

PAPER • OPEN ACCESS

Correlation functions as a tool to study collective behaviour phenomena in biological systems

To cite this article: Tomás S Grigera 2021 *J. Phys. Complex.* **2** 045016

View the [article online](#) for updates and enhancements.

You may also like

- [Benchmarking embedded chain breaking in quantum annealing](#)
Erica Kelley Grant and Travis S Humble
- [Mapping global research on agricultural insurance](#)
Shalika Vyas, Tobias Dalhaus, Martin Kropff et al.
- [The variability of the narrow-line Seyfert 1 galaxies from the Pan-STARRS's view](#)
Hongtao Wang, Yanping Su, Xue Ge et al.

OPEN ACCESS

PAPER



Correlation functions as a tool to study collective behaviour phenomena in biological systems

RECEIVED
2 July 2021

REVISED
15 September 2021

ACCEPTED FOR PUBLICATION
28 September 2021

PUBLISHED
30 November 2021

Tomás S Grigera*

Instituto de Física de Líquidos y Sistemas Biológicos (IFLYSIB), CONICET y Universidad Nacional de La Plata, Calle 59 no. 789, B1900BTE La Plata, Argentina

CCT CONICET La Plata, Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina

Departamento de Física, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Argentina

* Author to whom any correspondence should be addressed.

E-mail: tgrigera@iflysib.unlp.edu.ar

Keywords: correlation functions, collective behaviour, criticality

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Abstract

Much of interesting complex biological behaviour arises from collective properties. Important information about collective behaviour lies in the time and space structure of fluctuations around average properties, and two-point correlation functions are a fundamental tool to study these fluctuations. We give a self-contained presentation of definitions and techniques for computation of correlation functions aimed at providing students and researchers outside the field of statistical physics a practical guide to calculating correlation functions from experimental and simulation data. We discuss some properties of correlations in critical systems, and the effect of finite system size, which is particularly relevant for most biological experimental systems. Finally we apply these to the case of the dynamical transition in a simple neuronal model.

1. Introduction

Biological systems behave in complex ways. This is particularly evident at the level of whole organisms: animals can adapt to a variety of situations, and this capacity is related to the large number of states (spatiotemporal patterns) that their brains can explore. These states are realised on some physical substrate (neurons in the case of the brain), but it is quite plausible that the same kind of behaviour can be produced by another system, composed of units that interact among themselves in a similar way, but microscopically of a very different nature. For example, robot birds could interact to form flocks that look macroscopically like starling murmurations, or a collection of neurons simulated in software, or built with electronics, could produce the same kind of patterns and behaviour as a brain (this has actually been attempted using simulated neurons wired with the connectome of the nematode *Caenorhabditis elegans* to drive a robot, see Urbina *et al* (2020)). The point is that many of the interesting phenomenology of biological systems stems from collective behaviour (Kaneko 2006), and that much of it may arise independently of the details of the biological units making up the system. A first description of collective properties starts of course with computation of average global quantities. But in complex systems there is much information on how fluctuations around those averages are structured in space and time. Correlation functions, the subject of this article, can capture information about these fluctuations, and thus constitute one of the basic statistical physics tools to describe collective properties. To be sure, there is more to collective behaviour than what is captured by the two-point correlation functions we discuss here. But correlation functions should be in the toolbox of any researcher attempting to understand the behaviour of complex biological systems, and it is probably fair to say that, although they have been used and studied for a long time in statistical physics, they have not been exploited enough in biology.

Among systems exhibiting nontrivial collective behaviour, critical systems deserve special mention. At a critical point (i.e. at the border between two static or dynamic phases) of a continuous transition, long-range correlated fluctuations dominate the large-scale phenomenology, so that many critical systems look alike at large scales: this is called universality. Criticality is rare, in the sense that systems are critical only in a negligible

region of the space of their control parameters. However, criticality appears to play an important role in biological systems (Bak 1996, Mora and Bialek 2011 and Muñoz 2018) and in particular the brain (Chialvo 2010), with different signs of critical behaviour having been found in systems as diverse as proteins (Mora *et al* 2010 and Tang *et al* 2017), membranes (Honerkamp-Smith *et al* 2009 and Veatch *et al* 2007), bacterial colonies (Dombrowski *et al* 2004 and Zhang *et al* 2010), social amoebas (De Palo *et al* 2017), networks of neurons (Beggs and Plenz 2003 and Schneidman *et al* 2006), brain activity (Fraiman and Chialvo 2012 and Haimovici *et al* 2013), midge swarms (Attanasi *et al* 2014 and Cavagna *et al* 2017), starling flocks (Cavagna *et al* 2010a) and sheep herds (Ginelli *et al* 2015). Since the odds that the system's parameters are tuned to critical just by chance are vanishing, the question arises of why critical systems appear to be so common in biology. An answer may be that criticality is the simple path to complexity (Chialvo 2018) and thus the reason why a functioning brain, for instance, needs to be 'at the boundary between being nearly dead and being fully epileptic' (Mora and Bialek 2011). Two-point correlation functions are an important tool in the exploration of critical systems. Again, they are not the whole story, but a necessary instrument nevertheless.

The purpose of this contribution is to present in a brief but self-contained form, the definitions of several two-point correlation functions together with a discussion on how to compute them from experimental or simulation data. We aim to give enough practical detail to enable students or researchers with no strong background in statistical physics to start to calculate correlation functions on their own data. At the end we consider a simple neural model, to illustrate how correlations can be studied on neuronal networks, and to argue that a simultaneous study of space and time correlations can yield useful information, and perhaps an alternative view, of their dynamic behaviour.

In the following, we present the theoretical definition of two-point correlation functions (section 2), followed by a discussion on their computation from experimental data (section 3). Then we discuss the definition and computation of the correlation length and time (section 4) and some properties of the correlation functions of near-critical systems (section 5). Finally we present some results on space and time correlations on a neuronal model (section 6).

2. Definitions

2.1. Correlation and covariance

Let x and y be two random variables and $p(x, y)$ their joint probability density. We use $\langle \dots \rangle$ to represent the appropriate averages, e.g. the mean of x is $\langle x \rangle = \int x p(x) dx$ (the probability distribution of x can be obtained from the joint probability, $p(x) = \int dy p(x, y)$, and in this context is called marginal probability) and its variance is $\text{Var}_x = \langle (x - \langle x \rangle)^2 \rangle$. We define

$$C_{xy} = \langle xy \rangle = \int xy p(x, y) dx dy, \quad (1)$$

$$\text{Cov}_{x,y} = \langle (x - \langle x \rangle) (y - \langle y \rangle) \rangle = \langle xy \rangle - \langle x \rangle \langle y \rangle. \quad (2)$$

We call C_{xy} the *correlation* and $\text{Cov}_{x,y}$ the *covariance* of x and y . The covariance is bounded by the product of the standard deviations (Priestley 1981, section 2.12),

$$\text{Cov}_{x,y}^2 \leq \text{Var}_x \text{Var}_y, \quad (3)$$

and is related to the variance of the sum:

$$\text{Var}_{x+y} = \text{Var}_x + \text{Var}_y + 2 \text{Cov}_{x,y}. \quad (4)$$

The *Pearson correlation coefficient* is defined as

$$r_{x,y} = \frac{\text{Cov}_{x,y}}{\sqrt{\text{Var}_x \text{Var}_y}}, \quad (5)$$

and, as a consequence of (3), is bounded: $-1 \leq r_{x,y} \leq 1$. It can be shown that the equality holds only when the relationship between x and y is linear (Priestley 1981, section 2.12).

The variables are said to be *uncorrelated* if their covariance is null:

$$\text{Cov}_{x,y} = 0 \iff \langle xy \rangle = \langle x \rangle \langle y \rangle \quad (\text{uncorrelated}). \quad (6)$$

Absence of correlation is weaker than *independence*: independence means that $p(x, y) = p(x)p(y)$ (and clearly implies absence of correlation). For uncorrelated variables it holds, because of (4), that the variance of the sum is the sum of the variances, but $\text{Cov}_{x,y} = 0$ is equivalent to independence only when the joint probability $p(x, y)$

is Gaussian. The covariance, or the correlation coefficient, are said to measure the degree of *linear association* between x and y , because it is possible to build a nonlinear dependence of x on y that yields zero covariance (see chapter 2 of Priestley 1981).

2.2. Fluctuating quantities as stochastic processes

Consider now an experimentally observable quantity, a , that provides some useful information on a property of a system of interest. We assume here for simplicity that a is a scalar, but the present considerations can be rather easily generalised to vector quantities. a can represent the magnetization of a material, the number of bacteria in a certain region, neural activity (e.g. as measured by fMRI), etc. We assume that a is a local quantity, i.e. that its value is defined at a particular position in space and time, so that we deal with a function $a(\mathbf{r}, t)$ (for the case where only space or time dependence is available see section 2.3.1).

We are interested in cases where a is *noisy*, i.e. subject to random fluctuations that arise because of our incomplete knowledge of the variables affecting the system's evolution, or because of our inability control them with enough precision (e.g. we do not know all variables that can affect the variation of the price of a stock market asset, we do not know all the interactions in a magnetic system, we cannot control all the microscopic degrees of freedom of a thermal bath). We wish to compare the values of a measured at different positions and times, but a statement like ' $a(\mathbf{r}_1, t_1)$ is larger than $a(\mathbf{r}_2, t_2)$ ' is useless in practice, because even if it is meaningful for a particular realization of the measurement process, the noisy character of the observable means that a different repetition of the experiment, under the same conditions, will yield a different function $a(\mathbf{r}, t)$. Actually repeating the experiment may or may not be feasible depending on the case, but we assume that we know enough about the system to be able to assert that a hypothetical repetition of the experiment would not exactly reproduce the original $a(\mathbf{r}, t)$. For clarity, it may be easier to imagine that several copies of the system are made and let evolve in parallel under the same conditions, each copy then producing a signal slightly different from that of the other copies. The quantity $a(\mathbf{r}, t)$ is then a random variable, and to compare it at different values of its argument we will use *correlation functions*, and this section is devoted to stating their precise definitions. (For a broader discussion of fluctuating variables in the natural sciences we can refer the reader to the book by Sornette (2006)).

The expression 'under the same conditions' deserves a comment. Clearly we expect that two strictly identical copies of a system evolving under exactly identical conditions will produce the same signal $a(\mathbf{r}, t)$. The 'same conditions' must be understood in a statistical way: the system is prepared by choosing a random initial configuration extracted from a well-defined probability distribution, or two identical copies evolve with a random dynamics with known statistical properties (e.g. coupled to a thermal bath at given temperature and pressure). We are excluding from consideration cases where the fluctuations are mainly due to experimental errors. If that were the only source of noise, one could in principle repeat the measurement enough times so that the average $\langle a(\mathbf{r}, t) \rangle$ is known with enough precision. Then $\langle a(\mathbf{r}, t) \rangle$ would be an accurate description of the system's actual spatial and temporal variations, and the correlations we are about to study would be dominated by properties of the measurement process rather than by the dynamics of the system itself. Instead we are interested in the opposite case: experimental error is negligible, and the fluctuations of the observable are due to some process intrinsic to the system. Indeed in many cases (such as macroscopic observables of systems in thermodynamic equilibrium) the average of the signal is uninteresting (it is a constant), but the correlation functions unveil interesting details of the system's spatial structure and dynamics.

