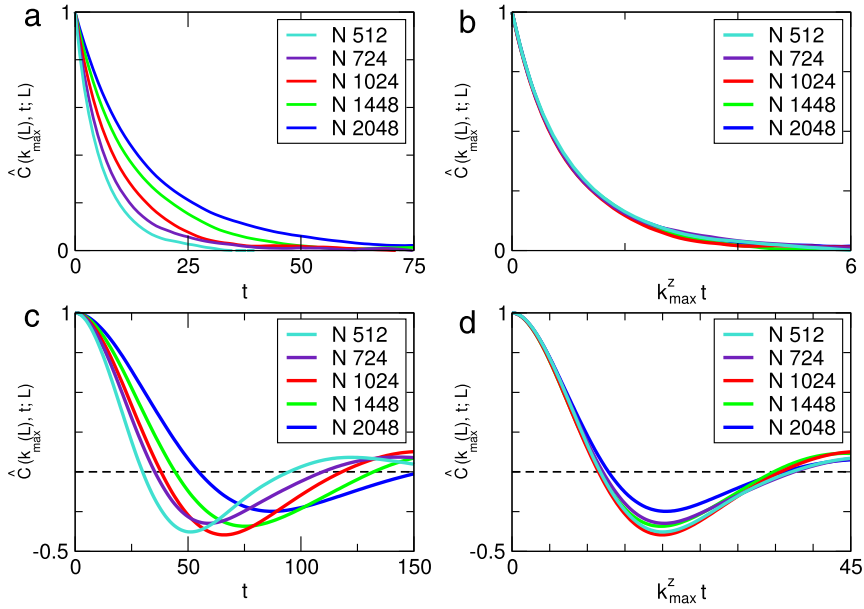
Fig. 18. Numerical simulations of the VM (panels a and b) and ISM (panels c and d) models in $d = 3$. Left panels: correlation function $\hat{C}(k, t) = C(k, t)/C_0(k)$ as a function of time with kL fixed. Right panels: scaling plots showing collapse of correlations at different sizes and wavenumbers. $z_{\text{Vicsek}} = 2.13$, $z_{\text{ISM}} = 1.15$.
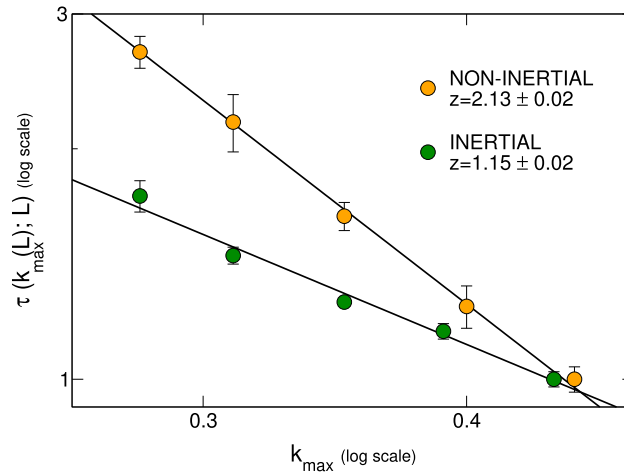*Source:* From Ref. [29].

Fig. 19. Time scales of the inertial (ISM) and non-inertial (VM) models.
Source: From Ref. [29].

hypothesis (171) to collapse the correlation functions to obtain z. However, the case of flocks analyzed here and considered in Ref. [29] is peculiar because in the deeply ordered phase (zero-temperature fixed point) we do not expect a renormalization of the prefactors of the dispersion relation; in addition here the approximate treatment of Section 4.3 is valid, hence we know the dispersion relation exactly. The conclusion is that one actually *can* read z from the dispersion relation (160) *only because we are in the deeply ordered phase*, where the coefficients of the dispersion relation do not renormalize. In this case the exponent z read from the dispersion relation must also work in the dynamic scaling hypothesis, giving a collapse of the correlation functions. This is why we have found $z \approx 1$ for inertial systems (corresponding to the observation of oscillations with a nearly linear dispersion relation) and $z \approx 2$ for overdamped systems.

In general, then, the dynamical critical exponent $z_{T=0}$ that we determine for flocks in their deeply ordered phase ($T = 0$ fixed point) can differ from the dynamical critical exponent $z_{T=T_c}$ that holds close to the ordering transition ($T = T_c$ fixed point). This last exponent may be relevant for disordered systems, as insect swarms [65], but not for flocks. A similar investigation near the ordering transition is clearly in order, although the value of $z_{T=T_c}$ is much harder to guess for the

reasons given above (renormalization of the coefficients in the dispersion relation). For this reason, it is also not clear whether $z_{T=T_c}$ will be different in the inertial and non-inertial cases. For systems living near the ordering transition the derivative of the time correlation function for finite k at $t = 0$ may prove to be a better fingerprint of the presence of inertia than the dynamical exponent.

4.5. The issue of space–time correlations in flocks: wing-flapping and experimental resolution

The scaling method we have discussed in the previous Section can be applied to experimental data-sets, as velocity time correlation functions are experimentally accessible from the 3-d trajectories. This has been done explicitly for midge swarms in [65]. But this is also the kind of data we have for starling flocks, hence the reader may expect us to do this next. Unfortunately, we will not. As we have seen, the spatio-temporal correlation functions provide important information both at long and at short times, where short means not unreasonably short, but of the same order or somewhat smaller than the relaxation time. Similarly, when analyzing correlation functions, space and time are obviously connected, hence medium–short times correspond to medium–short distances; therefore, while for the purely static correlation functions analyzed in the first Section we were simply interested in the long-distance behavior (correlation length and how it scales with the system's size L), when studying spatio-temporal correlations we also have to take care of smaller distances. Such short scales, both in time and in space, are problematic in flocks. Let us briefly explain why.

The first nontrivial issue is wing flapping (WF). Starlings move their wings at about 10hz, hence with a time resolution of 170 frames-per-seconds as in our experiments, wing flapping is clearly captured in the trajectories (see Fig. 1 of [36]). Such motion creates a zig-zag effect of the trajectories, with consequent large fluctuations of the velocity, that is superimposed to the main directional adjustment of each bird. Our observations show that WF is uncorrelated between birds, hence it acts as a noise that tends to destroy the mean natural correlation between the birds, making problematic the determination of the correlation function at short scales. What one needs to do is to filter out WF, as we have done in [36] for the determination of the mutual turning delays of the birds; however, in that case we only had to use the long-time behavior of the filtered trajectories, while for spatio-temporal correlation we also would use the short-time behavior. Hence, filtering becomes a sensitive issue, to be performed with care.

Moreover, short space scales are more subject to noise because of segmentation issues (see [66] and in particular [67]), because of the inevitable noise in the image detail when we focus on details below a few pixels. However, a smooth determination of the trajectories also for short scales is quite important, if we want to distinguish *bona fide* WF from random noise. Our current tracking methods can be improved regarding short-scales smoothness, hence we hopefully will overcome these issues in the next generation of data analysis. For now, we are afraid the reader must be content with the numerical simulations of the previous section.

5. Final remarks and outlook

5.1. The need to keep it simple

One could say that computing correlations, which is what we have done all along this work, is quite a basic thing to do and that it blatantly lacks the complexity and diversity of many other methods. Biology is a rather vast field, and collective behavior is not its easiest alley, hence the reader could have expected our tools of analysis to be as diverse and complex as such a multiform field deserves. More specifically, one may object that by restricting ourselves to correlation in extracting information about the equations ruling the systems, we are limiting too much our possibilities. This may be true, but there are nevertheless a couple of considerations to make.

First, we hope this work has convinced the reader that, basic as it may be, the conceptual framework of correlation functions is still capable of developing a rich score of mathematical and physical consequences, whose complexity is sometimes too much to digest, rather than too little. Yes, correlation is just one concept, and yet it has deep and solid roots, so that its branches are very diverse and far-reaching. We have seen that correlation is a two-way path that leads down in the microscopic direction to the mechanisms of inter-individual interaction and the dynamical equations ruling the system, as well up in the macroscopic direction to group-level information transfer and collective response. Correlation is therefore a key bridge between scales, which is all the more useful since collective behavior is really a mystery about how the microscopic scale can change so much the macroscopic one. This up-scaling, by the way, is the very essence of statistical physics, hence it is unsurprising that the physics of collective behavior has invaded so indiscreetly the foreign field of biology.

Second, when we leave the well-defined arena of correlations, we certainly gain in the number of tools of analysis and in terms of things to observe experimentally and to calculate mathematically, but we do so at such a level that some caution is in order. As we have seen, despite the obvious crucial differences between statistical mechanics and biology, when we use correlations we still enjoy the conceptual overtones of a well-established theoretical framework, a sort of intellectual map that is effective at least at signaling our worst blunders. For example, we do not need to know much about a system to expect that the correlation must decay to zero for sufficiently large distance and/or time. Similarly, we have a lot of venerable theory telling us what should be the connection between correlation length, susceptibility and fluctuations. It is clearly not the case that all this theory (which is statistical mechanics) must hold no matter what; but at least, when we do find something anomalous, we know anomalous respect to *what;* we have a frame of reference. When we use other, more sophisticated

and exotic means, we usually have no background reference theory to gauge our results, no zeroth order result, no trivial expectation. This is absolute freedom and in such a context every result is likely to be new and exciting. But, of course, it may be harder to assess its relevance and some times even its *bona fide* validity.

Collective behavior in biological systems *is* an utterly complex field, in which one studies systems as diverse as bird flocks, fish schools, insect swarms and cell colonies. Our opinion is that, in face of such complexity of the phenomenon and novelty of the field, it may be wise to limit the complexity and novelty of the theoretical framework of analysis, lest the data analysis explodes in our face. This is why, within the approach to collective behavior developed in this work, the tendency to 'keep it simple' is so strong.

5.2. Beyond flocks

It goes without saying that there is a large variety of biological systems displaying collective behavior beyond flocks. What about them? the reader may ask. Will I need to read a different review for each one of those? That would be exhausting, if not positively depressing. Our own group has been gathering data not only on bird flocks, but also, more recently, on midge swarms, which are very different from flocks: swarms are not polarized, hence do not enjoy the simplifications of the spin wave approximation we have heavily used in this work; moreover, the lack of polarization means that rotational symmetry is not spontaneously broken, so one does not expect long-range correlation of the directions of motion on the basis of a Goldstone argument (see Section 2.2). This lack of order in swarms may push some observer as far as to guess that no collective behavior is present, and that the swarm is simply the result of independent insects trying to keep their positions near some environmental landmark. In other words, in absence of group-level order, it seems hard to assess or even define the existence of collective behavior.