From this discussion it is clear that $a(\mathbf{r}, t)$ is not an ordinary function. Rather, at each particular value of its arguments $a(\mathbf{r}, t)$ is a random variable: $a(\mathbf{r}, t)$ is therefore treated mathematically as a family of random variables indexed by \mathbf{r} and t , called *stochastic process* or *random field*. We do not need to delve into them here (but see appendix A for a brief introduction). To define correlation functions we only need to assume that the joint distributions

$$P_2(a_1, \mathbf{r}_1, t_1, a_2, \mathbf{r}_2, t_2) \quad (7)$$

are defined for all possible values of $\mathbf{r}_1, t_1, \mathbf{r}_2$, and t_2 . These distributions characterise only partially the stochastic process, but they are enough to define the correlation functions we consider here.

2.3. Definition of space-time correlation functions

The correlation function is the correlation of the random variables $a(\mathbf{r}_0, t_0)$ and $a(\mathbf{r}_0 + \mathbf{r}, t_0 + t)$,

$$C(\mathbf{r}_0, t_0, \mathbf{r}, t) = \langle a^*(\mathbf{r}_0, t_0) a(\mathbf{r}_0 + \mathbf{r}, t_0 + t) \rangle = \int da_1 da_2 P(a_1, \mathbf{r}_0, t_0, a_2, \mathbf{r}_0 + \mathbf{r}, t_0 + t) a_1^* a_2, \quad (8)$$

where the star stands for complex conjugate. The time difference t is sometimes called *lag*, and the function is called *self correlation* or *autocorrelation* to emphasize the fact that it is the correlation of the same observable quantity measured at different points. However that from the point of view of probability theory $a(\mathbf{r}_0, t_0)$ and

$a(\mathbf{r}_0 + \mathbf{r}, t_0 + t)$ are two *different* (though not independent) random variables. The *cross-correlations* of two observables is similarly defined:

$$C_{ab}(\mathbf{r}_0, t_0, \mathbf{r}, t) = \langle a^*(\mathbf{r}_0, t_0) b(\mathbf{r}_0 + \mathbf{r}, t_0 + t) \rangle. \quad (9)$$

The star again indicates complex conjugate. From now on however we shall restrict ourselves to real quantities and omit it in the following equations.

It is often useful to concentrate on the *fluctuation* $\delta a(\mathbf{r}, t) = a(\mathbf{r}, t) - \langle a(\mathbf{r}, t) \rangle$, especially when the average is independent of position and/or time. The correlation of the fluctuations is

$$C_c(\mathbf{r}_0, t_0, \mathbf{r}, t) = \langle \delta a(\mathbf{r}_0, t_0) \delta a(\mathbf{r}_0 + \mathbf{r}, t_0 + t) \rangle = C(\mathbf{r}_0, t_0, \mathbf{r}, t) - \langle a(\mathbf{r}_0, t_0) \rangle \langle a(\mathbf{r}_0 + \mathbf{r}, t_0 + t) \rangle, \quad (10)$$

and is called the *connected* correlation function (in diagrammatic perturbation theory, the connected correlation is obtained as the sum of connected Feynman diagrams only, see e.g. Binney *et al* 1992, chapter 8). From (6) it is clear that this function is zero when the variables are uncorrelated. For this reason it is often more useful than the correlation (8), which for uncorrelated variables takes the value $\langle a(\mathbf{r}_0, t_0) \rangle \langle a(\mathbf{r}_0 + \mathbf{r}, t_0 + t) \rangle$.

Sometimes a normalised connected correlation is defined so that its absolute value is bounded by 1, in analogy with the Pearson coefficient:

$$\rho(\mathbf{r}_0, t_0, \mathbf{r}, t) = \frac{C_c(\mathbf{r}_0, t_0, \mathbf{r}, t)}{\sqrt{C_c(\mathbf{r}_0, t_0, 0, 0) C_c(\mathbf{r}_0 + \mathbf{r}, t_0 + t, 0, 0)}}. \quad (11)$$

The names employed here are usual in the physics literature (e.g. Binney *et al* 1992 and Hansen and McDonald 2005). In the mathematical statistics literature, the connected correlation function is called *auto-covariance* (in fact it is just the covariance of $a(\mathbf{r}_0, t_0)$ and $a(\mathbf{r}_0 + \mathbf{r}, t_0 + t)$), while the name autocorrelation is reserved for the normalized connected correlation (11).

The correlations defined here are also known as *two-point* correlation functions, to distinguish them from higher-order correlations that can be studied but are outside the scope of this paper. Higher-order correlation functions are higher-order moments of $a(\mathbf{r}, t)$, while higher-order connected correlations correspond to the *cumulants* of $a(\mathbf{r}, t)$ (Itzykson and Drouffe 1989a, chapter 7) (the distinction between cumulants and moments is only relevant beyond third order, see e.g. van Kampen 2007, section 2.2).

2.3.1. Global and static quantities

The full spatial and temporal evolution may be not accessible or not relevant (e.g. because time evolution is very slow and the quantity is static at the experimental time-scales). In those cases one defines static space correlation functions or global time correlation functions which can be obtained as particular cases of (8) and (10).

The static space correlation function is just the space-time correlation evaluated at $t = 0$,

$$C(\mathbf{r}_0, \mathbf{r}) = C(\mathbf{r}_0, t_0, \mathbf{r}, 0) = \langle a(\mathbf{r}_0, t_0) a(\mathbf{r}_0 + \mathbf{r}, t_0) \rangle, \quad (12)$$

and analogously for the connected correlation. Time-only correlation functions are defined from a signal that can be regarded as a space integral of a local quantity,

$$A(t) = \int_V a(\mathbf{r}, t) d\mathbf{r}, \quad (13)$$

so that the time correlation function is

$$C(t, t_0) = \langle A(t_0) A(t_0 + t) \rangle = \left\langle \int d\mathbf{r}_0 a(\mathbf{r}_0, t_0) \int d\mathbf{r}_1 a(\mathbf{r}_1, t_0 + t) \right\rangle = \int d\mathbf{r}_0 d\mathbf{r} C(\mathbf{r}_0, t_0, \mathbf{r}, t). \quad (14)$$

2.3.2. Symmetries

Knowledge of the symmetries of the system under study, which are reflected in invariances of the stochastic process used to represent it, is very important. Apart from their significance in our theoretical understanding, at the practical level symmetries are useful in the estimation of statistical quantities, because they afford a way of obtaining additional sampling of the stochastic signal. For example, if the system is translation-invariant, one can perform measurements at different places of the same system and treat them as different samples of the same process, because one knows the statistical properties are independent of position (section 3.1).

If a system is time-translation invariant (TTI), then

$$\langle a(\mathbf{r}, t) \rangle = \langle a(\mathbf{r}, t + s) \rangle, \quad (15a)$$

$$\langle a(\mathbf{r}_1, t_1) a(\mathbf{r}_2, t_2) \rangle = \langle a(\mathbf{r}_1, t_1 + s) a(\mathbf{r}_2, t_2 + s) \rangle \quad (15b)$$

for all $\mathbf{r}_1, \mathbf{r}_2, t_1, t_2$ and s . In this case the system is said to be *stationary*.

In stationary systems the time origin is irrelevant: no matter *when* one performs an experiment, the statistical properties of the observed signal are always the same. In particular, this implies that the average $\langle a(\mathbf{r}, t) \rangle$ is independent of time (but does not mean that time correlations are trivial). Setting $s = -t_1$ one sees that the second moment depends only on the time difference $t_2 - t_1$, so that the time correlation depends only on one time (in our definition, the second time, or lag),

$$C(t) = \langle A(0)A(t) \rangle, \quad (\text{stationary}) \quad (16)$$

and differs from the connected correlation by a constant:

$$C_c(t) = \langle A(0)A(t) \rangle - \langle A \rangle^2 = C(t) - \langle A \rangle^2, \quad (\text{stationary}). \quad (17)$$

This is the situation that holds for a system in thermodynamic equilibrium.

Translation invariance in space similarly means that

$$\langle a(\mathbf{r}, t) \rangle = \langle a(\mathbf{r} + \mathbf{s}, t) \rangle, \quad (18a)$$

$$\langle a(\mathbf{r}_1, t_1)a(\mathbf{r}_2, t_2) \rangle = \langle a(\mathbf{r}_1 + \mathbf{s}, t_1)a(\mathbf{r}_2 + \mathbf{s}, t_2) \rangle \quad (18b)$$

for all positions and \mathbf{s} . In a translation-invariant, or *homogeneous* system, the space origin is irrelevant: no matter *where* the experiment is performed, the statistical properties are the same. In particular $\langle a(\mathbf{r}, t) \rangle$ is independent of \mathbf{r} , and the second moments depend only on the displacement $\mathbf{r}_2 - \mathbf{r}_1$. Space correlations then depend only on the second argument:

$$C(\mathbf{r}) = \langle a(0, t_0)a(\mathbf{r}, t_0) \rangle, \quad (\text{homogeneous}) \quad (19)$$

$$C_c(\mathbf{r}) = \langle \delta a(0, t_0)\delta a(\mathbf{r}, t_0) \rangle = \langle a(0, t_0)a(\mathbf{r}, t_0) \rangle - \langle a(0, t_0) \rangle^2 = C(\mathbf{r}) - \langle a(t_0) \rangle^2, \quad (\text{homogeneous}). \quad (20)$$

If in addition the system is *isotropic*, then the process will be *rotation invariant*, and the second moment depends only on the distance $r = |\mathbf{r}|$ between the two points, so that $C(\mathbf{r}) = C(r\mathbf{n})$ for any unit vector \mathbf{n} .

For stationary and homogeneous systems, the normalised correlation can be obtained simply by dividing by its value at the origin,

$$\rho(\mathbf{r}, t) = \frac{C_c(\mathbf{r}, t)}{C_c(0, 0)}, \quad (\text{stationary, homogeneous}). \quad (21)$$

2.3.3. Some properties

It is easy to see from the definitions that in the stationary and homogeneous case it holds that

$$C_c(0, 0) = \langle a(\mathbf{r}_0, t_0)^2 \rangle - \langle a(\mathbf{r}_0, t_0) \rangle^2 = \text{Var}_a, \quad (22)$$

$$|C_c(\mathbf{r}, t)| \leq C_c(0, 0) \quad \forall \mathbf{r}, t, \quad (23)$$

and that the correlation is even in \mathbf{r} and t if $a(\mathbf{r}, t)$ is real-valued.

Also, for sufficiently long lags and distances the variables will become uncorrelated, so that

$$C(\mathbf{r} \rightarrow \infty, t) \rightarrow \langle a \rangle^2, \quad C_c(\mathbf{r} \rightarrow \infty, t) \rightarrow 0, \quad (24)$$

$$C(\mathbf{r}, t \rightarrow \infty) \rightarrow \langle a \rangle^2, \quad C_c(\mathbf{r}, t \rightarrow \infty) \rightarrow 0. \quad (25)$$

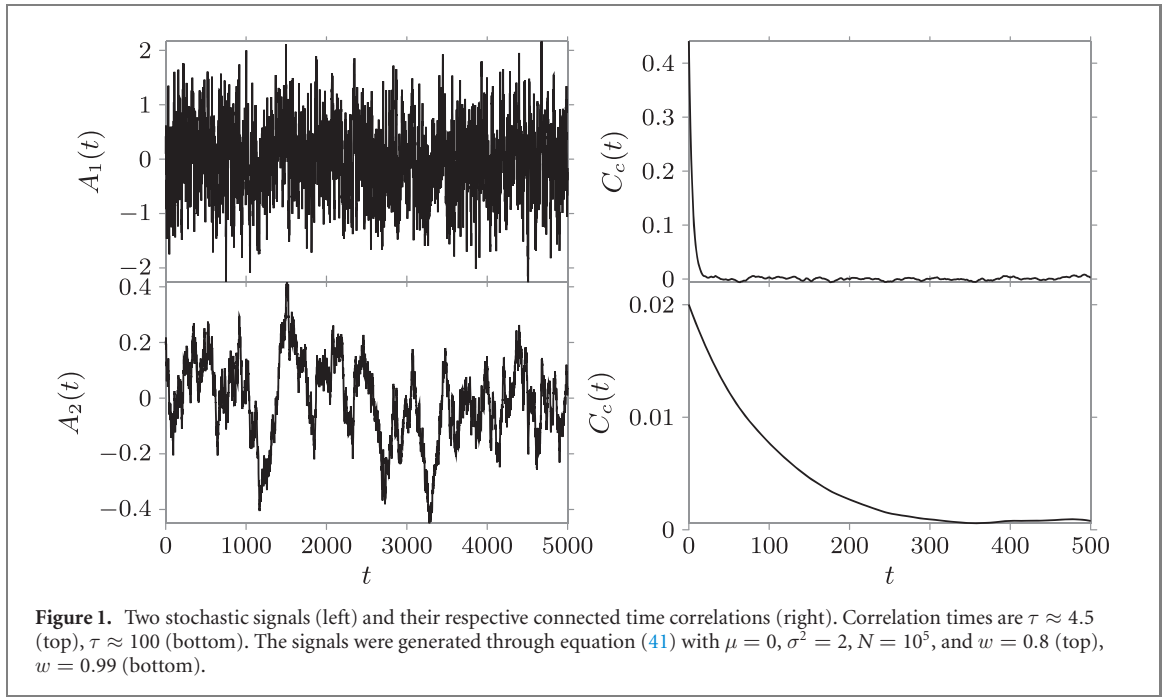
Thus the connected correlation will have a Fourier transform (section 2.4).

2.3.4. Example

Let us conclude this section of definitions with an example. Figure 1 shows on the left two synthetic stochastic signals, generated with the random process described in section 3.1.2. On the right there are the corresponding connected correlations. The two signals look different, and this difference is captured by $C_c(t)$. We can say that fluctuations for the lower signal are more persistent: when some value of the signal is reached, it tends to stay at similar values for longer, compared to the other signal. This is reflected in the slower decay of $C_c(t)$.

2.4. Correlations in Fourier space

Correlation functions are often studied in frequency space, either because they are obtained experimentally in the frequency domain (e.g. in techniques like dielectric spectroscopy), because the Fourier transforms are easier to compute or handle analytically, or because they provide an easier or alternative interpretation of the fluctuating process. Although the substance is the same, the precise definitions used can vary. One must



pay attention to (i) the convention used for the Fourier transform pairs and (ii) the exact variables that are transformed.

Here we define the Fourier transform pairs as

$$\tilde{f}(\mathbf{k}, \omega) = \int_{-\infty}^{\infty} dt \int d^d r f(\mathbf{r}, t) e^{-i\mathbf{k}\cdot\mathbf{r} + i\omega t}, \quad f(\mathbf{r}, t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \int \frac{d^d r}{(2\pi)^d} \tilde{f}(\mathbf{k}, \omega) e^{i\mathbf{k}\cdot\mathbf{r} - i\omega t}, \quad (26)$$

but some authors choose the opposite sign for the forward transform and/or a different placement for the $1/2\pi$ factor (sometimes splitting it between the forward and inverse transforms). Depending on the convention a factor 2π can appear or disappear in some relations.

Let us consider time-only correlations first. These are defined in terms of two times, t_1 and t_2 (which we chose to write as t_0 and $t_0 + t$). One can transform one or both of t_1 and t_2 or t_0 and t . Let us mention two different choices, useful in different circumstances. First, take the connected time correlation (10) and do a Fourier transform on t :

$$C_c(t_0, \omega) = \int dt e^{i\omega t} C_c(t_0, t_0 + t) = e^{-i\omega t_0} \langle \delta A(t_0) \delta \tilde{A}(\omega) \rangle, \quad (27)$$

where $\delta \tilde{A}(\omega)$ stands for the Fourier transform of $\delta A(t)$. This definition is convenient when there is an explicit dependence on t_0 but the evolution with t_0 is slow, as in the physical ageing of glassy systems (see e.g. Cugliandolo 2004); one then studies a time-dependent spectrum. If on the other hand C_c is stationary, it is more convenient to write $t_1 = t_0$, $t_2 = t_0 + t$ and do a double transform in t_1, t_2 :

$$\begin{aligned} \tilde{C}_c(\omega_1, \omega_2) &= \int dt_1 dt_2 e^{i\omega_1 t_1} e^{i\omega_2 t_2} C_c(t_1, t_2 - t_1) \\ &= \int dt_1 dt_2 e^{i\omega_1 t_1} e^{i\omega_2 t_2} \langle \delta A(t_1) \delta A(t_2) \rangle = \langle \delta \tilde{A}(\omega_1) \delta \tilde{A}(\omega_2) \rangle \\ &= \int dt_1 dt_2 e^{i(\omega_1 + \omega_2)t_1} e^{i\omega_2(t_2 - t_1)} C_c(t_2 - t_1) = (2\pi) \delta(\omega_1 + \omega_2) \tilde{C}_c(\omega_2), \end{aligned} \quad (28)$$

where $\tilde{C}_c(\omega)$ is the Fourier transform of the stationary connected correlation with respect to t and we have used the integral representation of Dirac's delta, $\delta(\omega - \omega') = (2\pi)^{-1} \int_{-\infty}^{\infty} e^{it(\omega - \omega')} dt$. As (28) shows, the transform is zero unless $\omega_1 = -\omega_2$, so that it is useful to define the *reduced* correlation,

$$C_c^R(\omega) = \langle \delta \tilde{A}(-\omega) \delta \tilde{A}(\omega) \rangle = \langle \delta \tilde{A}^*(\omega) \delta \tilde{A}(\omega) \rangle, \quad (29)$$

where the rightmost equality holds when $A(t)$ is real.

The transform $\tilde{C}_c(\omega)$ is a well-defined function, because the connected correlation decays to zero (and usually fast enough). We can then say

$$\tilde{C}_c(\omega) = \frac{1}{2\pi\delta(0)} C_c^R(\omega). \quad (30)$$

This relation can be exploited to build a very fast algorithm to compute $C_c(t)$ in practice (appendix D). The at first sight baffling infinite factor relating the reduced correlation to $\tilde{C}_c(\omega)$ originates in the fact that $\delta\tilde{A}(\omega)$ cannot exist as an ordinary function, since we have assumed that $A(t)$ is stationary. This implies that the signal has an infinite duration, and that, its statistical properties being always the same, it cannot decay to zero for long times as required for its Fourier transform to exist. The Dirac delta can be understood as originating from a limiting process where one considers signals of duration T (with suitable, usually periodic, boundary conditions) and then takes $T \rightarrow \infty$. Then $2\pi\delta(\omega = 0) = \int dt e^{i\omega t}|_{\omega=0} = \int dt = T$. These considerations can be made rigorous by defining the signal's *power spectrum*,

$$h(\omega) = \lim_{T \rightarrow \infty} \frac{1}{T} \langle A_T(\omega) A_T^*(\omega) \rangle, \quad A_T(\omega) = \int_{-T/2}^{T/2} A(t) e^{i\omega t} dt, \quad (31)$$

and then proving that $h(\omega) = \tilde{C}_c(\omega)$ (Priestley 1981, sections 4.7 and 4.8).

The same considerations apply to space or space-time correlations. For reference, the relations one finds are

$$\tilde{C}_c(\mathbf{k}_1, \omega_1, \mathbf{k}_2, \omega_2) = \int d^d r_1 dt_1 \int d^d r_2 dt_2 e^{-i\mathbf{k}_1 \cdot \mathbf{r}_1 + i\omega_1 t_1} e^{-i\mathbf{k}_2 \cdot \mathbf{r}_2 + i\omega_2 t_2} C_c(\mathbf{r}_1, t_1, \mathbf{r}_2 - \mathbf{r}_1, t_2 - t_1), \quad (32a)$$

$$\tilde{C}_c(\mathbf{k}, \omega) = \int_{-\infty}^{\infty} dt \int d^d r e^{-i\mathbf{k} \cdot \mathbf{r} + i\omega t} C_c(\mathbf{r}, t), \quad (32b)$$

$$C_c^R(\mathbf{k}, \omega) = \langle \delta a(-\mathbf{k}, -\omega) \delta a(\mathbf{k}, \omega) \rangle, \quad (32c)$$

$$\tilde{C}_c(\mathbf{k}_1, \omega_1, \mathbf{k}_2, \omega_2) = (2\pi)^{d+1} \delta^{(d)}(\mathbf{k}_1 + \mathbf{k}_2) \delta(\omega_1 + \omega_2) C_c^R(\mathbf{k}_1, \omega_1), \quad (32d)$$

$$\tilde{C}_c(\mathbf{k}, \omega) = \frac{1}{(2\pi)^{d+1} \delta^{(d)}(0) \delta(0)} C_c^R(\mathbf{k}, \omega) = \frac{1}{(2\pi)^{d+1} VT} C_c^R(\mathbf{k}, \omega). \quad (32e)$$

3. Computing correlation functions from experimental data

The theoretical definitions of correlation functions are given as averages over some probability distribution, or ensemble. To compute them in practice, given data recorded in an experiment or produced in numerical simulation, there are two issues to consider. First, experimental data will be discrete in time as well as in space, either as a result of sampling, or because the spatial resolution is high enough to measure the actual discrete units that make up our system ('particles', i.e. birds, cells, neurons, etc.). Second, we do not have direct access to the probability distribution, but only to a set of samples, i.e. results from experiments, distributed randomly according to an unknown distribution. Thus what we actually compute are *estimators* (as they are called in statistics) of the averages we want (the correlation functions). We will discuss estimators corresponding to different situations, and for clarity it will be convenient to consider first the case of a global quantity (correlation in time only, section 3.1) and only later introduce the additional complications brought in by the space structure (section 3.2).