However, despite these differences, we have now quite solid evidence that collective behavior *is* present in swarms as much as in flocks, since the collective correlation functions, both static and dynamic, are very strong [11,12]. Our own understanding of natural swarms is not as mature as that of bird flocks (which is why we limited the present work to flocks), and yet we believe we can draw a first general conclusion: the behavior of both flocks and swarms, in terms of correlation functions, is quite similar, suggesting that strong (connected) correlation, rather than strong order, is the true landmark of collective behavior. In both cases the inter-individual interaction is based on an imitation mechanism, not unlike that described by the Vicsek model; different level of noise and different values of the control parameters determine the fact that flocks live in the ordered phase, while swarms in the disordered one, but in both cases correlations are large.

Other groups have calculated correlations in different biological systems, most notably in neural assemblies [68], bacterial clusters [69,70], cell colonies [71], and proteins [72]. In all this wildly diverse cases correlations of the fluctuations turned out to be strong, as in flocks and swarms. This fact is so remarkable to suggest that it cannot be simply chance and that biological systems are poised at criticality for some fundamental reason. This is a vast and highly controversial topic of discussion that we cannot examine here; the interested reader can find a very good account in [73]. Irrespective of the issue of criticality, we can remark that the method of analysis based on correlation functions is simple and general enough to be adapted to very different systems (though with care), hence providing a common conceptual framework to discuss collective behavior at the most general level.

5.3. Correlation: What is it good for?

The question may seem a provocation, after such a long ode to correlation. But our point here is: what is the *biological* purpose, if any, of strong correlations in a group? We have followed in this work a strict mechanistic philosophy, typical of the physicist's approach: do not ask *why*, stick to *how*. However, at least in the outlook section we can briefly indulge and ponder on the biological function (again: if any) of the strong correlations that seem to be so common in biological systems displaying collective behavior. First, one must be aware of the fact that in certain cases strong correlations may be simply the result of math; as we have seen, a flock is a system that breaks a continuous symmetry (rotation), hence on the basis of Goldstone's theorem one expects scale-free correlations of the directions of motion. Correlation here appears not because of a biological bonus, but due to an accident of global ordering; hence, it has no 'function'. On the other hand, we have also seen that long range correlations also hold for the speed, and this is not a consequence of Goldstone's theorem or of any other piece of known math. Similarly, in midge swarms strong correlations of the directions cannot be due to a symmetry breaking mechanism, because no symmetry is broken in swarms (they are disordered 'paramagnets' as one would call them in a magnetic context), hence the origin of long range spatial correlation is not merely mathematical. Similar cases occur in other systems. Hence, in all these cases our question is valid: what is the biological purpose (if any) of strong correlations within a group?

In the case of flocks, the answer seems rather simple: as we have seen all along this work, a scale-free correlation function is a necessary condition (not sufficient, though) to transfer information across the group without the need of encoding/decoding, independently of the group's size. We have seen how this is crucial when a flock changes collectively direction of motion, but it is equally crucial when the group undergoes a sudden change of speed: in absence of scale-free correlation the information to change some collective property does not reach the whole group, putting its cohesion at risk. Hence, strong correlation seems functional to having an effective information transfer, which is probably a good thing at the biological level.

However, it is not easy to generalize the link between correlation and information transfer, because in many systems it is not clear what kind of information would be transferred. Consider again swarms: correlations of the directions of motion are large, but why would the swarm propagate a change of direction when there is no net direction of motion of the group? In other systems the situation is even trickier as there is no clear spatial structure, hence propagation of information in its classic sense is hard to define. For example, in neural assemblies, where each neuron is connected to each other, one cannot even define a correlation length, hence connecting it to information transfer becomes problematic.

In all cases when correlation is strong, even though it is not clearly connected to spatial information transfer, it is the physicist's second nature to guess that correlation is relevant because a strongly correlated system responds strongly to external perturbations. In other words, as physicists, when we discuss correlations, susceptibility and fluctuations we always have in the back of our mind the fluctuation–dissipation theorem. This seems to make sense: a biological system is strongly correlated so that it can respond, or react, more effectively to environmental perturbations (a big deal in biology). The problem is that, intuitive as this scenario may be, it should be proved more convincingly than by invoking a statistical mechanics principle. In fact, one may very well argue in the opposite direction: if a highly correlated systems is (according to the physicist) also a highly susceptible one, it is far from ideal with respect to its biological stability. The capacity of a system to maintain such stability is called homeostasis, which is also quite a big deal in biology.

As usual, the question should be settled by experiments, and this is our true outlook. We believe that the next generation of experiments in collective behavior should focus on the perturbation–response relationship, with the aim to validate or reject the hypothesis that more correlated systems respond 'better' to perturbations (and where a careful, experimentally checkable, definition of 'better' should be provided). A critical issue in this context is how to choose the pair of conjugated variables upon which building this analysis. In physics, every observable A that we measure has its own conjugated variable b that we can manipulate to obtain a change of A (classic pairs are volume and pressure, magnetization and magnetic field, energy and temperature, and so on). In a perturbation–response experiment we know that we have to measure the extent of change of A, given some (possibly small) change of b. This ratio should be connected, both statically and dynamically, to the correlation function of A. Doing this within a biological group is far from trivial, simply because in general we do not have a Hamiltonian formalism dictating us the conjugated pairs. Hence, to compare the correlation of the orientations in a flock with a perturbation experiment, what variable should we be controlling? What is the field conjugated to the velocity in a natural way? Similarly in swarms and other systems: we observe the large correlation, but we do not know *a priori* what is the right perturbation–response pair to compare with. This is problematic, but also quite exciting, as it suggests that there may be more in store to discover than simply a connection between correlation and response, as perhaps we will be able to define for collective behavior in biological systems a full set of 'canonically' conjugated variables as we do in statistical mechanics. The next decade or so will tell whether or not this is just wishful thinking.

Acknowledgments

Appendix A. Two possible definitions of the velocity field

The definition of correlation function given in Eq. (6) describes an average of pair correlation performed at fixed mutual distance. Looking at properties in terms of their dependence on distance rather than on individual identities is necessary when dealing with systems that evolve off-lattice, and this is what is usually done in the study of liquids. From the point of view of the theory of fluids it is thus natural to ask how our correlation function is related to the correlation of the velocity field.

There are at least two different ways to define the velocity field. Consider a system where particles or organisms move in space, with V indicating the system's volume and N the global number of individuals. To define a field we can average $\mathbf{v}_i$ over all particles within a small volume δw around the position $\mathbf{r}$:

$$\mathbf{v}(\mathbf{r}) = \frac{1}{\delta n} \sum_{i \in \delta w} \mathbf{v}_i \xrightarrow[\delta w \to 0]{} \mathbf{v}(\mathbf{r}) = \frac{\sum_i \mathbf{v}_i \delta(\mathbf{r}_i - \mathbf{r})}{\sum_k \delta(\mathbf{r}_k - \mathbf{r})}, \tag{A.1}$$

where δn is the number of particles in the volume δw. One could also choose to average over the volume δw:

$$\mathbf{v}^{(S)}(\mathbf{r}) = \frac{1}{\rho_0} \frac{1}{\delta w} \sum_{i \in \delta w} \mathbf{v}_i \xrightarrow[\delta w \to 0]{} \mathbf{v}^{(S)}(\mathbf{r}) = \frac{1}{\rho_0} \sum_i \mathbf{v}_i \delta(\mathbf{r}_i - \mathbf{r}), \tag{A.2}$$

where we divided by the average density $\rho_0 = N/V$ to obtain a field with the dimensions of a velocity.

The definition $\mathbf{v}^{(S)}(\mathbf{r})$ is simpler, and it is standard when constructing local fields from single-particle observables [13]. However it has the disadvantage that it entangles density and velocity fluctuations: in the case where all the $\mathbf{v}_i$ are equal,

$\mathbf{v}^{(S)}(\mathbf{r})$ will be higher in regions of higher density, while $\mathbf{v}(\mathbf{r})$ would have the same value throughout. The local velocity of a tracer particle is given by $\mathbf{v}(\mathbf{r})$ and not by $\mathbf{v}^{(S)}(\mathbf{r})$. Although using $\mathbf{v}(\mathbf{r})$ results in slightly more complex expressions for the correlation functions, as we show below, it is the definition we will adopt in the rest of the discussions outside this subsection, because the ability to disentangle (at least partially) density and velocity fluctuations is crucial to interpret the correlation functions.[5]

The two fields lead to two different measures in Euclidean space. Since the sum of the velocities is an important quantity (it is proportional to the order parameter), we want to define an integration measure $\mu(\mathbf{r})$ in such a way that the integral over space of the continuous field is equal to the sum over all sites of the discrete variables:

$$\int d\mu(\mathbf{r})\mathbf{v}(\mathbf{r}) \equiv \sum_i \mathbf{v}_i. \tag{A.3}$$

This requirement, combined with Eqs. (A.1) and (A.2) immediately gives

$$d\mu(\mathbf{r}) = \rho(\mathbf{r})\,d\mathbf{r}, \tag{A.4}$$

$$d\mu^{(S)}(\mathbf{r}) = \rho_0\,d\mathbf{r}, \tag{A.5}$$

where

$$\rho(\mathbf{r}) = \sum_i \delta(\mathbf{r} - \mathbf{r}_i) \tag{A.6}$$

is the local number density [13]. Space averages over the fields thus require in both cases a normalization factor $1/N$:

$$\mathbf{V} = \frac{1}{N}\int d\mu(\mathbf{r})\,\mathbf{v}(r) = \frac{1}{N}\sum_i \mathbf{v}_i = \frac{1}{N}\int d\mu^{(S)}(\mathbf{r})\,\mathbf{v}^{(S)}(r). \tag{A.7}$$

Having defined the local fields and the corresponding integration measures, we can proceed to define space correlation functions. We introduce the fluctuation of the fields,

$$\delta\mathbf{v}(\mathbf{r}) = \mathbf{v}(\mathbf{r}) - \mathbf{V}, \tag{A.8}$$

$$\delta\mathbf{v}^{(S)}(\mathbf{r}) = \mathbf{v}^{(S)}(\mathbf{r}) - \frac{\rho(\mathbf{r})}{\rho_0}\mathbf{V}, \tag{A.9}$$

where, in Eq. (A.9), we subtract $(\rho(\mathbf{r})/\rho_0)\mathbf{V}$ to measure fluctuations due to velocity (rather than density) deviations. Then we consider fluctuations of the field at two points $\mathbf{r}_0$ and $\mathbf{r}_1$ in space and look at the mutual correlation between such field fluctuations. If we assume translational invariance then the correlation only depends on the distance vector $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_0$ and we can therefore perform a space average over $\mathbf{r}_0$, i.e.

$$C(\mathbf{r}) = \frac{1}{N}\int d\mu(\mathbf{r}_0)\,\delta\mathbf{v}(\mathbf{r}_0) \cdot \delta\mathbf{v}(\mathbf{r}_0 + \mathbf{r}), \tag{A.10}$$

$$C^{(S)}(\mathbf{r}) = \frac{1}{N}\int d\mu^{(S)}(\mathbf{r}_0)\,\delta\mathbf{v}^{(S)}(\mathbf{r}_0) \cdot \delta\mathbf{v}^{(S)}(\mathbf{r_0} + \mathbf{r}). \tag{A.11}$$