Before moving to the estimation of correlations, let us examine the estimator for the mean, which we will need to estimate the connected correlation functions, and which will allow us to make a couple of general points. It is clearly hopeless to attempt to estimate an ensemble average unless it is possible to obtain many samples under the same conditions (i.e., if one is throwing dice, one should throw many times the *same* dice). So we assume that we are given a set of M samples $\{a^{(n)}\}$, $n = 1, \dots, M$ obtained in M equivalent experiments. From these samples we can estimate the mean as

$$\bar{a} = \frac{1}{M} \sum_{n=1}^M a^{(n)}. \quad (33)$$

This well-known estimator has the desirable property that it 'gets better and better' when the number of samples M grows, if the ensemble has finite variance. More precisely, $\bar{a} \rightarrow \langle a \rangle$ as $M \rightarrow \infty$. This property is called *consistency*, and is guaranteed because $\langle \bar{a} \rangle = \langle a \rangle$ (i.e. the estimator is *unbiased*) and the variance of \bar{a} vanishes for $M \rightarrow \infty$ (Priestley 1981, sections 5.2 and 5.3).

How far off the estimate can be expected to be from the actual value (i.e. the ‘error’), is proportional to its variance¹. A result called the Cramer–Rao inequality implies that there is a lower bound to the variance of any estimator, given the number of samples (Priestley 1981, section 5.2). So in practice one chooses a good known estimator such as (33), but the only way to improve on the error is to acquire more samples, since the variance of (33) goes as $\sim 1/\sqrt{M}$. Unfortunately $1/\sqrt{M}$ is not a very fast-decreasing function (a tenfold increase in the number of samples reduces the error only by about a third), so one is always pressed to obtain the largest possible number of samples. But in experiments (especially in biology) the number of samples may be scarce, and more so in the case of correlations, which are defined in terms of pairs of values, separated by a fixed distance or time.

To make the most out of the available samples, one must take advantage of the known symmetries of the system. Space translation invariance allows us to treat samples measured at different points of a large sample as different equivalent experiments, and the same goes for samples obtained at different points in time (even if the experiment has not been restarted) if the system is stationary. The samples obtained this way are in general *not* independent (they will be in fact correlated), but the estimate (33) does not need to be used on independent samples. It is still unbiased and consistent, only that its variance will be larger than it would be for the same number of independent samples (see equation (38)).

Accordingly, whenever in what follows we use n to index samples or experiments, it is understood that the averages over n can actually be implemented as averages over a time or space index if the system is stationary or homogeneous, respectively. We refer then to *time averages* or *space averages* as replacing ensemble averages when building an estimator.

3.1. Estimation of time correlation functions

Assume that the experiment records a scalar signal with a uniform sampling interval Δt , so that we are handled N real-valued and time-ordered values forming a sequence A_i , with $i = 1, \dots, N$. It is understood that if the data are digitally sampled from a continuous time process, proper filtering has been applied². In what follows we shall measure time in units of the sampling interval, so that in the formulae below we shall make no use of Δt . To recover the original time units one must simply remember that $A_i = A(t_i)$ with $t_i = t_0 + (i - 1)\Delta t$. For the stationary case we shall write $C_k = C(t_k)$ where t_k is the time difference, $t_k = k\Delta t$ and $k = 0, \dots, N - 1$, and in the non-stationary case $C_{i,k} = C(t_i, t_k)$.

If the process is not stationary, nothing can be estimated from a single sequence because the successive values correspond to different, nonequivalent experiments, as the conditions have changed from one sample to another (i.e. the system has evolved in a way that has altered the characteristic of the stochastic process). In this case the correlation depends on two times and the mean itself can depend on time. The only way to estimate the mean or the correlation function is to obtain many sequences $A_i^{(n)}$, $n = 1, \dots, M$ reinitializing the system to the same macroscopic conditions each time (in a simulation, one can for example restart the simulation with the same parameters but changing the random number seed). It is not possible to resort to time averaging, and the estimates are obtained by replacing the ensemble average by averages across the different sequences, i.e. the (time-dependent) mean is estimated as

$$\langle A(t_i) \rangle \approx \bar{A}_i = \frac{1}{M} \sum_{n=1}^M A_i^{(n)}, \quad (34)$$

and the time correlation as

$$C_c(t_i, t_k) \approx \hat{C}_{i,k}^{(c)} = \frac{1}{M} \sum_n \delta A_i^{(n)} \delta A_{i+k}^{(n)}, \quad \delta A_i^{(n)} = A_i^{(n)} - \bar{A}_i, \quad (35)$$

where the hat distinguishes the estimate from the actual quantity. Both estimators are consistent and unbiased, i.e. $\hat{C}_{i,k}^{(c)} \rightarrow C(t_i, t_i)$ and $\overline{A(t_i)} \rightarrow \langle A(t_i) \rangle$ for $M \rightarrow \infty$.

If the system is stationary, one can resort to time averaging to build estimators (like (37) below) that only require one sequence. But let us remark that it is always sensible to check whether the sequence is actually stationary. A first check on the mean can be done dividing the sequence in several sections and computing

¹ Being a function of random variables, the estimator has a probability distribution of its own, and in particular mean and variance.

² According to the Nyquist–Shannon sampling theorem, if the signal has frequency components higher than half the sampling frequency (i.e. if the Fourier transform is nonzero for $\omega \geq \pi/\Delta t$) then the signal cannot be accurately reconstructed from the discrete samples; in particular the high frequencies will ‘pollute’ the low frequencies (an effect called aliasing). Thus the signal should be passed through an analog low-pass filter before sampling. See Press et al (1992) (section 12.1) for a concise self-contained discussion, or Priestley (1981) (section 7.1).

the average of each section, then looking for a possible systematic trend. If this check passes, then one should compute the time correlation of each section and then check that all of them show the same decay (using fewer sections than in the first check, as longer sections will be needed to obtain meaningful estimates of the correlation function). It is important to note that this second check is necessary even if the first one passes; as it is possible for the mean to be time-independent (or its time-dependence undetectable) while the time correlation still depends on two times.

In the stationary case, then, we can estimate the mean with

$$\bar{A} = \frac{1}{N} \sum_{i=1}^N A_i, \quad (36)$$

and the connected correlation function with

$$\hat{C}_k^{(c)} = \frac{1}{N-k} \sum_{j=1}^{N-k} \delta A_j \delta A_{j+k}, \quad \delta A_j = A_j - \bar{A}. \quad (37)$$

In the last equation, the true sample mean $\langle A \rangle$, if known, can be used instead of the estimate \bar{A} . If the true mean is used, the estimator (37) is unbiased, otherwise it is *asymptotically* unbiased, i.e. the bias tends to zero for $N \rightarrow \infty$, provided the Fourier transform of $C_c(t)$ exists. More precisely, $\langle \hat{C}_k^{(c)} \rangle = C_c(t_k) - \alpha/N$, where $\alpha = 2\pi \text{Var}_A \tilde{C}_c(\omega = 0)$. The variance of $\hat{C}_k^{(c)}$ is $O(\frac{1}{N-k})$ (Priestley 1981, section 5.3). This is sufficient to show that, at fixed k , the estimator is consistent, i.e. $\hat{C}_k^{(c)} \rightarrow C_c(t_k)$ for $N \rightarrow \infty$. However, the variance grows with increasing k , and thus the tail of $\hat{C}_k^{(c)}$ is considerably noisy. In practice, for k near N the estimate is mostly useless, and the usual rule of thumb is to use $\hat{C}_k^{(c)}$ only for $k \leq N/2$.

Knowing the time correlation, one can make the statement that the variance of the estimate (36) is larger than that of (34) quantitative: the variance of the estimate with correlated samples is (Sokal 1997)

$$\text{Var}_{\bar{a}} \approx \frac{2\tau_{\text{int}}}{N} \left[\langle A^2 \rangle - \langle A \rangle^2 \right], \quad (38)$$

i.e. $2\tau_{\text{int}}$ times larger than the variance in the independent sample case, where τ_{int} is the integral correlation time defined below (58). In this sense $N/2\tau_{\text{int}}$ can be thought of as the number of ‘effectively independent’ samples (see also Sornette 2006, chapter 8).

We have not written an estimator for the nonconnected correlation function. It is natural to propose

$$\hat{C}_k = \frac{1}{N-k} \sum_{j=1}^{N-k} A_j A_{j+k}. \quad (39)$$

However, although \hat{C}_k is unbiased, it can have a large variance when the signal has a mean value larger than the typical fluctuation (see section 3.1.1). As a general rule, it is not a good idea estimate the connected correlation by using (39) and then subtracting \bar{A}^2 . Instead, the connected estimator (37) should be used. A possible exception may be the case when an experiment can only furnish many short independent samples of the signal (see the examples in section 3.1.2).

If several sequences sampled from a stationary system are available, it is possible to combine the two averages: one *first* computes the stationary correlation estimate (37) for each sequence, and *then* averages the different correlation estimates (over n at fixed lag k). It is clearly wrong to average the sequences themselves before computing the correlation, as this will tend to approach the (stationary) ensemble mean $\langle A \rangle$ for all times, destroying the dynamical correlations.

There is another asymptotically unbiased estimator of the connected correlation, that can be obtained by using $1/N$ instead of $1/(N-i)$ as prefactor of the sum in (37). Calling this estimator $C_{c,k}^*$, it holds that $\langle C_{c,k}^* \rangle = C_c(t_k) - \alpha/N - kC(t_k)/N - \alpha k/N^2$, where α is defined as before, and again $\alpha = 0$ if the exact sample mean is used (Priestley 1981, section 5.3). This has the unpleasant property that the bias depends on k , while the bias of $\hat{C}_{c,k}$ is independent of its argument, and smaller in magnitude. The advantage of $C_{c,k}^*$ is that its variance is $O(1/N)$ independent of k , thus it has a less noisy tail. Some authors prefer $C_{c,k}^*$ due to its smaller variance and to the fact that it strictly preserves properties of the correlation, which may not hold exactly for $\hat{C}_{c,k}$. Here we stick to $\hat{C}_{c,k}$, as usual in the physics literature (e.g. Allen and Tildesley 1987, Newman and Barkema 1999 and Sokal 1997), so that we avoid worrying about possible distortions of the shape. In practice however, it seems that as long as N is greater than $\sim 10\tau$ (a necessary requirement in any case, see below), and for $k \leq N/2$, there is little numerical difference between the two estimates.

3.1.1. Two properties of the estimator

We must mention two important features of the estimator that are highly relevant when attempting to compute numerically the time correlation. The first is that the difference between the non-connected and connected estimators is not \bar{a}^2 , but

$$\hat{C}_k - \hat{C}_{c,k} = \bar{a} \left[\frac{1}{N-k} \sum_j^{N-k} a_j + \frac{1}{N-k} \sum_j^{N-k} a_{j+k} - \bar{a} \right], \quad (40)$$

as it is not difficult to compute. The difference does tend to \bar{a}^2 for $N \rightarrow \infty$, but in a finite sample fluctuations are important, especially at large k . Since fluctuations are amplified by a factor \bar{a} , when the signal mean is large with respect to its variance, the estimate \hat{C}_k is completely washed out by statistical fluctuations. This is why, while $C(t)$ and $C_c(t)$ differ by a constant, in practice it is a bad idea to compute \hat{C}_k and subtract \bar{a}^2 to obtain an estimate of the connected correlation. Instead, *one computes the connected correlation directly* by estimating first the mean and then using (37).

Another important fact is that the estimator of the connected correlation *will always have zero*, whatever the length of the sample used, *even if N is much shorter than the correlation time*. As shown in appendix B, for any time series one has that $\hat{C}_0^{(c)} > 0$ and that $\hat{C}_{c,k}^{(c)}$ will *change sign* at least once for $k \geq 1$ (this applies to the case when the mean and connected correlation are estimated with a single time series). The practical consequence is that when N is of the order of τ , or smaller, the estimate $\hat{C}_k^{(c)}$ will suffer from strong finite-size effects, and its shape will be quite different from the actual $C_c(t)$. In particular, since $\hat{C}_k^{(c)}$ will intersect the x -axis, it will look like the series has decorrelated when in fact it is still highly correlated. Be suspicious if $\hat{C}_k^{(c)}$ changes sign once and stays very anticorrelated. Anticorrelation may be a feature of the actual $C_c(t)$, but if the sample is long enough, the estimate should decay until correlation is lost (noisily oscillating around zero). One must always perform tests estimating the correlation with different values of N : if the shape of the correlation at short times depends on N , then N must be increased until one finds that estimates for different values of N coincide for lags up to a few times the correlation time (we show an example next). Once a sample-size-independent estimate has been obtained, the correlation time can be estimated (section 4.1), and it must be checked that self-consistently N is several times larger than τ .

3.1.2. Example

To illustrate the above considerations on estimating $C_c(t)$ we use a synthetic correlated sequence generated from the recursion

$$a_i = wa_{i-1} + (1-w)\xi_i, \quad (41)$$

where $w \in [0, 1]$ is a parameter and the ξ_i are uncorrelated Gaussian random variables with mean μ and variance σ^2 . Assuming the stationary state has been reached it is not difficult to find

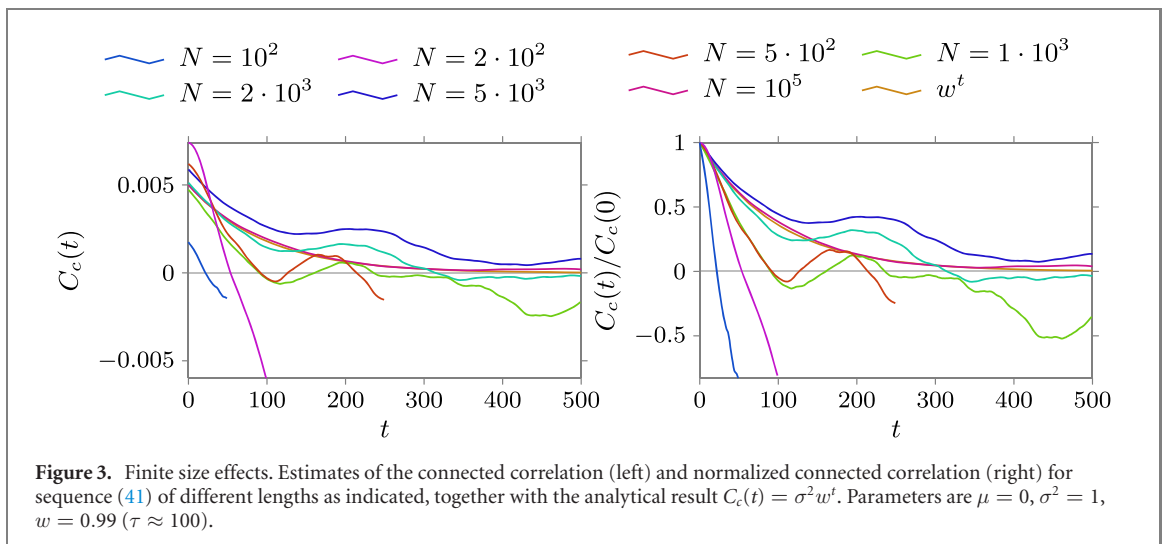
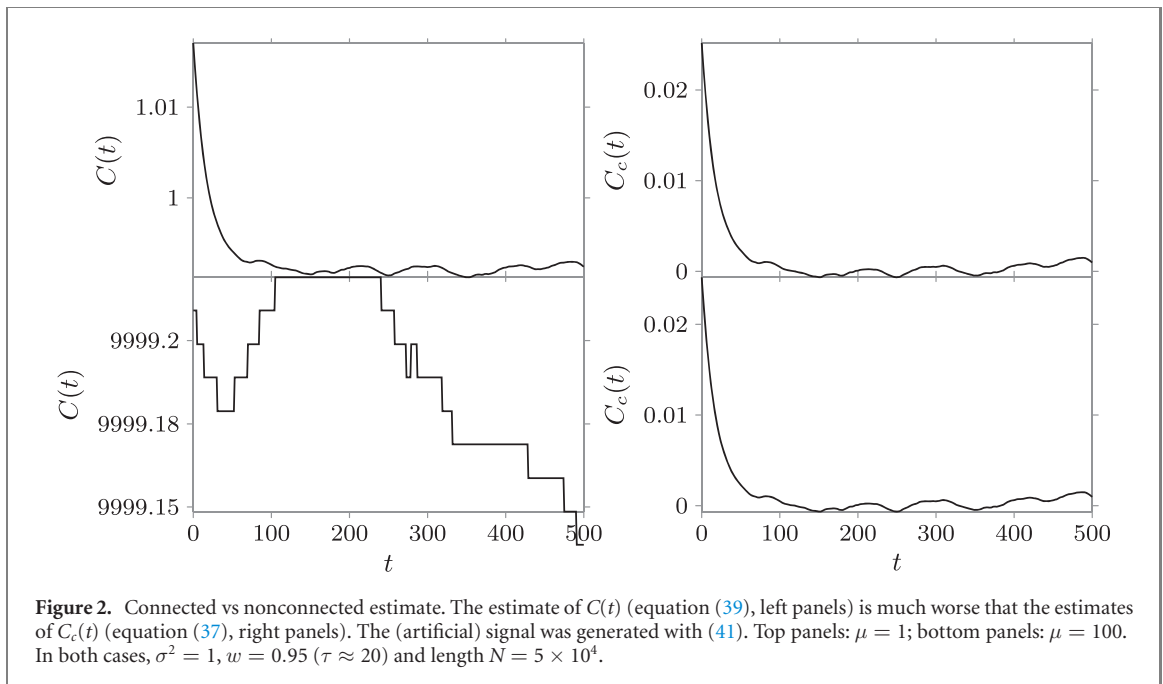
$$\langle a \rangle = \mu, \quad \langle (a - \mu)^2 \rangle = \frac{1-w}{1+w} \sigma^2, \quad C_k^{(c)} = \langle (a_i - \mu)(a_{i+k} - \mu) \rangle = \sigma^2 w^k, \quad (42)$$

so that the correlation time is $\tau = -1/\log w$.

We used the above recursion to generate artificial sequences and computed their time correlation functions with the estimates discussed above. Figure 2 shows the problem that can face the non-connected estimate. When the average of the signal is smaller than or of the order of the noise amplitude (as in the top panels), one can get away with using (39). However if $\mu \gg \sigma$, the non-connected estimate is swamped by noise, while the connected estimate is essentially unaffected (bottom panels). Hence, if one is considering only one sequence, one should always use the connected estimator.

In figure 3 we see how using samples that are too short affects the correlation estimates. The same artificial signal was generated with different lengths. For the shorter lengths, it is seen that the correlation estimate crosses the t axis (as we have shown it must) but does not show signs of losing correlation. One might hope that $N = 1000 \approx 10\tau$ is enough (the estimate starts to show a flat tail), but comparing to the result of doubling the length shows that it is still suffering from finite-length effects. For this particular sequence, it is seen that at least 20τ to 50τ is needed to get the initial part of the normalized connected correlation more or less right, while a length of about 1000τ is necessary to obtain a good estimate up to $t \sim 5\tau$. The unnormalized estimator suffers still more from finite size due to the increased error in the determination of the variance (left panel).

If the experimental situation is such that it is impossible to obtain sequences much longer than the correlation time, one can get some information on the time correlation if it is possible to repeat the experiment so as to obtain several *independent* and *statistically equivalent* sample sequences. In figure 4 we take several sequences with the same parameters as in the previous example, but quite short ($N = 200 \approx 2\tau$). As we have shown, it is not possible to obtain a moderately reasonable estimate of $C_c(t)$ using one such sequence (as is also clear from

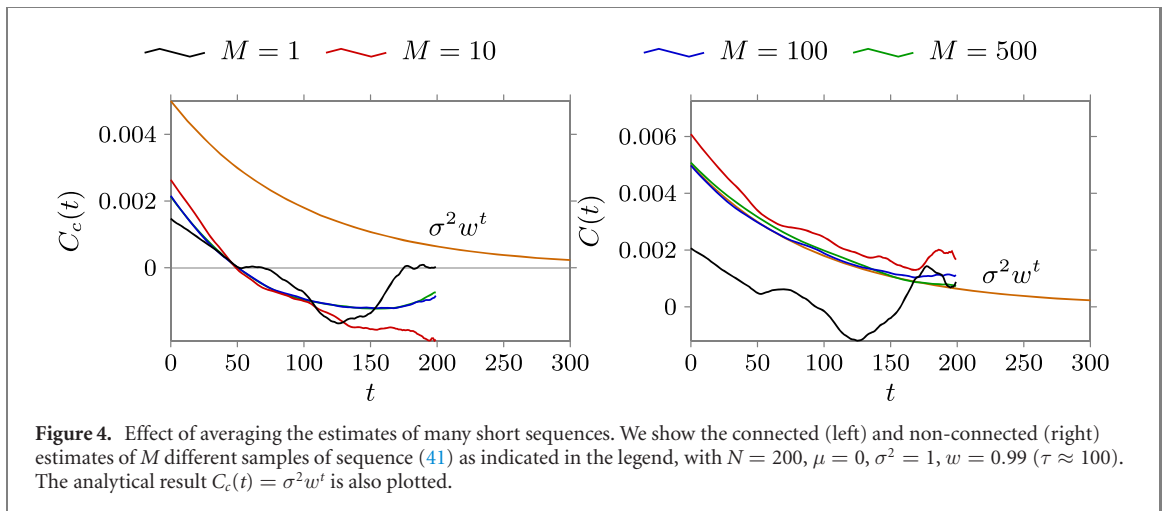


the $M = 1$ case of figure 4). However, the figure shows how one may benefit from the multiple repetitions of the (short) experiment by averaging together all the estimates. The averaged connected estimates are always far from the actual correlation function, even for $M = 500$ (the case which contains in total 10^5 points, which proved quite satisfactory in the previous example): this is consequence of the fact that all connected estimates must become negative. Instead, the averaged non-connected estimates approach quite closely the analytical result even though not reaching the decorrelated region³. Although it is tricky to try to extract a correlation time from this estimate (due to lack of information on the last part of the decay), this procedure at least offers a way to obtain some dynamical information in the face of experimental limitations.