These two correlations lead to slightly different expressions. Using the definitions of the fields and the corresponding measures, we get

$$C(\mathbf{r}) = \frac{\sum_{ij} \delta\mathbf{v}_i \cdot \delta\mathbf{v}_j \delta(\mathbf{r} - \mathbf{r}_{ij})}{\sum_{kl} \delta(\mathbf{r} - \mathbf{r}_{kl})}, \tag{A.12}$$

$$C^{(S)}(\mathbf{r}) = \frac{1}{N\rho_0}\sum_{ij} \delta\mathbf{v}_i \cdot \delta\mathbf{v}_j \delta(\mathbf{r} - \mathbf{r}_{ij}) \tag{A.13}$$

where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$. If the system is isotropic the correlation only depends on $r = |\mathbf{r}|$ and the correlations becomes

$$C(r) = \frac{\sum_{ij} \delta\mathbf{v}_i \cdot \delta\mathbf{v}_j \delta(r - r_{ij})}{\sum_{kl} \delta(r - r_{kl})}, \tag{A.14}$$

$$C^{(S)}(r) = \frac{1}{N\rho_0}\frac{1}{4\pi r^2}\sum_{ij} \delta\mathbf{v}_i \cdot \delta\mathbf{v}_j \delta(r - r_{ij}). \tag{A.15}$$

Thus the correlation that we introduced in Eq. (6) is equivalent to the correlation function obtained using the field (A.1). As we already noticed, this definition separates the contribution of the particle velocities from the trivial contributions of

---

[5] Coupling between density and velocity is a characteristic of active matter systems. In some systems like the metric Vicsek model, a small region of higher density has a stronger effective interaction that leads to higher local ordering. In this sense, both definitions of the field will be coupled to density fluctuations, but $\mathbf{v}^{(S)}(\mathbf{r})$ picks up also trivial contributions such as random density fluctuations in a system of non-interacting particles with uniform velocity.

particle density. We therefore expect this correlation function to be less influenced by the presence of density fluctuations. This can also be seen by noting that

$$C^{(S)}(r) = g(r)C(r), \tag{A.16}$$

where $g(r) = (1/N\rho_0)\sum_{ij}\delta(r - r_{ij})/(4\pi r^2)$ is the radial distribution function[6] measuring the average number of particle pairs at mutual distance r. In a completely homogeneous system $g(r) = 1$ and the two definitions are equivalent. However in general $g(r) \neq 1$ (i.e. in the presence of non-trivial density correlations) and the two definitions differ.

We also notice that the correlation (A.15) is reasonably defined only in the bulk and might be subject to spurious boundary effects. Indeed, it basically estimates the number of particles at distance r from a given particle as $4\pi \rho_0 r^2 dr$, so that – even disregarding density fluctuations – this estimate is wrong if r is of the order of the linear size of the system. On the contrary, definition (A.14), by counting explicitly the number of pairs at distance r, is less sensitive to this boundary bias.

Summarizing, we expect the two correlations defined in this section to be rather similar on scales where the system does not exhibit strong density correlations and far from the boundaries. Since our main purpose is to capture the amount of directional correlations in real natural groups – where the system is necessarily finite and boundary effects can be important – and we want to decouple as much as possible from the density fluctuations, we adopt the definition of correlation function given in Eq. (6) and Eq. (A.14).

One interesting feature of defining the correlations in terms of velocity fields is that we can very naturally derive the correlation function in Fourier space. Let us focus now on the $\mathbf{v}(\mathbf{r})$ fields, the most relevant for active systems. The Fourier transform of the field is

$$\mathbf{v}(\mathbf{k}) = \int d\mu(\mathbf{r})\,e^{-i\mathbf{k}\cdot\mathbf{r}}\mathbf{v}(\mathbf{r}) = \sum_i \mathbf{v}_i e^{-i\mathbf{k}\cdot\mathbf{r}_i}. \tag{A.17}$$

We then define the Fourier-space correlation by

$$C(\mathbf{k}) = \frac{1}{N}\delta\mathbf{v}(\mathbf{k})\cdot\delta\mathbf{v}(-\mathbf{k}) = \frac{1}{N}\sum_{ij}\delta\mathbf{v}_i\cdot\delta\mathbf{v}_j e^{i\mathbf{k}\cdot(\mathbf{r}_i-\mathbf{r}_j)}, \tag{A.18}$$

which in the isotropic case is written, after averaging over the polar angles of $\mathbf{k}$,

$$C(k) = \frac{1}{N}\sum_{i,j}^{N}\int_{-1}^{+1}d(\cos\theta)e^{ikr_{ij}\cos(\theta)}\,\delta\mathbf{v}_i\cdot\delta\mathbf{v}_j$$

$$= \frac{1}{N}\sum_{i,j}^{N}\frac{\sin kr_{ij}}{kr_{ij}}\delta\mathbf{v}_i\cdot\delta\mathbf{v}_j. \tag{A.19}$$

The correlations (A.18) and (A.19) are equivalent to (35) and (36) of the main text, but were here derived starting from the velocity fields. As noticed in the main text, the relationship between $C(k)$ and $C(r)$ is

$$C(k) = \rho_0\int d\mathbf{r}\,g(r)\,e^{-i\mathbf{k}\cdot\mathbf{r}}C(r). \tag{A.20}$$

Within the present derivation, it is now clear that the factor $g(r)$ has its origin in the integration measure $d\mu(\mathbf{r})$ in the definition of the Fourier transform (A.17).

Finally, these definitions can be easily extended to the space–time correlation functions. In this case, we are interested in comparing fluctuations at different positions and times. We therefore generalize (A.10) and (A.18) to

$$C(\mathbf{r}, t) = \left\langle \frac{1}{N}\int d\mu(\mathbf{r}_0)\,\delta\mathbf{v}(\mathbf{r}_0, t_0)\cdot\delta\mathbf{v}(\mathbf{r}_0 + \mathbf{r}, t_0 + t) \right\rangle_{t_0}$$

$$= \left\langle \sum_{i,j}^{N} \frac{\delta\mathbf{v}_i(t_0)\cdot\delta\mathbf{v}_i(t_0 + t)\delta\left[\mathbf{r} + \mathbf{r}_i(t_0) - \mathbf{r}_j(t_0 + t)\right]}{\sum_{k,l}\delta\left[\mathbf{r} + \mathbf{r}_k(t_0) - \mathbf{r}_l(t_0 + t)\right]} \right\rangle_{t_0}, \tag{A.21}$$

and

$$C(\mathbf{k}, t) = \frac{1}{N}\langle\delta\mathbf{v}(\mathbf{k}, t_0)\cdot\delta\mathbf{v}(-\mathbf{k}, t_0 + t)\rangle_{t_0}$$

$$= \frac{1}{N}\left\langle \sum_{i,j}^{N}\delta\mathbf{v}_i(t_0)\cdot\delta\mathbf{v}_j(t_0 + t)e^{-i\mathbf{k}\cdot[\mathbf{r}_i(t_0)-\mathbf{r}_j(t_0+t)]} \right\rangle_{t_0}, \tag{A.22}$$

---

[6] More precisely, in the literature the radial distribution function $g^R(r)$ is often defined as the contribution to the two point correlation function at $r \neq 0$, i.e. $g(r) = g^R(r) + \delta(r)/(4\pi\rho_0 r^2)$, see [13].

where all spatial coordinates are expressed in the flock's reference frame. Again, we obtain the same expressions (148) and (151) of the main text, but starting from the velocity fields and their Fourier transforms.

Appendix B. The spin-wave expansion

When considering systems in their polarized phase, such as bird flocks or fish schools, which can have average polarization as high as $\Phi \sim 0.9$ [10], there is a well-defined mean velocity and individual orientations deviate weakly from this mean. When this happens it is possible to simplify the mathematical description of the system by expanding all non-linear quantities at the leading order in these weak fluctuations. This is called *spin-wave expansion* and we will describe it next.

We can assume that the mean velocity points along the unit vector $\mathbf{n}_x = (1, 0, 0)$. Each velocity $\mathbf{v}_i$ can be decomposed into a longitudinal component, let us call it v_i^x, along the direction of motion $\mathbf{n}_x$ and a transverse component, which is a $(d-1)$-dimensional vector $\boldsymbol{\pi}_i$ perpendicular to $\mathbf{n}_x$,

$$\mathbf{v}_i = v_i^x \mathbf{n}_x + \boldsymbol{\pi}_i. \tag{B.1}$$

The transverse components $\boldsymbol{\pi}_i$ have the physical dimension of a velocity and by construction they satisfy

$$\sum_i \boldsymbol{\pi}_i = 0. \tag{B.2}$$

The spin-wave expansion is most useful when considering cases where the individual speed is fixed, or can be approximated as fixed, $|\mathbf{v}_i| = v_0$. Then the longitudinal component can be written as a function of the transverse one,

$$v_i^x = \sqrt{v_0^2 - \pi_i^2}, \tag{B.3}$$

and when the polarization is large all velocities will be mainly along the mean direction of motion, implying $\pi_i^2 \ll v_0^2$. This is the spin-wave approximation, which yields,

$$v_i^x \approx v_0 \left(1 - \frac{1}{2}\pi_i^2/v_0^2\right), \tag{B.4}$$

and

$$\mathbf{v}_i \approx \mathbf{n}_x v_0 \left(1 - \frac{1}{2}\pi_i^2/v_0^2\right) + \boldsymbol{\pi}_i. \tag{B.5}$$

It is sometimes convenient to write the transverse components $\boldsymbol{\pi}_i$, in terms of dimensionless angles expressing the departure of each $\mathbf{v}_i$ from the mean direction of motion,

$$\pi_i^y = v_0 \sin \varphi_i^z \sim v_0 \, \varphi_i^z, \tag{B.6}$$
$$\pi_i^z = v_0 \sin \varphi_i^y \sim v_0 \, \varphi_i^y. \tag{B.7}$$

To understand these relations, recall that to create a y component of the velocity one needs to rotate $\mathbf{v}_i$ around the z axis, and vice-versa. These transverse angles φ_i^z and φ_i^y are the key degrees of freedom in a polarized system and they are called *phases*. They simply represent the (small) angular deviations of each individual $\mathbf{v}_i$ with respect to the mean velocity of the group.

From relations (B.5)–(B.7) we can work out an expression of the scalar polarization in terms of the phases,

$$\Phi = \left| \frac{1}{N} \sum_i \frac{\mathbf{v}_i}{v_0} \right| = 1 - \frac{(d-1)}{2N} \sum_i \varphi_i^2, \tag{B.8}$$

from which we see that the limit of large polarization, $\Phi \sim 1$, is equivalent to the limit of small phases, $\varphi_i^2 \ll 1$. In $d = 3$ we get

$$\Phi = 1 - \frac{1}{N} \sum_i \varphi_i^2. \tag{B.9}$$

Finally, in this approximation we can express the velocity fluctuations (4) using (B.6) and (B.7); we obtain,

$$\delta\mathbf{v}_i = v_0 \left(0, \ \varphi_i^z, \ \varphi_i^y\right), \tag{B.10}$$

up to linear order in the phases. This equation embodies the fact that in a polarized system the fluctuations are strongly dominated by their transverse components.