3.2. Estimation of space correlation functions

To compute space correlations one needs experimental data with spatial resolution. We can assume that each experiment provides us a set of values $\{a_i^{(n)}\}$ together with a set of positions $\{\mathbf{r}_i^{(n)}\}$ which specify where in the sample the i th value has been recorded. The experiment label, $n = 1, \dots, M$ can indicate experiments on different samples, or for a stationary system, it can be a time index.

³ Note that in this case $\mu = 0$ so that fluctuations are larger than the average. If the average were very large, one may attempt to compute a connected correlation estimate by using all sequences to estimate the average, then subtracting this same average to all sequences.



The positions $\mathbf{r}_i^{(n)}$ may be fixed by the experiment (e.g. when sampling in space with an array of sensors, in which case they are likely independent of n) or might be themselves part of the experimental result (as when measuring the positions of moving organisms). In the former case, a naive implementation of definition (12) will pick up the structure of the experimental array itself, introducing correlations that are clearly an artefact. In the latter case, the positions may encode nontrivial structural features, which will show up in the correlation function of a together with a -specific effects. For example, particles with excluded volume will have non-trivial density correlations at short range, while there might be long-range correlations due to an ordering of a . It is then wise to untangle the two effects as much as possible. Thus in any case it is convenient to implement the definition in a way that separates structural correlations.

Structural effects can be studied through density correlation functions. One such function is the *pair distribution function* (Hansen and McDonald 2005, section 2.5), defined for homogeneous systems as

$$g(\mathbf{r}) = \frac{1}{\rho_0 N} \left\langle \sum_{i \neq j} \delta^{(d)}(\mathbf{r} - \mathbf{r}_{ij}) \right\rangle = \frac{1}{N \rho_0} \left\langle \sum_{i \neq j} \frac{\delta(|\mathbf{r}| - |\mathbf{r}_{ij}|)}{4\pi |\mathbf{r}|^2} \right\rangle, \quad (43)$$

where $\delta^{(d)}(\mathbf{r})$ is the d -dimensional Dirac's delta, $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ is the pair displacement and $\rho_0 = N/V$ is the average density. The second equality is valid for isotropic systems, and $g(r)$ is often called *radial distribution function* in this case. Defining the density of the configuration $\{\mathbf{r}_i\}$ by $\rho(\mathbf{r}) = \sum_i \delta(\mathbf{r} - \mathbf{r}_i)$, the pair distribution function is found to be related to the density-density correlation function by

$$\langle \rho(\mathbf{r}_0) \rho(\mathbf{r}_0 + \mathbf{r}) \rangle = \rho_0^2 g(\mathbf{r}) + \rho_0 \delta(\mathbf{r}) = \rho_0^2 g(|\mathbf{r}|) + \rho_0 \frac{\delta(|\mathbf{r}|)}{4\pi |\mathbf{r}|^2}, \quad (44)$$

where again the second equality applies to isotropic systems. The Dirac delta term arises because in $\langle \rho(\mathbf{r}_0) \rho(\mathbf{r}_0 + \mathbf{r}) \rangle$ one includes the $i = j$ terms that were excluded from the sum in (43). If the $\{\mathbf{r}_i\}$ are obtained with periodic boundary conditions (as is often the case in simulation), (43) can be used to estimate $g(r)$ almost directly: it suffices to replace the Dirac delta by a discrete version (i.e. turn it into a binning function) and the average by a (normalised) sum over all available statistically equivalent configurations. For the isotropic case,

$$\hat{g}_k = \frac{1}{M} \sum_n \frac{1}{\rho_0 N} \sum_{ij} \frac{\Delta [r_{ij}^{(n)} - (k + 1/2) \Delta r]}{V_k}, \quad (45a)$$

$$\Delta(r) = \begin{cases} 1 & \text{if } -\Delta r/2 < r \leq \Delta r/2, \\ 0 & \text{otherwise,} \end{cases} \quad (45b)$$

where Δr is the bin width and V_k is the volume of the k th bin, in three dimensions $V_k = \frac{4}{3} \pi (\Delta r)^3 ((k + 1)^3 - k^3)$. The bin width must not be chosen too small or some bins will end up having very few or no particles. In any case, it must be larger than the experimental resolution, otherwise $g(r)$ will pick up spurious correlations arising from the fact that due to finite spatial resolution, all positions effectively lay on the sites of a square lattice.

If borders are non-periodic, as in experimental data, using (49) will introduce bias at large r , because the number of neighbours of a given particle stops growing as r^2 (the volume of the bin) as soon as r is of the order of the distance of the particle to the nearest border. In this case an unbiased estimator is

$$\hat{g}_k = \frac{1}{\rho_0 N_k} \sum_{i \in S_k} \sum_j \frac{\Delta(k - r_{ij}/\Delta r)}{V_k}, \tag{46}$$

where $i \in S_k$ means that the sum runs only over the set of particles that are at least at a distance $(k + 1/2)\Delta r$ from the nearest border, and N_k is the number of such particles. In other words, for a given point i , the pair ij only enters the sum if \mathbf{r}_j lies on a bin (centred at i) that is completely contained within the system's volume. This is known as the unweighted Hanisch method (Hanisch 1984).

We stress that in systems with non-periodic boundaries it is essential to deal adequately with borders. Border effects are not limited to correlation functions, but also affect other observables, such as the average density. The Hanisch method (46) is an unbiased estimator for the radial distribution function, but to apply it one first needs to define the borders, which are to a certain degree arbitrary in the case of a discrete system. The *convex hull* algorithm can define an unambiguous border by finding the smallest convex set that includes all the system. However, when obvious concavities or 'holes' are present, the convex hull can greatly overestimate the volume (leading to underestimating the average density). To deal with concavities one can use the α -shape algorithm (Edelsbrunner and Mücke 1994), at the price of introducing an arbitrary parameter (see Cavagna *et al* 2008, for discussion on using these algorithms to determine the borders of biological groups). Details on the correct determination of borders are also discussed by Villegas *et al* (2021), who study the spatial structure of a rain forest, and show how valuable insight can be obtained from the combined study of structural correlations together with the spatial-scale dependence of density fluctuations.

To conclude our remarks on structural correlation functions, let us note that if correlations in Fourier space are of interest, it is typically better to estimate them directly, rather than going through the real space functions and applying a numerical Fourier transform. For example, the *structure factor*, which is proportional to the reduced density correlation,

$$S(\mathbf{k}) = \frac{1}{N} \langle \rho(\mathbf{k})\rho(-\mathbf{k}) \rangle = 1 + \rho_0 \tilde{g}(\mathbf{k}), \tag{47}$$

where $\tilde{g}(\mathbf{k})$ is the Fourier transform of $g(\mathbf{r})$, can be computed directly from the $\mathbf{r}_i^{(n)}$ through

$$S(\mathbf{k}) = \frac{1}{M} \frac{1}{N} \sum_n \sum_{j,l} e^{i\mathbf{k}\cdot\mathbf{r}_{jl}^{(n)}} = 1 + \frac{1}{NM} \sum_n \sum_{i \neq j} \frac{\sin |\mathbf{k}| r_{ij}^{(n)}}{|\mathbf{k}| r_{ij}^{(n)}}, \tag{48}$$

where the second equality applies to isotropic system and is obtained by analytically averaging the first expression over all angles. For further discussion of density correlation functions we refer the reader to condensed matter texts such as Hansen and McDonald (2005) or Ashcroft and Mermin (1976).

Turning now to the correlations of a , and focusing on the homogeneous case, we seek an estimator of the connected correlation function. How to write it depends on exactly what one means by $a(\mathbf{r})$ (it can be defined as a density or as specific quantity, i.e. per unit volume or per particle), and on the fact that one wants to separate structural correlations. These issues are discussed at length in Cavagna *et al* (2018). Here it will suffice to state that a good estimator is

$$\hat{C}^{(c)}(\mathbf{r}) = \frac{\sum_{ij} \delta a_i \delta a_j \delta(\mathbf{r} - \mathbf{r}_{ij})}{\sum_{kl} \delta(\mathbf{r} - \mathbf{r}_{kl})} = \frac{\sum_{ij} \delta a_i \delta a_j \delta(r - r_{ij})}{\sum_{kl} \delta(r - r_{kl})}, \tag{49}$$

where again $\delta a_i = a_i - \bar{a}$ and the last expression is for the isotropic case. We have written δ functions for clarity, but clearly one uses binning as in (45). These expressions do a good job of disentangling correlations of the property a from structural correlations. In the case of a fixed lattice, $\hat{C}^{(c)}(\mathbf{r})$ is completely emancipated from the effects of the lattice structure, while in the case of moving particles, at least the uncorrelated effects of structure will be taken care of. Also, thanks to the denominator, (49) is free from boundary effects, so that one does need to worry about the border bias that requires the use of the Hanisch method for $g(r)$.

As written, (49) can be computed using a single configuration. Since the system is homogeneous, the mean $\langle a \rangle$ can also be estimated with a space average,

$$\bar{a} = \frac{1}{N} \sum_i a_i. \tag{50}$$

When only one configuration is available, there is no choice but to use that configuration to compute both \bar{a} and $\hat{C}^{(c)}(r)$. If several configurations, statistically equivalent and with the same boundary conditions, are

available, there are two possibilities. One is to use all configurations to compute *first* an estimate of the mean, and then compute the correlation. In this case we shall call the estimate $\hat{C}_{\text{ph}}^{(c)}(r)$, for phase average, because this procedure is closest to actually performing a phase-space, or ensemble, average. The other possibility is to use each configuration to compute the mean and correlation function (with the mean estimated with the same single configuration), and afterwards average over configurations. We shall write $\hat{C}_{\text{sp}}^{(c)}(r)$ (for space average) in this case.

An important fact about the estimate (49) in the space-average case is that $\hat{C}_{\text{sp}}^{(c)}(\mathbf{r})$ will always have a zero, exactly like the estimator for the connected time correlation (37). The reason is similar, and it can be seen considering that since $\sum_i \delta a_i = 0$, then

$$0 = \frac{1}{N} \sum_{ij} \delta a_i \delta a_j = \rho_0 \int d\mathbf{r} g_F(\mathbf{r}) C(\mathbf{r}), \quad (51)$$

where $g_F(\mathbf{r})$ is defined as the pair distribution function (43) but without excluding the case $i = j$ from the double sum. Since $g_F(r) > 0$, it follows that $\hat{C}_{\text{sp}}^{(c)}(\mathbf{r})$ must change sign, so there is always an r_0 such that $\hat{C}_{\text{sp}}^{(c)}(r_0) = 0$. This is an artefact, but it can be exploited to obtain a proxy of the correlation scale in critical systems (section 5.2).

Let us conclude by writing the Fourier-space counterpart of $\hat{C}^{(c)}(r)$. We introduce it as proportional to the reduced correlation,

$$\hat{C}(\mathbf{k}) = \frac{1}{N} \delta a(\mathbf{k}) \delta(-\mathbf{k}) = \frac{1}{N} \sum_{ij} \delta a_i \delta a_j e^{i\mathbf{k} \cdot \mathbf{r}_{ij}} = \frac{1}{N} \sum_{ij} \frac{\sin |\mathbf{k}| r_{ij}}{|\mathbf{k}| r_{ij}} \delta a_i \delta a_j. \quad (52)$$

It is related to the real space correlation by

$$\hat{C}(k) = \rho_0 \int d\mathbf{r} g_F(r) e^{-i\mathbf{k} \cdot \mathbf{r}} \hat{C}^{(c)}(r). \quad (53)$$

The function $g_F(r)$ appears in the integral (in effect reintroducing structural effects in $\hat{C}(k)$) because (52) is more convenient computational-wise than the Fourier transform of (49), and is consistent with an integration measure chosen so that the integral of $a(r)$ equals $\sum_i a_i$, which is often the global order parameter (see Cavagna *et al* 2018, appendix A).

3.3. Estimation of space-time correlations

Space-time correlation estimation brings together the issues encountered in the time and space cases. The previous discussion should allow the reader to generalise the estimators written above to the space-time case. For brevity, let us just write the space-time generalisation of (49) and (52) (for homogeneous and stationary systems):

$$\hat{C}^{(c)}(\mathbf{r}, t) = \frac{1}{T} \sum_{t_0} \frac{\sum_{ij} \delta a_i(t_0) \delta a_j(t_0 + t) \delta(\mathbf{r} - \mathbf{r}_i(t_0) + \mathbf{r}_j(t_0 + t))}{\sum_{ij} \delta(\mathbf{r} - \mathbf{r}_i(t_0) + \mathbf{r}_j(t_0 + t))}, \quad (54)$$

$$\hat{C}(\mathbf{k}) = \frac{1}{TN} \sum_{t_0} \sum_{ij} \delta a_i(t_0) \delta a_j(t_0 + t) e^{-i\mathbf{k} \cdot [\mathbf{r}_i(t_0) - \mathbf{r}_j(t_0 + t)]}. \quad (55)$$

4. Correlation time and correlation length

A connected space correlation function measures how correlation is gradually lost as one considers two locations further and further apart. Similarly, the connected time correlation function measures how correlation is gradually lost as time elapses. One often seeks for a summary of this detailed information, in the form of a (space or time) *scale* that measures the interval it takes for significant decorrelation to happen: these are the *correlation length* ξ and the *correlation time*⁴ τ .

⁴ Sometimes the term relaxation time is used as synonymous with correlation time. Actually, the relaxation time is the time scale for the system to return to stationarity after an external static perturbation is applied or removed. These two scales are equal for systems in thermodynamic equilibrium, as a consequence of the fluctuation-dissipation theorem (Kubo *et al* 1998), so the exchange of terms is admissible in that case. However, this is not to be taken for granted in the general out-of-equilibrium case.

Sometimes these are interpreted as the distance or time after which the connected correlation has descended past a prescribed level. Although this can be a useful operational definition when working with a set of correlation functions of similar shape, these scales are better understood as the asymptotic exponential rate of the decay. The precise numerical value (which will depend on the details of the computation method) is less important than their trend with the relevant control parameters (temperature, concentration, . . .): in this sense one can say that ξ and τ are most useful to compare the correlation range as some environmental condition is changed.

Of course, correlation functions are typically more complicated than a simple exponential, and it may make sense to describe them using more than one scale (e.g. as a superposition of exponential decays). The correlation length refers to the rate of the last exponential to die out. More precisely (Amit and Martín-Mayor 2006, sections II.2–3)

$$\xi = \lim_{r \rightarrow \infty} \frac{r}{-\log C_c(r)}. \quad (56)$$

This definition, with emphasis on the long-range behaviour, is particularly suited to the study of systems in the critical region. Note that $\xi = \infty$ does not mean that the correlation never decays, but only that it does so more slowly than an exponential.

Similarly, the correlation time is (Sokal 1997),

$$\tau = \lim_{t \rightarrow \infty} \frac{t}{-\log C_c(t)}. \quad (57)$$

These definitions work well when the decay is a combination of exponentials and power laws. Exponentials of powers, on the other hand, never yield a finite scale according to (56) or (57), we defer discussion of those to appendix C.

Clearly definitions (56) and (57) are not directly applicable to extract ξ or τ from correlation functions computed in experiments or simulations, which are of finite range. The rest of this section is devoted to discussing ways to obtain ξ or τ from finite data.

A possibility is to fit the decay to some function, and apply the definitions to this function. This may be fine, as long as one can avoid proliferation of parameters⁵. However, it can be difficult to fit both the short- and long-range regimes, so an alternative is to attempt fit $\log [x^\alpha C_c(x)]$ (where x stands for r or t) for large x with a straight line $-Ax$. If the fit is good then one has $\xi = 1/A$ or $\tau = 1/A$. One expects $\alpha = 0$ for the time case (at least in equilibrium), and in the space case $\alpha \approx 0$ in $d = 2$ and $\alpha \approx 1$ in $d = 3$ according to the scaling form (66). This power has no effect on the definition (56), but is important to obtain a reasonable fit on relatively short range.

Nevertheless, fitting may not be the best strategy, especially if the linear region in the fit is short and the slope is very sensitive to the choice of fitting interval. We present next some alternative procedures.

4.1. Correlation time

A quite general way to define a correlation time is from the integral of the normalized connected correlation,

$$\tau_{\text{int}} = \int_0^\infty \rho(t) dt. \quad (58)$$

Clearly for a pure exponential decay $\tau_{\text{int}} = \tau$. In general, if $\rho(t) = f(t/\tau)$ then $\tau_{\text{int}} = \text{const}\tau$, so that τ_{int} has the same dependence on control parameters as τ . This would change if $\rho(t) = (t/t_0)^\alpha f(t/\tau)$ (Sokal 1997, section 2), but one expects $\alpha = 0$ in equilibrium. In any case, τ_{int} has a meaning of its own, as the scale that corrects the variance of the estimate of the mean when computed with correlated samples (section 3.1).

With some care, τ_{int} can be computed from experimental or simulation data, avoiding the difficulties associated with thresholds or fitting functions. The following procedure (Sokal 1997) is expected to work well if long enough sequences are at disposal and the decay does not display strong oscillations or anticorrelation. If $\hat{C}_k^{(c)} \approx C_c(k\Delta t)$ is the estimate of the stationary connected correlation (37), then the integral can be approximated by a discrete sum. However, the sum cannot run over all available values $k = 0, \dots, N-1$, because the variance of $\hat{C}_k^{(c)}$ for k near $N-1$ is large (section 3.1), so that the sum $\sum_{k=0}^{N-1} \hat{C}_k^{(c)}$ is dominated by statistical noise (more precisely, its variance does not go to zero for $N \rightarrow \infty$). A way around this difficulty is (Sokal 1997, section 3) to cut-off the integral at a time $t_c = c\Delta t$ such that $c \ll N$ but the correlation is already small at $t = t_c$

⁵ ‘With four parameters I can fit an elephant, and with five I can make him wiggle his trunk.’ Attributed to John von Neumann (Dyson 2004).

(i.e. t_c is a few times τ). Thus τ_{int} is defined self-consistently as

$$\tau_{\text{int}} = \int_0^{\alpha\tau_{\text{int}}} \rho(t) dt, \quad (59)$$

where α should be chosen larger than about 5, and within a region of values such that $\tau_{\text{int}}(\alpha)$ is approximately independent of α . Longer tails will require larger values of α ; we have found adequate values to be as large as 20 (Cavagna *et al* 2012). To solve (59) one can compute $\tau(M) = \sum_k^M \hat{C}_k^{(c)} / \hat{C}_0^{(c)}$ starting with $M = 1$ and increasing M until $\alpha\tau(M) > M$.

Another useful definition of correlation time can be obtained from $\tilde{\rho}(\omega)$, the Fourier transform of $\rho(t)$. Normalisation of $\rho(t)$ implies that $1 = \rho(0) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \tilde{\rho}(\omega)$. Then a characteristic frequency ω_0 (and a characteristic time $\tau_0 = 1/\omega_0$) can be defined such that half of the spectrum of $\tilde{\rho}(\omega)$ is contained in $\omega \in [-\omega_0, \omega_0]$ (Halperin and Hohenberg 1969), i.e.

$$\int_{-\omega_0}^{\omega_0} \frac{d\omega}{2\pi} \tilde{\rho}(\omega) = \frac{1}{2}. \quad (60)$$

This definition of can be expressed directly in the time domain writing

$$\frac{1}{2} = \int_{-\omega_0}^{\omega_0} \frac{d\omega}{2\pi} \int_{-\infty}^{\infty} dt \rho(t) e^{i\omega t} = 2 \int_0^{\infty} dt \rho(t) \int_{-\omega_0}^{\omega_0} \frac{d\omega}{2\pi} e^{i\omega t} = \frac{2}{\pi} \int_0^{\infty} dt \rho(t) \frac{\sin \omega_0 t}{t}, \quad (61)$$

where we have used the fact that $\rho(t)$ is even. Then the correlation time is defined by

$$\int_0^{\infty} \frac{dt}{t} \rho(t) \sin\left(\frac{t}{\tau_0}\right) = \frac{\pi}{4}. \quad (62)$$

It can be seen that if $\rho(t) = f(t/\tau)$, then τ_0 is proportional to τ (it suffices to change the integration variable to $u = t/\tau$ in the integral above). An advantage of this definition is that it copes well with the case when inertial effects are important and manifest in (damped) oscillations of the correlation function (see figure 10 in appendix C).