Appendix C. Linear response

In thermodynamics one defines *linear susceptibilities,* quantities that measure, to a first approximation, how some thermodynamic state variable changes under a variation of some control variable. For example, the magnetization m will change under an applied magnetic field h, and one defines the linear magnetic susceptibility as $\chi_m = \partial m/\partial h|_{h=0}$. Linear response theory generalizes this idea to space- and time-dependent variations of the quantity of interest.

Consider a field $\phi(\mathbf{x}, t)$ that under certain circumstances assumes a stationary value $\phi_0(\mathbf{x})$. Let us now change these circumstances in a way that can be modeled by a scalar field $h(\mathbf{x}, t)$ (which we can assume different from zero only for $t \geq 0$ to fix a time origin). The *impulse response function,* or simply response function, is defined as

$$R(\mathbf{x}, t; \mathbf{x}', t') = \left.\frac{\delta \phi(\mathbf{x}, t)}{\delta h(\mathbf{x}', t')}\right|_{h=0}, \tag{C.1}$$

where $\delta(\ldots)/\delta(\ldots)$ denotes the functional derivative

$$\frac{\delta \phi(\mathbf{x}, t)}{\delta h(\mathbf{x}', t')} \equiv \lim_{\epsilon \to 0} \frac{\phi[h(\mathbf{x}, t) + \epsilon \delta(\mathbf{x} - \mathbf{x}')\delta(t - t')] - \phi[h(\mathbf{x}, t)]}{\epsilon}. \tag{C.2}$$

From definition (C.1) we can write

$$\phi(\mathbf{x}, t) - \phi_0(\mathbf{x}) = \int d\mathbf{x}'\, dt'\, R(\mathbf{x}, t; \mathbf{x}', t')h(\mathbf{x}', t') + \cdots, \tag{C.3}$$

where the ellipsis stands for terms of second or higher order in $h(\mathbf{x}, t)$.

Since we have assumed that $\phi(\mathbf{x}, t)$ is stationary when $h(\mathbf{x}, t)$ is zero, it is mostly safe to assume that the time variations of $\phi(\mathbf{x}, t)$ after h is turned on are due to the effect of h. It follows that if h is turned on at a different time, the variations of ϕ will be identical but shifted in time by the same amount. This is equivalent to assuming that the response function is a function only of the time difference,

$$R(\mathbf{x}, t; \mathbf{x}', t') = R(\mathbf{x}, \mathbf{x}', t - t'). \tag{C.4}$$

In addition, we will assume *causality,* i.e. that the applied field does not affect the value of ϕ at times before being applied. This corresponds mathematically to the condition $R(\mathbf{x}, \mathbf{x}', t < 0) = 0$, so that the effective upper limit of the time integral in (C.3) is t.

If the perturbing field is a step function, $h(\mathbf{x}, t) = \hat{h}(\mathbf{x})\Theta(t)$, then the linear response is

$$\phi(\mathbf{x}, t) - \phi_0(\mathbf{x}) = \int d\mathbf{x}' \int_0^t dt'\, R(\mathbf{x}, \mathbf{x}', t - t')\hat{h}(\mathbf{x}'), \tag{C.5}$$

In this case the evolution of ϕ is described by the *dynamic susceptibility* or *integrated response,*

$$\chi(\mathbf{x}, \mathbf{x}', t) = \left.\frac{\delta \phi(\mathbf{x}, t)}{\delta \hat{h}(\mathbf{x}')}\right|_{\hat{h}=0} = \int_0^t R(\mathbf{x}, \mathbf{x}', t')\, dt'. \tag{C.6}$$

The static (thermodynamic) susceptibility is then the limit of the integrated response,

$$\chi(\mathbf{x}, \mathbf{x}') = \lim_{t \to \infty} \chi(\mathbf{x}, \mathbf{x}', t). \tag{C.7}$$

All of the above definitions are quite general, and their usefulness depends only on whether the applied field in the situation under study is strong enough to produce significant nonlinear effects. The next and last result, known as the static fluctuation–dissipation theorem holds strictly for equilibrium systems, i.e. when $\phi(\mathbf{x}, t)$ is a quantity averaged over microscopic degrees of freedom weighted with Boltzmann's factor. If $h(\mathbf{x}, t)$ is a field that couples linearly to some observable $\phi(\mathbf{x}, t)$, it is easy to show [15] that when equilibrium has been attained and ϕ is stationary,

$$\chi(\mathbf{x}, \mathbf{x}') = \beta \langle \phi(\mathbf{x})\phi(\mathbf{x}') \rangle - \langle \phi(\mathbf{x}) \rangle \langle \phi(\mathbf{x}') \rangle \equiv \beta C(\mathbf{x}, \mathbf{x}'), \tag{C.8}$$

where C is the *connected* correlation function.

Appendix D. Green's function method

Let $D_{\mathbf{x}, t}$ be a differential operator involving partial derivatives of time and position. The (partial) differential equation

$$D_{\mathbf{x}, t}\phi(\mathbf{x}, t) = f(\mathbf{x}, t), \tag{D.1}$$

considered in a region $\mathbf{x} \in \mathcal{R} \subset \mathbb{R}^d, t \in [t', \infty)$ may be expected to have a unique solution only after specifying in addition certain initial and boundary conditions [74,75]. The Green's function, or fundamental solution, $G(\mathbf{x}, t; \mathbf{x}', t')$ associated with (D.1) is the solution to

$$D_{\mathbf{x}, t}G(\mathbf{x}, t; \mathbf{x}', t') = \delta(\mathbf{x} - \mathbf{x}')\delta(t - t'), \qquad \lim_{t \to t'^+} G(\mathbf{x}, t; \mathbf{x}', t') = 1, \tag{D.2}$$

and subject to homogeneous boundary conditions of the same kind as (D.1). The Green's function is identical to the linear response defined in Appendix C: if $\phi_f(\mathbf{x}, t)$ is the solution for a given f, then applying $D_{\mathbf{x},t}$ to the linear response $R(\mathbf{x}, t; \mathbf{x}', t') = \delta\phi(\mathbf{x}, t)/\delta f(\mathbf{x}', t')$,

$$D_{\mathbf{x},t} R(\mathbf{x}, t; \mathbf{x}', t') = \frac{\delta D_{\mathbf{x},t}\phi(\mathbf{x}, t)}{\delta f(\mathbf{x}', t')} = \frac{\delta f(\mathbf{x}, t)}{\delta f(\mathbf{x}', t')} = \delta(\mathbf{x} - \mathbf{x}')\delta(t - t'), \tag{D.3}$$

i.e. R is a solution of the equation that defines the Green's function.

If $D_{\mathbf{x},t}$ is linear, then the superposition principle applies and the Green's function, in addition to being the linear response, can be used to write the solution of (D.1) for a general right-hand side as a convolution,

$$\phi(\mathbf{x}, t) = \int d\mathbf{x}' \int_0^\infty dt' \, G(\mathbf{x}, t; \mathbf{x}', t')f(\mathbf{x}', t') + u_h(\mathbf{x}, t) \tag{D.4}$$

where $u_h(\mathbf{x}, t)$ is a solution of the homogeneous equation (i.e. with $f(\mathbf{x}, t) = 0$), as can be checked by applying $D_{\mathbf{x},t}$ to the formal solution.

We are interested in cases where $D_{\mathbf{x},t}$ has the form

$$D_{\mathbf{x},t} = \chi \frac{\partial^2}{\partial t^2} + \eta \frac{\partial}{\partial t} - L_{\mathbf{x}}, \tag{D.5}$$

with $L_{\mathbf{x}}$ a linear, Hermitian, negative definite operator (i.e. a restoring force). This implies, in whole space at least, that $L_{\mathbf{x}}$ involves only even derivatives and ensures the $\mathbf{x} \longrightarrow -\mathbf{x}$ symmetry. We can write a formal expression of the Green's function in this case by expanding it in the base of eigenfunctions of $-L_{\mathbf{x}}$,

$$-L_{\mathbf{x}} u_n(\mathbf{x}) = \lambda_n u_n(\mathbf{x}), \tag{D.6}$$

with $\lambda_n \geq 0$ (i.e. $L_{\mathbf{x}}$ negative). In this base we write the Green's function as (we can put $t' = 0$)

$$G(\mathbf{x}, t; \mathbf{x}') = \sum_n G_n(t; \mathbf{x}')u_n(\mathbf{x}), \qquad G_n(t; \mathbf{x}') = (u_n, G) \tag{D.7}$$

where (u_n, G) is the inner product. Doing the inner product of u_n with (D.2) and using the fact that $L_{\mathbf{x}}$ is Hermitian, we obtain an ordinary differential equation for $G_n(t; \mathbf{x}')$:

$$\chi \ddot{G}_n + \eta \dot{G}_n + \lambda_n G_n = u_n^*(\mathbf{x}')\delta(t), \qquad G_n(t \to 0^-) = 0. \tag{D.8}$$

It is convenient for what follows to solve this using the Fourier–Laplace transform of $G_n(t)$,

$$G_n(\omega) = \int_0^\infty dt \, e^{i\omega} G_n(t), \qquad G_n(t) = \int_{-\infty}^\infty \frac{d\omega}{2\pi} e^{-i\omega t} G_n(\omega). \tag{D.9}$$

Introducing this in (D.8) one readily obtains

$$G_n(\omega; \mathbf{x}') = \frac{u_n^*(\mathbf{x}')}{-\chi\omega^2 - i\eta\omega + \lambda_n}. \tag{D.10}$$

From this we find $G_n(t; \mathbf{x}')$ by integration in the complex plane. We can close the path with a semi-circle of infinite radius. Due to the exponential factor, we must choose the path that encloses $\text{Im}\,\omega < 0$ ($\text{Im}\,\omega > 0$) half plane when $t > 0$ ($t < 0$). The poles of $G_n(\omega)$ are located at

$$\omega_\pm = -i\gamma \pm \omega_n, \qquad \gamma \equiv \frac{\eta}{2\chi}, \qquad \omega_n \equiv \sqrt{\frac{\lambda_n}{\chi} - \gamma^2}. \tag{D.11}$$

If $\lambda_n/\chi \geq \gamma^2$ then ω_n is real and both poles have negative imaginary part. If ω_n is imaginary, it is clear that $\text{Im}\,\omega_-$ is negative, while $\text{Im}\,\omega_+$ is negative provided $\lambda_n > 0$. Thus $\lambda_n > 0$, as we have assumed, means that both poles are always in the lower half-plane: this *ensures causality* of the response function, because for $t < 0$ the integration path includes the upper half-plane and all expansion coefficients vanish. So for $t \geq 0$ we have

$$G_n(t; \mathbf{x}') = \int_{-\infty}^\infty \frac{d\omega}{2\pi} \frac{e^{-i\omega t} u_n^*(\mathbf{x}')}{-\chi(\omega - \omega_+)(\omega - \omega_-)} = -i[\text{Res}(\omega_+) + \text{Res}(\omega_-)] \tag{D.12}$$

$$= \frac{iu_n^*(\mathbf{x}')}{\chi(\omega_+ - \omega_-)} \left[e^{-i\omega_+ t} - e^{-i\omega_- t} \right] \tag{D.13}$$

$$= u_n^*(\mathbf{x}') \, e^{-\gamma t} \frac{\sin \omega_n t}{\chi\omega_n}. \tag{D.14}$$

When ω_n is imaginary, it may be more convenient to define $\Gamma_n = i\omega_n$ and write instead

$$G_n(t; \mathbf{x}') = \frac{u_n^*(\mathbf{x}')}{2\chi\Gamma_n} \left[e^{-(\gamma - \Gamma_n)t} - e^{-(\gamma + \Gamma_n)t} \right]. \tag{D.15}$$

Finally the Green's function expansion in eigenvalues of $L_{\mathbf{x}}$ reads

$$G(\mathbf{x}, t; \mathbf{x}', t') = \sum_n u_n(\mathbf{x}) u_n^*(\mathbf{x}') \, e^{-\gamma t} \, \frac{\sin \omega_n(t - t')}{\chi \omega_n}. \tag{D.16}$$

If the region of interest is the whole space, the linear equation can be solved by Fourier transform, which is equivalent to saying that the eigenfunctions are $u_{\mathbf{k}}(\mathbf{x}) = e^{i\mathbf{k}\cdot\mathbf{x}}$, with $\mathbf{k} \in \mathcal{R}^d$, and (D.16) reads

$$G(\mathbf{x}, t; \mathbf{x}', t') = \frac{1}{(2\pi)^d} \int_{-\infty}^{\infty} d\mathbf{k} \, e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')} \, e^{-\gamma t} \, \frac{\sin \omega_{\mathbf{k}}(t - t')}{\chi \omega_{\mathbf{k}}}, \tag{D.17}$$

where the factor $(2\pi)^d$ appears because the plane waves are not normalized. This formula shows that $G(\mathbf{x}, t; \mathbf{x}', t')$ is a function of the difference $\mathbf{x} - \mathbf{x}'$, a manifestation of space translation invariance (resulting from the translation invariance of operator $L_{\mathbf{x}}$ plus the absence of borders). Let us finally write explicitly the Fourier transform of $G(\mathbf{x}, t)$ for later reference: it is equivalent to (D.10) with $\mathbf{x}' = 0$,

$$G(\mathbf{k}, \omega) = \frac{1}{-\chi \omega^2 - i\eta\omega + \lambda_{\mathbf{k}}}. \tag{D.18}$$

Appendix E. Dispersion relation

The dispersion relation is a functional relationship $\omega(\mathbf{k})$ between frequency and wavelength stating which plane waves can propagate in the medium described by a given differential equation. This relation, as we shall show, is a direct consequence of the pole structure of the Green's function in Fourier space. Consider

$$\chi \frac{\partial^2 \phi}{\partial t^2} + \eta \frac{\partial \phi}{\partial t} - L_{\mathbf{x}} \phi(\mathbf{x}, t) = f(\mathbf{x}, t), \tag{E.1}$$

i.e. (D.1) with $D_{\mathbf{x},t}$ given by (D.5). The dispersion relation is obtained by finding the poles, i.e. solving

$$G^{-1}(\mathbf{k}, \omega) = -\chi \omega^2 - i\eta\omega + \lambda_{\mathbf{k}} = 0. \tag{E.2}$$

To see why plane waves must obey the dispersion relation, we write the solution as the inverse transform of the product of $G(\mathbf{k}, \omega)$ and the Fourier transform of the source $f(\mathbf{x}, t)$:

$$\phi(\mathbf{x}, t) = \frac{1}{(2\pi)^{d+1}} \int d\mathbf{k} \, d\omega \, e^{i\mathbf{k}\cdot\mathbf{x} - i\omega t} G(\mathbf{k}, \omega) f(\mathbf{k}, \omega). \tag{E.3}$$

Unless $f(\mathbf{x}, t)$ is exponentially divergent in time, $f(\mathbf{k}, \omega)$ will not have poles in $\operatorname{Im} \omega < 0$, so that the integral over ω can be done in the complex plane exactly as in Appendix D, and the only relevant poles are those of $G(\mathbf{k}, \omega)$. One obtains

$$\phi(\mathbf{x}, t) = \frac{1}{(2\pi)^d} \int d\mathbf{k} \, \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{i\chi} \left[\frac{e^{-i\omega_+(\mathbf{k})t}}{\omega_+(\mathbf{k}) - \omega_-(\mathbf{k})} f\left(\mathbf{k}, \omega_+(\mathbf{k})\right) + \frac{e^{-i\omega_-(\mathbf{k})t}}{\omega_-(\mathbf{k}) - \omega_+(\mathbf{k})} f\left(\mathbf{k}, \omega_-(\mathbf{k})\right) \right]$$

$$= \frac{1}{(2\pi)^d} \int d\mathbf{k} \, \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{i\chi \omega_{\mathbf{k}}} e^{-\gamma t} \left[e^{-i\omega_{\mathbf{k}} t} f(\mathbf{k}, -i\gamma + \omega_{\mathbf{k}}) - e^{i\omega_{\mathbf{k}} t} f(\mathbf{k}, -i\gamma - \omega_{\mathbf{k}}) \right], \tag{E.4}$$

where, we remind, $\omega_{\mathbf{k}} = \sqrt{\lambda_{\mathbf{k}}/\chi - \gamma^2}$ (see Eq. (D.11)). This expression makes it clear that the solution is a superposition of (over)damped plane waves obeying the dispersion relation and with amplitudes given by the Fourier components of the source.

A more convenient expression for $\phi(\mathbf{x}, t)$ can be written, which is manifestly real for real $f(\mathbf{x}, t)$ and that shows the different behavior of damped and overdamped waves. ω_k will be real (pure imaginary) in a region $\mathcal{R}_r$ ($\mathcal{R}_i$). Assuming $\lambda_{\mathbf{k}}$ is an even function of $\mathbf{k}$ (as must be the case when $L_{\mathbf{x}}$ has only even derivatives), both regions must be symmetric with respect to the origin (i.e. $\mathbf{k} \in \mathcal{R}_r$ if and only if $-\mathbf{k} \in \mathcal{R}_r$, and similarly for $\mathcal{R}_i$). Since $f(\mathbf{k}, t)$ is real, the following conditions hold:

$$f^*(\mathbf{k}, \omega_+) = f(-\mathbf{k}, \omega_-), \qquad\qquad\qquad \mathbf{k} \in \mathcal{R}_r, \tag{E.5}$$

$$f^*(\mathbf{k}, \omega_+) = f(-\mathbf{k}, \omega_+), \qquad\qquad\qquad \mathbf{k} \in \mathcal{R}_i, \tag{E.6}$$

$$f^*(\mathbf{k}, \omega_-) = f(-\mathbf{k}, \omega_-), \qquad\qquad\qquad \mathbf{k} \in \mathcal{R}_i. \tag{E.7}$$

Using these conditions, the solution can be transformed as follows: For $\mathbf{k} \in \mathcal{R}_r$, take the second term in (E.4) and change the integration variable from $\mathbf{k}$ to $-\mathbf{k}$. Then it is seen that the second term is the complex conjugate of the first term, the difference of the two yielding a pure imaginary quantity which becomes real after dividing by $i\chi$. For $\mathbf{k} \in \mathcal{R}_i$, divide the integral in the two symmetric parts around the origin and change $\mathbf{k}$ to $-\mathbf{k}$ in one of them (say the negative part). Then combining the integrals again results in the sum of an expression with its complex conjugate, resulting in a real quantity, while the denominator is now real because $\omega_{\mathbf{k}}$ is imaginary. Defining $i\Gamma_{\mathbf{k}} = \omega_{\mathbf{k}}$ and $f(\mathbf{k}, \omega) = f'(\mathbf{k}, \omega) + if''(\mathbf{k}, \omega)$, the final result is

$$\phi(\mathbf{x}, t) = \frac{1}{(2\pi)^d} \int_{\mathcal{R}_r} d\mathbf{k} \, \frac{2e^{-\gamma t}}{\chi \omega_{\mathbf{k}}} \left[f''(\mathbf{k}, \omega_{\mathbf{k}} - i\gamma) \cos(\mathbf{k} \cdot \mathbf{x} - \omega_{\mathbf{k}} t) + \right.$$

$$f'(\mathbf{k}, \omega_{\mathbf{k}} - i\gamma)\sin(\mathbf{k} \cdot \mathbf{x} - \omega_{\mathbf{k}}t)\Bigg] +$$

$$\frac{1}{(2\pi)^d}\int_{\mathcal{R}_i^+}d\mathbf{k}\,\frac{2e^{-(\gamma+\Gamma_{\mathbf{k}})t}}{\chi\,\Gamma_{\mathbf{k}}}\Bigg[f'(\mathbf{k}, -i\Gamma_k - i\gamma)\cos(\mathbf{k} \cdot \mathbf{x})$$

$$f''(\mathbf{k}, -i\Gamma_k - i\gamma)\sin(\mathbf{k} \cdot \mathbf{x})\Bigg] -$$

$$\frac{1}{(2\pi)^d}\int_{\mathcal{R}_i^+}d\mathbf{k}\,\frac{2e^{-(\gamma-\Gamma_{\mathbf{k}})t}}{\chi\,\Gamma_{\mathbf{k}}}\Bigg[f'(\mathbf{k}, i\Gamma_k - i\gamma)\cos(\mathbf{k} \cdot \mathbf{x})$$

$$f''(\mathbf{k}, i\Gamma_k - i\gamma)\sin(\mathbf{k} \cdot \mathbf{x})\Bigg]. \tag{E.8}$$

Appendix F. Computation of $C(r, t)$ in ordered phase of VM and ISM

To obtain the time correlation of the field $\varphi(\mathbf{x}, t)$ ruled by Eq. (157) or (158) one first writes the field explicitly in terms of the external (random) force, then computes the correlation $\langle\varphi(\mathbf{x}, t)\varphi(0, 0)\rangle$ by averaging over the realizations of the random force. The averaging is actually simple because in practice only the second moment of the force is needed.

Since the equations are linear in $\varphi(\mathbf{x}, t)$, they can be solved by finding the Green's function, which can be done expanding in eigenfunctions (Appendix D). Eqs. (157) and (158) are of the form (D.5) with $L_{\mathbf{x}} = Jn_c a^2\nabla_{\mathbf{x}}^2$. The eigenfunctions of $-L_{\mathbf{x}}$ are plane waves, and the corresponding eigenvalues are $\lambda_{\mathbf{k}} = Jn_c a^2 k^2$. The Fourier–Laplace transform of the Green's function is obtained by plugging the eigenvalues into (D.18):

$$G(k, \omega) = \begin{cases} \left(-i\eta\omega + Jn_c a^2 k^2\right)^{-1}, & \text{VM} \\ \left(-\chi\omega^2 - i\eta\omega + Jn_c a^2 k^2\right)^{-1}. & \text{ISM} \end{cases} \tag{F.1}$$

Now the Green's function can be used to write the full time evolution of the field. It is convenient to place the initial conditions at $t = -\infty$; then the system loses memory of them and then fields depend only on the random force,

$$\varphi(\mathbf{x}, t) = \int_{-\infty}^{\infty}dt'\int d\mathbf{x}'\,G(\mathbf{x} - \mathbf{x}', t - t')\zeta(\mathbf{x}', t'), \tag{F.2}$$

where the Green's function appropriate for each model must be used. Now the formal solution (F.2) can be used in (159) to obtain

$$C(\mathbf{x}, t) = \int_{-\infty}^{\infty}dt_1\,dt_2\int d\mathbf{x}_1\,d\mathbf{x}_2\,G(\mathbf{x} - \mathbf{x}_1, t - t_1)G(0 - \mathbf{x}_2, 0 - t_2)\,\langle\zeta(\mathbf{x}_1, t_1)\zeta(\mathbf{x}_2, t_2)\rangle,$$

$$= 2\eta Ta^3\int_{-\infty}^{\infty}dt_1\int d\mathbf{x}_1\,G(\mathbf{x} - \mathbf{x}_1, t - t_1)G(-\mathbf{x}_1, -t_1), \tag{F.3}$$

where the average could be performed immediately since the correlation of the random force is known (it is delta-correlated in space and time). To deal with the convolution it is convenient then to introduce the Fourier transform and write finally

$$C(\mathbf{k}, t)\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}e^{-i\omega t}C(\mathbf{k}, \omega), \tag{F.4}$$

$$C(\mathbf{k}, \omega) = 2\eta Ta^3 G(\mathbf{k}, \omega)G(-\mathbf{k}, -\omega). \tag{F.5}$$

Formulae (F.4) and (F.5) are the expressions that allow us to analyze the qualitative features of $C(\mathbf{k}, t)$. The shape of the correlation depends crucially on the pole structure of the Green's function in Fourier space $G(\mathbf{k}, \omega)$. A general property of the case when the eigenvalues $\lambda_{\mathbf{k}}$ are positive is that $G(\mathbf{k}, \omega)$ has poles only for $\text{Im}\,\omega < 0$, which ensures that the response is causal (see Appendix D). Additionally, in the present case, the eigenvalues are quadratic in k, which causes $G(\mathbf{k}, \omega)$ to be even in $\mathbf{k}$, as is obvious from (F.1). Then from (F.5) we see that the poles of the correlation $C(\mathbf{k}, \omega)$ are symmetric for $\text{Im}\,\omega < 0$ and $\text{Im}\,\omega > 0$. This implies that $C(\mathbf{k}, t)$ is even in *time,* because the integral in (F.4) can be solved by integrating in the complex ω-plane using a path that encloses the negative imaginary semi-plane if $t > 0$, or the positive imaginary semi-plane if $t < 0$. But since the poles are symmetric, the result will be identical in both cases (at fixed k of course).

To work out $C(\mathbf{k}, t)$ one needs explicitly the poles of $C(\mathbf{k}, \omega)$, so we rewrite $G(k, \omega$ and $C(k, \omega)$ in a way that makes evident the location of the poles and the value of the residue. Recalling the definitions of the convenience constants (163) and (162), we have, for the Vicsek case,

$$G(\mathbf{k}, \omega) = \frac{1}{-i\eta(\omega + i\tau_0^{-1})}, \qquad \text{VM}, \tag{F.6}$$

$$C(\mathbf{k}, \omega) = \frac{2Ta^3}{\eta}\frac{1}{\omega^2 + \tau_0^{-2}}, \qquad \text{VM}, \tag{F.7}$$

and for the ISM,

$$G(\mathbf{k}, \omega) = \frac{1}{-\chi(\omega - \omega_+)(\omega - \omega_-)}, \qquad \text{ISM,} \tag{F.8}$$

$$C(\mathbf{k}, \omega) = \frac{2a^3\eta T}{\chi^2} \frac{1}{(\omega^2 - \omega_+^2)(\omega^2 - \omega_-^2)}, \qquad \text{ISM.} \tag{F.9}$$

From (F.7) and (F.9) we can finally compute $C(\mathbf{k}, t)$ by integrating over ω in the complex plane as done in Appendix D for the Green's function (see comments after Eq. (D.10)). The symmetric placement of the poles means that the result involves half of the poles of $C(\mathbf{k}, \omega)$ (one for Vicsek, two for ISM), since the integration contour selects only half of the imaginary plane. For instance, for the VM and $t > 0$ the integration path encloses the lower semi-plane, and the relevant pole is $\omega_p = -i\tau_0^{-1}$. The integral is then computed as (the first minus sign arises because the residue theorem uses a path that runs counterclockwise which takes the real ω axis from ∞ to $-\infty$):

$$C(\mathbf{k}, t) = -\frac{2Ta^3}{\eta} \oint \frac{d\omega}{2\pi} e^{-i\omega t} \frac{1}{\omega^2 + \tau_0^2} = -\frac{2Ta^3}{2\pi\eta} 2\pi i \text{Res}|_{-i\tau_0} = -\frac{2Ta^3}{\eta} i e^{-t/\tau_0} \frac{1}{-2i\tau_0^{-1}}. \tag{F.10}$$

This gives (164). For the ISM the same procedure, now involving two poles (ω_+ and ω_- for $t > 0$) yields (165) after some algebra.

Appendix G. Structure of the correlation function in the complex ω-plane

To interpret the non-exponential form of $C(k, t)$ it is useful to reason in terms of the poles of its Fourier transform $C(k, \omega)$ in the complex ω-plane, as their structure reflects the dispersion relation of the system and thus the underlying equation of motion [76]. What we will prove here is that exponential relaxation in time derives from a *single* pole of $C(k, \omega)$ on the positive imaginary semi-plane, while a vanishing first derivative of the temporal correlation implies the existence of *two*, or more, poles of $C(k, \omega)$ in the positive imaginary semi-plane. From the Fourier relation

$$C(t) = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{-i\omega t} C(\omega), \tag{G.1}$$

we have that the time derivative of the correlation function is given by

$$\dot{C}(t) = -\int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{-i\omega t} F(\omega), \qquad F(\omega) = i\omega C(\omega). \tag{G.2}$$

From the physical condition $C(t) = C(-t)$, and therefore $C(\omega) = C(-\omega)$, we obtain that the poles of $C(\omega)$ must have a symmetric structure,

$$C(\omega) = \frac{1}{\prod_{i=1}^{K}(\omega - \omega_i)^{n_i}(\omega + \omega_i)^{n_i}}, \tag{G.3}$$

where we admit that some pole may have multiplicity n_i larger than one.

The $t \to 0^+$ limit of $\dot{C}(t)$ in (G.2) can be computed with the residue theorem by integrating $F(\omega)$ along the path in Fig. G.20. Because $F(-\omega) = -F(\omega)$, we have,

$$\text{Res}\,(F(\omega), +\omega_i) = \text{Res}\,(F(\omega), -\omega_i) \qquad \forall i = 1, \dots, K$$

so that, after some algebra, we obtain

$$\lim_{t \to 0^+} \dot{C}(t) = \frac{1}{2} \sum_{i=1}^{K} \left[\text{Res}\,(F(\omega), +\omega_i) + \text{Res}\,(F(\omega), -\omega_i) \right]. \tag{G.4}$$

The sum of all the residues of $F(\omega)$ coincides with its residue at infinity, $\text{Res}\,(F(\omega), \infty)$, which can be computed as the residue in $z = 0$ of the function $\hat{F}(z) = F(1/z)/z^2$,

$$\text{Res}\left(\hat{F}(z), 0\right) = \lim_{\epsilon \to 0} \oint_{\mathcal{C}(\epsilon)} dz \frac{z^{(2\sum_i n_i - 3)}}{\prod_i (z^2 - 1/\omega_i^2)^{n_i} \prod_i \omega_i^{2n_i}},$$

where $\mathcal{C}(\epsilon)$ is a circle of radius ϵ centered in the origin. The integral above is easily calculated, so that (G.4) becomes,

$$\lim_{t \to 0^+} \dot{C}(t) = \text{Res}\left(\hat{F}(z), 0\right) = \begin{cases} 1 & \text{if } \sum_i n_i = 1, \\[2mm] 0 & \text{if } \sum_i n_i \geq 2. \end{cases}$$

We conclude that a single pole in the positive semi-plane implies a non-zero first derivative of the time correlation function; more precisely, in this case $C(\omega)$ is a Lorentzian, so that $C(t)$ is purely exponential. On the other hand, a vanishing first
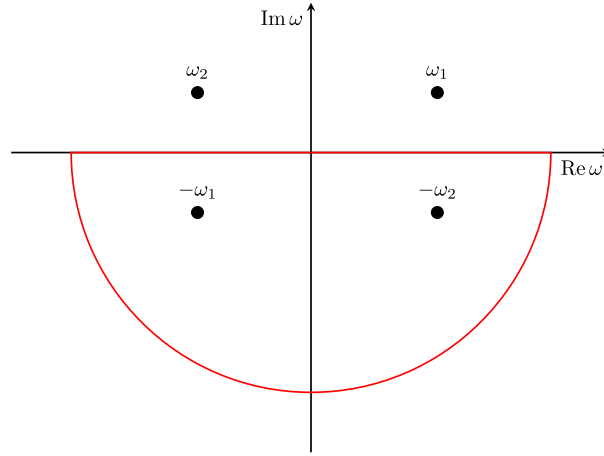
Fig. G.20. The integration path contains all poles of $C(\omega)$ with non-positive imaginary part (assuming there are two such poles).

derivative of the time correlation function $C(t)$ for $t \to 0$ is caused by the existence of two, or more, poles of its Fourier transform $C(\omega)$ in the positive imaginary semi-plane.

The structure of these poles reflects the structure of the dispersion polynomial of the theory; in particular, multiple poles with a non-zero real part are the most distinctive hallmark of propagating spin-waves [54]. In the overdamped, paramagnetic phase the real part of the spin-wave poles vanishes and the poles move onto the imaginary axis. Yet, their multiple structure (namely, the fact that they are more than one), remains as a remnant of the spin-wave phase and, as we have seen here, this remnant shows up as a zero derivative of the time correlation function. When we push a paramagnetic system deeply into its overdamped phase, i.e. down to the hydrodynamic phase, some of these poles becomes so large (high frequencies) that we no longer have the experimental resolution to see their effect in the derivative of $C(t)$, and we observe purely exponential relaxation.

Appendix H. Damped harmonic oscillator as toy model for velocity correlations

To get an intuitive grasp on how a time correlation function that is generated by a dynamics with multiple poles, and hence flat at $t = 0$, can crossover to a correlation with finite derivative as the damping is increased, we can study the over-damped stochastic harmonic oscillator [55]. It is described by

$$\chi \ddot{u}(t) + \eta \dot{u}(t) + \kappa u(t) = \zeta(t). \tag{H.1}$$

This is identical to the space Fourier transform of (158) evaluated at a fixed $k^2 = \kappa / J n_c a^2$, so that we can read the time correlation from (165):

$$\hat{C}(t) \equiv \frac{C(t)}{C(t = 0)} = e^{-\gamma t} \left[\cos \hat{\omega} t + \frac{\gamma}{\hat{\omega}} \sin \hat{\omega} t \right], \tag{H.2}$$

with

$$\gamma = \frac{\eta}{2\chi}, \qquad \omega_0^2 = \frac{\kappa}{\chi}, \qquad \hat{\omega} = \sqrt{\omega_0^2 - \gamma^2}. \tag{H.3}$$

The Green's function can be obtained in the same way from (160) but is also easily found directly to be

$$G(\omega) = \frac{1}{-\chi \omega^2 - i\omega \eta + \kappa} = \frac{1}{-\chi (\omega - \omega^+)(\omega - \omega^-)}, \qquad \omega^\pm = -i\gamma \pm \hat{\omega} \tag{H.4}$$

We can distinguish three regimes:

1. *Underdamped regime:* This is when $\gamma / \omega_0 < 1$ and inertia dominates over viscosity. The two poles of the Green's function have a large nonzero real part and a small imaginary part, $\hat{\omega}$ is real and $\hat{C}(t)$ displays a clear oscillatory behavior (Fig. H.21a). This regime is the analogous of the propagating spin-wave phase of ferromagnets [54].
2. *Critically damped regime:* When $\gamma / \omega_0 = 1$ inertia and viscosity exactly balance each other, the two poles coincide on the imaginary axis, and $\hat{C}(t)$ does not oscillate but retains a clear non-exponential form with a flat correlation for small times (Fig. H.21b). This is analogous to the critical point between ferromagnetic and paramagnetic phase.
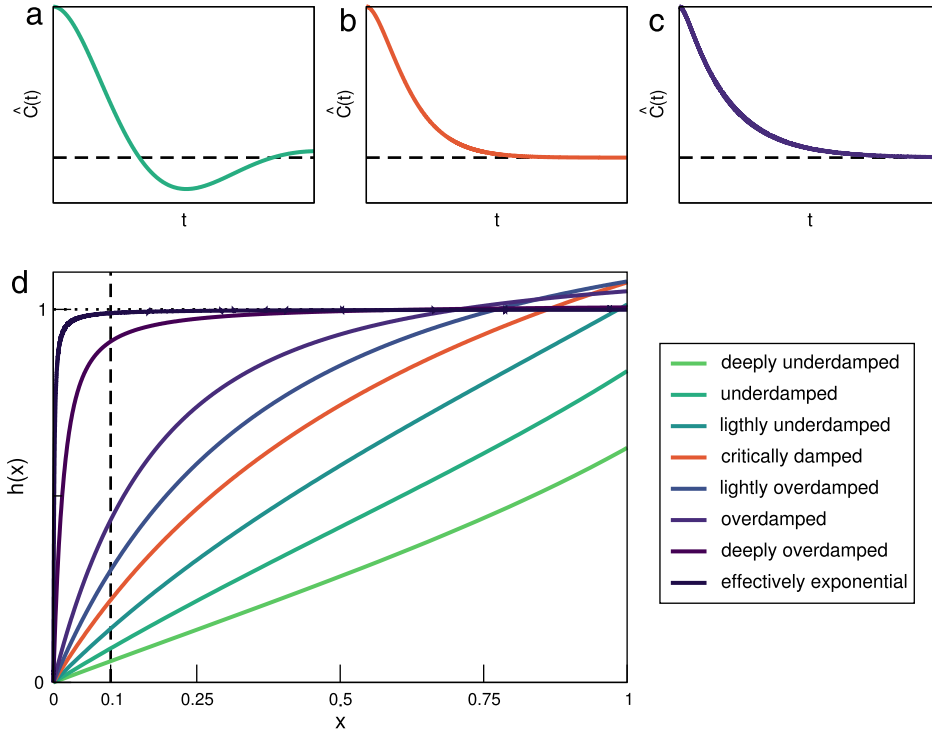
Fig. H.21. Normalized correlation function for the harmonic oscillator in the under-damped (a), critically damped (b) and over-damped (c) regimes. Panel (d) shows the function $h(x)$, Eq. (H.6) in the different regimes. For an exponential decay, $h(x) \to 1$ for $x \to 0$, so when the t/τ is less than the experimental resolution (dotted vertical line), the decay is effectively exponential.

3. *Over-damped regime:* When $\gamma/\omega_0 > 1$ both poles are imaginary and separated, with one of them approaching the origin ($\omega^- \sim i\omega_0^2/2\gamma \ll 1$ and $\omega^+ \sim 2i\gamma$) and the time correlation becomes more and more exponential. This regime is the analogous of the paramagnetic phase (Fig. H.21c).

Thus upon increasing the damping $\hat{C}(t)$ turns from an oscillatory, far-from-exponential behavior in the under-damped regime to a non-oscillatory, nearly-exponential function behavior in the over-damped regime (Fig. H.21d). Yet it is straightforward to check from (H.2) that

$$\lim_{t \to 0^+} \frac{d\hat{C}(t)}{dt} = 0 \tag{H.5}$$

for all η. This is in accordance with the general result that when the Green's function has more than one pole in the complex ω-plane, the first derivative of $C(t)$ vanishes at $t = 0$ (Appendix G). So, how does it happen that the correlation becomes "more and more exponential" in the over-damped regime?

What happens for increasing damping can be understood by following the evolution of the function

$$h(x) \equiv -\frac{1}{x} \log \hat{C}(x), \qquad x \equiv t/\tau. \tag{H.6}$$

Since $\hat{C}(0) = 1$, a zero first derivative for $t \to 0$ implies

$$\lim_{x \to 0} h(x) = 0, \tag{H.7}$$

while for pure exponential one would have

$$\lim_{x \to 0} h(x) = 1. \tag{H.8}$$

In practice, then, the crossover from (H.7) to (H.8) happens as displayed in Fig. H.21: although $h(x)$ is always zero at exactly $x = 0$, by increasing the damping the value of $x = t/\tau$ where $h(x)$ departs from 1 becomes smaller and smaller. Our experimental apparatus will always have a finite time resolution (it is unphysical to think to be able to resolve the correlation for $x = t/\tau$ arbitrarily small). Say that this resolution is $t/\tau = \epsilon$. This means that beyond a certain damping we are doomed to observe $h(x \sim \epsilon) \sim 1$ within our experimental resolution, and the time correlation becomes therefore purely exponential

for all practical purposes. On the other hand, in the weakly over-damped regime the departure from the exponential case is strong: the limit of $h(x)$ form small x is clearly far from 1 even within our experimental resolution $x > \epsilon$. This is the mechanism underlying the existence of paramagnetic spin-wave remnants: although all the explicit oscillatory phenomena of spin waves are absent, the strong non-exponential character of the correlation function *in the experimentally relevant time regime* $t \sim \tau$ is clear evidence that the original equation of motion admits spin-waves in a certain region of the parameter space and it is therefore second order in time.

References

[1] D.J. Sumpter, Collective Animal Behavior, Princeton University Press, 2010.
[2] T. Vicsek, A. Zafeiris, Collective motion, Phys. Rep. 517 (3) (2012) 71–140.
[3] S. Ramaswamy, The mechanics and statistics of active matter, Annu. Rev. Condens. Matter Phys. 1 (2010) 323.
[4] M. Marchetti, J. Joanny, S. Ramaswamy, T. Liverpool, J. Prost, M. Rao, R.A. Simha, Hydrodynamics of soft active matter, Rev. Modern Phys. 85 (3) (2013) 1143.
[5] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, O. Shochet, Novel type of phase transition in a system of self-driven particles, Phys. Rev. Lett. 75 (6) (1995) 1226–1229.
[6] J. Toner, Y. Tu, Long-range order in a two-dimensional dynamical XY model: How birds fly together, Phys. Rev. Lett. 75 (23) (1995) 4326–4329.
[7] J. Toner, Y. Tu, S. Ramaswamy, Hydrodynamics and phases of flocks, Ann. Physics 318 (1) (2005) 170–244.
[8] T. Mora, A.M. Walczak, L. Del Castello, F. Ginelli, S. Melillo, L. Parisi, M. Viale, A. Cavagna, I. Giardina, Local equilibrium in bird flocks, Nat. Phys. 12 (12) (2016) 1153–1157.
[9] A. Cavagna, A. Cimarelli, I. Giardina, A. Orlandi, G. Parisi, A. Procaccini, R. Santagati, F. Stefanini, New statistical tools for analyzing the structure of animal groups, Math. Biosci. 214 (1–2) (2008) 32–37.
[10] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, M. Viale, Scale-free correlations in starling flocks, Proc. Natl. Acad. Sci. USA 107 (26) (2010) 11865–11870.
[11] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, S. Melillo, L. Parisi, O. Pohl, B. Rossaro, E. Shen, E. Silvestri, et al., Collective behaviour without collective order in wild swarms of midges, PLoS Comput. Biol. 10 (7) (2014) e1003697.
[12] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, S. Melillo, L. Parisi, O. Pohl, B. Rossaro, E. Shen, E. Silvestri, et al., Finite-size scaling as a way to probe near-criticality in natural swarms, Phys. Rev. Lett. 113 (23) (2014) 238102.
[13] J.-P. Hansen, I.R. McDonald, Theory of Simple Liquids, Elsevier, 1990.
[14] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, V. Zdravkovic, Empirical investigation of starling flocks: a benchmark study in collective animal behaviour, Anim. Behav. 76 (2008) 201–215.
[15] S. keng Ma, Modern theory of critical phenomena, in: Advanced book classics, Perseus Pub, 2000.
[16] G. Parisi, Statistical field theory, in: Frontiers in Physics, Addison-Wesley, Redwood City, CA, 1988. URL: https://cds.cern.ch/record/111935.
[17] A. Cavagna, I. Giardina, A. Orlandi, G. Parisi, A. Procaccini, The starflag handbook on collective animal behaviour: 2. Three-dimensional analysis, Anim. Behav. 76 (2008) 237–248.
[18] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, et al., Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study, Proc. Natl. Acad. Sci. 105 (4) (2008) 1232–1237.
[19] A. Cavagna, L. Del Castello, S. Dey, I. Giardina, S. Melillo, L. Parisi, M. Viale, Short-range interactions versus long-range correlations in bird flocks, Phys. Rev. E 92 (1) (2015) 012705.
[20] N. Goldenfeld, Lectures on Phase Transitions and the Renormalization Group, Perseus Books, Reading, Massachusetts, 1992.
[21] M. Le Bellac, Quantum and Statistical Field Theory, Clarendon Press Oxford, 1991.
[22] I.D. Couzin, J. Krause, R. James, G.D. Ruxton, N.R. Franks, Collective memory and spatial sorting in animal groups, J. Theoret. Biol. 218 (1) (2002) 1–11.
[23] G. Grégoire, H. Chaté, Onset of collective and cohesive motion, Phys. Rev. Lett. 92 (2) (2004) 025702.
[24] H. Chaté, F. Ginelli, G. Grégoire, F. Peruani, F. Raynaud, Modeling collective motion: variations on the Vicsek model, Eur. Phys. J. B 64 (3–4) (2008) 451–456.
[25] H. Chaté, F. Ginelli, G. Grégoire, F. Raynaud, Collective motion of self-propelled particles interacting without cohesion, Phys. Rev. E 77 (4 Pt 2) (2008) 046113.
[26] C.K. Hemelrijk, H. Hildenbrandt, Scale-free correlations, influential neighbours and speed control in flocks of birds, J. Stat. Phys. 158 (3) (2015) 563–578.
[27] G. Barkema, M. Newman, Monte Carlo Methods in Statistical Physics, Oxford University Press, 2001.
[28] J.J. Binney, N. Dowrick, A. Fisher, M. Newman, The Theory of Critical Phenomena: An Introduction to the Renormalization Group, Oxford University Press, Inc., 1992.
[29] A. Cavagna, D. Conti, I. Giardina, T.S. Grigera, S. Melillo, M. Viale, Spatio-temporal correlations in models of collective motion ruled by different dynamical laws, Phys. Biol. 13 (2016) 065001.
[30] E.T. Jaynes, Information theory and statistical mechanics, Phys. Rev. 106 (1957) 620.
[31] E.T. Jaynes, Information theory and statistical mechanics. II, Phys. Rev. 108 (1957) 171.
[32] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, A.M. Walczak, Statistical mechanics for natural flocks of birds, Proc. Natl. Acad. Sci. USA 109 (13) (2012) 4786–4791.
[33] W. Bialek, A. Cavagna, I. Giardina, T. Mora, O. Pohl, E. Silvestri, M. Viale, A. Walczak, Social interactions dominate speed control in driving natural flocks toward criticality, 2013. ArXiv physics.bio-ph arXiv:1307.5563v1.
[34] S. Camazine, N.R. Franks, J. Sneyd, E. Bonabeau, J.-L. Deneubourg, G. Theraula, Self-Organization in biological systems, Princeton University Press, Princeton, NJ, USA, 2001.
[35] I. Couzin, J. Krause, Self-organization and collective behavior in vertebrates, Adv. Study Behav. 32 (2003) 1–75.
[36] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, T.S. Grigera, A. Jelić, S. Melillo, L. Parisi, O. Pohl, E. Shen, et al., Information transfer and behavioural inertia in starling flocks, Nat. Phys. 10 (9) (2014) 691–696.
[37] A. Cavagna, I. Giardina, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, V. Zdravkovic, The starflag handbook on collective animal behaviour: 1. Empirical methods, Anim. Behav. 76 (2008) 217–236.
[38] G.R. Conner, C.P. Grant, An extension of zermelo's model for ranking by paired comparisons, European J. Appl. Math. 11 (03) (2000) 225–247.
[39] D.V. Radakov, Schooling in the ecology of fish, 1973.
[40] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, A. Jelic, S. Melillo, L. Parisi, O. Pohl, E. Shen, M. Viale, Emergence of collective changes in travel direction of starling flocks from individual birds' fluctuations, J. R. Soc. Interface 12 (108) (2015) 20150319.
[41] H. Pomeroy, F. Heppner, Structure of turning in airborne rock dove (columba livia) flocks, The Auk (1992) 256–267.
[42] J. Krause, G.D. Ruxton, Living in Groups, Oxford University Press, 2002.

[43] F. Ginelli, The physics of the vicsek model, Eur. Phys. J. Spec. Top. 225 (11–12) (2016) 2099–2117.
[44] J. Niel, J. Zinn-Justin, Finite size effects in critical dynamics, Nuclear Phys. B 280 (1987) 355–384.
[45] A. Cavagna, I. Giardina, F. Ginelli, T. Mora, D. Piovani, R. Tavarone, A.M. Walczak, Dynamical maximum entropy approach to flocking, Phys. Rev. E 89 (4) (2014) 042707.
[46] A. Cavagna, L. Del Castello, I. Giardina, T. Grigera, A. Jelic, S. Melillo, T. Mora, L. Parisi, E. Silvestri, M. Viale, et al., Flocking and turning: a new model for self-organized collective motion, J. Stat. Phys. 158 (3) (2015) 601–627.
[47] Y. Tu, J. Toner, M. Ulm, Sound waves and the absence of galilean invariance in flocks, Phys. Rev. Lett. 80 (21) (1998) 4819–4822.
[48] J. Toner, Y. Tu, Flocks, herds, and schools: a quantitative theory of flocking, Phys. Rev. E 58 (4) (1998) 4828.
[49] J. Toner, Reanalysis of the hydrodynamic theory of fluid, polar-ordered flocks, Phys. Rev. E 86 (3 Pt 1) (2012) 031918.
[50] E. Bertin, M. Droz, G. Grégoire, Boltzmann and hydrodynamic description for self-propelled particles, Phys. Rev. E 74 (2006) 022101.
[51] E. Bertin, M. Droz, G. Grégoire, Hydrodynamic equations for self-propelled particles: microscopic derivation and stability analysis, J. Phys. A 42 (2009) 445001.
[52] A. Peshkov, S. Ngo, E. Bertin, H. Chaté, F. Ginelli, Continuous theory of active matter systems with metric-free interactions, Phys. Rev. Lett. 109 (2012) 098101.
[53] Y. Tu, J. Toner, M. Ulm, Sound waves and the absence of galilean invariance in flocks, Phys. Rev. Lett. 80 (21) (1998) 4819.
[54] P.C. Hohenberg, B.I. Halperin, Theory of dynamic critical phenomena, Rev. Modern Phys. 49 (3) (1977) 435.
[55] R. Zwanzig, Nonequilibrium Statistical Mechanics, Oxford University Press, USA, 2001.
[56] T. Matsubara, H. Matsuda, A lattice model of liquid helium, I, Progr. Theoret. Phys. 16 (6) (1956) 569–582.
[57] H. Matsuda, T. Matsubara, A lattice model of liquid helium, II, Progr. Theoret. Phys. 17 (1) (1957) 19–29.
[58] B. Halperin, P. Hohenberg, Hydrodynamic theory of spin waves, Phys. Rev. 188 (2) (1969) 898.
[59] A. Cavagna, I. Giardina, T.S. Grigera, A. Jelic, D. Levine, S. Ramaswamy, M. Viale, Silent flocks: constraints on signal propagation across biological groups, Phys. Rev. Lett. 114 (21) (2015) 218101.
[60] A.Z. Patashinskii, V.L. Pokrovskii, Zh. Eksp. Teor. Fiz. 64 (1973) 1445.
[61] A.Z. Patashinskii, V.L. Pokrovskii, Fluctuation Theory of Phase Transitions, Pergamon Press, 1979.
[62] J. Goldstone, Field theories with superconductor solutions, Il Nuovo Cimento (1955–1965) 19 (1) (1961) 154–164.
[63] B.D. Josephson, Relation between the superfluid density and order parameter for superfluid He near Tc, Phys. Lett. 21 (6) (1966) 608–609.
[64] B.I. Halperin, P.C. Hohenberg, Scaling laws for dynamic critical phenomena, Phys. Rev. 177 (1969) 952–971.
[65] A. Cavagna, D. Conti, C. Creato, L. Del Castello, I. Giardina, T.S. Grigera, S. Melillo, L. Parisi, M. Viale, Dynamic scaling in natural swarms, Nature Phys. 13 (9) (2017) 914–918.
[66] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, A. Jelić, S. Melillo, L. Parisi, F. Pellacini, E. Shen, E. Silvestri, et al., GReTA-a novel global and recursive tracking algorithm in three dimensions, IEEE Trans. Pattern Anal. Mach. Intell. 37 (12) (2015) 2451–2463.
[67] A. Cavagna, C. Creato, L. Del Castello, I. Giardina, S. Melillo, L. Parisi, M. Viale, Error control in the set-up of stereo camera systems for 3d animal tracking, Eur. Phys. J. Spec. Top. 224 (17) (2015) 3211–3232.
[68] E. Schneidman, M.J. Berry, R. Segev, W. Bialek, Weak pairwise correlations imply strongly correlated network states in a neural population, Nature 440 (7087) (2006) 1007–1012;
       A. Haimovici, E. Tagliazucchi, P. Balenzuela, D.R. Chialvo, Brain organization into resting state networks emerges at criticality on a model of the human connectome, Phys. Rev. Lett. 110 (17) (2013) 171801;
       D. Fraiman, D.R. Chialvo, What kind of noise is brain noise: anomalous scaling behavior of the resting brain activity fluctuations, Front. Physiol. 3 (2012) 307.
[69] C. Dombrowski, L. Cisneros, S. Chatkaew, R.E. Goldstein, J.O. Kessler, Self-concentration and large-scale coherence in bacterial dynamics, Phys. Rev. Lett. 93 (9) (2004) 098103.
[70] H.-P. Zhang, A. Beer, E.-L. Florin, H.L. Swinney, Collective motion and density fluctuations in bacterial colonies, Proc. Natl. Acad. Sci. 107 (31) (2010) 13626–13630.
[71] G. Duclos, S. Garcia, H. Yevick, P. Silberzan, Perfect nematic order in confined monolayers of spindle-shaped cells, Soft Matter 10 (14) (2014) 2346–2353.
[72] Q.-Y. Tang, Y.-Y. Zhang, J. Wang, W. Wang, D.R. Chialvo, Critical fluctuations in the native state of proteins, Phys. Rev. Lett. 118 (8) (2017) 088102.
[73] T. Mora, W. Bialek, Are biological systems poised at criticality? J. Stat. Phys. 144 (2) (2011) 268–302.
[74] G. Duff, N. D, Differential Equations of Applied Mathematics, Wiley, 1966.
[75] G. Duff, Hyperbolic differential equations and waves, in: Boundary Value Problems for Linear Evolution Partial Differential Equations, Springer, 1977, pp. 27–155.
[76] C. Lanczos, Linear Differential Operators, Vol. 393, SIAM, 1961.