4.2. Correlation length

A quite general procedure is to use the *second moment correlation length* ξ_2 (Amit and Martín-Mayor 2006, section 2.3.1),

$$\xi_2 = \sqrt{\frac{\int d\mathbf{r} r^2 C_c(\mathbf{r})}{\int d\mathbf{r} C_c(\mathbf{r})}}. \quad (63)$$

The quotient here ensures that ξ_2 scales with control parameters like ξ (a definition analogous to that of the integral correlation time would not have this property due to the power-law prefactor). ξ_2 can also be expressed in terms of the Fourier transform of $C_c(\mathbf{r})$ as

$$\xi_2^2 = -\frac{1}{\tilde{C}(\mathbf{k})} \left. \frac{\partial^2 \tilde{C}(\mathbf{k})}{\partial k_\alpha \partial k_\alpha} \right|_{k=0}. \quad (64)$$

The integrals in (63) can be computed numerically over the finite volume, but (64) may be more convenient (recall that $\hat{C}(\mathbf{k})$ can be computed directly from the data, section 3.2). Numerical derivatives should clearly be avoided; instead one can fit $\tilde{C}(k) = 1/(A + Bk^2)$ for small k , and obtain the correlation length as $\xi_2 = \sqrt{B/A}$ (which follows from (64) apart from a dimensional-dependent numerical prefactor). Another possibility is to obtain A and B using only $\tilde{C}(0)$ and $\tilde{C}(k_{\min})$ (Cooper *et al* 1982), where k_{\min} is the smallest wavenumber allowed by boundary conditions (e.g. $k_{\min} = 2\pi/L$ in a cubic periodic system), resulting in

$$\xi_2 = \frac{1}{k_{\min}} \sqrt{\frac{\hat{C}(0)}{\hat{C}(k_{\min})} - 1}. \quad (65)$$

When working with lattice systems, $2 \sin(k_{\min}/2)$ can be substituted for the k_{\min} outside the square root (Caracciolo *et al* 1993), rendering the expression exact for a Gaussian lattice model. This procedure works well when $\xi \ll L$, so that the small wavevectors are seeing an exponential behaviour for $r \sim L$.

If A/B is not much larger than L^{-2} , then ξ is approaching the system size L and the estimate is not very reliable; for ξ of the order of L or larger, A/B may even become negative. In situations like these (which include critical systems), the first root of $\hat{C}_{\text{sp}}^{(c)}(r)$, can provide a useful proxy for the correlation scale, but in that case it is necessary to compare systems of different sizes (see section 5.2).

5. Correlation functions in the critical region

The critical region refers to the volume of parameter space near the critical surface, which is a boundary between different phases of a system. The word phase may refer to a thermodynamic equilibrium phase, but it also applies to systems out of equilibrium, where different phases are characterised by qualitatively different properties (static or dynamic).

The presence of long-range correlations is a key feature of critical systems, so that correlation functions are a fundamental tool to study systems at or near criticality. Note, however, that long-range correlations are *not* synonymous with the critical point: in systems with a continuous symmetry (such as the classical Heisenberg model, in which spins are vectors on the sphere), the correlation length is infinite across all the broken-symmetry (low-temperature, magnetised) phase (Goldenfeld 1992, chapter 11).

Sufficiently close to the critical surface, the fact that the correlation is much larger than microscopic lengths allows to write the correlation functions for large enough distances in the form of so-called *scaling relations*. These are functional relations where the dependence on the microscopic parameters enters only through a few control parameters, which are often recast in terms of the correlation length. The relations involve an unknown function, which gives the shape of the correlation decay but is independent of the control parameters, or depends on a subset of them. We state first the scaling relations for the static and dynamic correlation functions as applicable to infinite systems, and afterwards consider the case of finite sizes.

Scaling relations as presented below apply and are well understood within equilibrium thermodynamics, where the system's (near) scale invariance can be exploited using the renormalization group to explain how scaling relations, as well as subleading corrections, arise (Amit and Martín-Mayor 2006, Binney *et al* 1992, Cardy 1996, Itzykson and Drouffe 1989b, Le Bellac 1991 and Wilson and Kogut 1974). In out-of-equilibrium, and in particular biological, systems our understanding is less firm, although the scale invariance found near critical points can still be exploited (see e.g. Tauber 2014). Scaling has been successfully applied to biological systems (although in this case it is typically essential to take into account the system's finite size, as discussed in section 5.2), but situations might be encountered where matters are more complicated than presented here, especially when disorder or networks with complicated topology are involved.

5.1. Scaling

The connected correlation function for an infinite system near criticality for the case of a single control parameter can be written as

$$C_c(r) = \frac{1}{r^{d-2+\eta}} f(r/\xi) = \frac{1}{\xi^{d-2+\eta}} \hat{f}(r/\xi), \quad (66)$$

where $f(x)$ is a function that tends to 0 faster than any power law for $x \rightarrow \infty$, and the second equality is just a different way of writing the same scaling law, using $\hat{f}(x) = x^{-d+2-\eta} f(x)$. The exponent of the power-law prefactor is conventionally written in this way because η is typically a small correction to $d - 2$, which is the exponent obtained from dimensional analysis (Goldenfeld 1992, chapter 7). The exponent η is called an *anomalous dimension*, and is one of the set of critical exponents that describe the singular behaviour of thermodynamic quantities near criticality (see e.g. Binney *et al* 1992).

As stated, the scaling law (66) applies for large r and finite ξ . At short distances, non-universal dependence on microscopic details may be observed (in particular note that the correlation must be finite for $r = 0$). Precisely at the critical point, the correlation does not decay exponentially but as a power law,

$$C(r) \sim r^{2-d-\eta}. \quad (67)$$

In other words, for $\xi = \infty$ the scaling function becomes a constant, even if it is not true that $f(0)$ is finite. This decay is scale-free, in the sense that if one chooses an arbitrary scale r_0 , $C(r)/C(r_0)$ is a function of r/r_0 , i.e. it can be scaled with the externally chosen length scale. In contrast, if the decay is e.g. exponential, $C(r)/C(r_0)$ will depend on r/r_0 and on r_0/ξ , i.e. the function has an intrinsic length scale ξ .

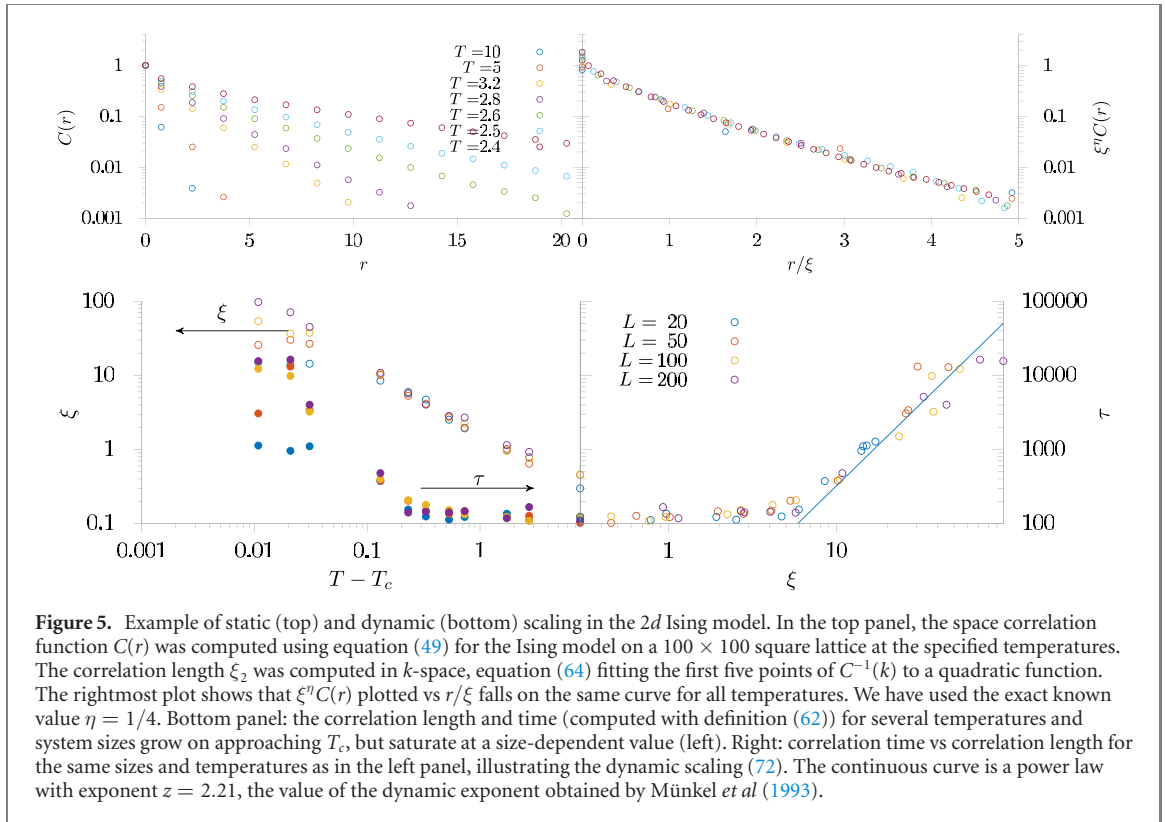
If there is more than one control variable that can take the system away from the critical surface then the scaling law can be more complicated than (66) (see Goldenfeld 1992, chapter 9). The scaling of $C_c(r)$ is illustrated in figure 5 for the $2d$ Ising model.

From (66) an equivalent scaling law can be written for the correlation in Fourier space:

$$\tilde{C}(k) = \xi^{2-\eta} F(k\xi) = k^{-2+\eta} \hat{F}(k\xi), \quad (68)$$

where $\hat{F}(x) = x^{2-\eta} F(x)$ and now the scaling law is valid for small k , with $F(x = 0)$ finite, so that $\hat{C}(k = 0)$ (which is proportional to the susceptibility) diverges when $\xi \rightarrow \infty$.

The space-time correlation can also be written in a scaling form, which ultimately leads to a link between ξ and τ . This *dynamic scaling* relation states that (Halperin and Hohenberg 1969, Hohenberg and Halperin



1977 and Tauber 2014)

$$\tilde{C}(k, \omega) = \tilde{C}(k) \frac{2\pi}{\omega_k} h(\omega/\omega_k, k\xi), \quad \omega_k = k^z \Omega(k\xi), \quad (69)$$

where $\tilde{C}(k)$ is the static correlation function and ω_k is the k -dependent inverse characteristic time defined through (60) applied to $\tilde{C}(k, \omega)$ at fixed k . Equation (69) introduces the *dynamic exponent* z and two scaling functions: $\Omega(x)$, which stays finite for $x \rightarrow \infty$, and the shape function $h(x)$, which obeys $\int_{-\infty}^{\infty} h(x) dx = 1$ and $\int_{-1}^1 h(x) dx = 1/2$ because $C(k) = C(k, t = 0) = \int_{-\infty}^{\infty} C(k, \omega) d\omega/2\pi$. In the time domain the scaling reads

$$C(k, t) = C(k) \hat{h}(t/\tau_k, k\xi), \quad \tau_k = k^{-z} \Omega^{-1}(k\xi). \quad (70)$$

In general the correlation time *and* the shape of the normalised k -dependent time correlation depend on k . If one compares systems with different correlation lengths but at different k such that the product $k\xi$ is held constant, then $C(k, t)/C(k)$ will be seen to depend on t and k only through the scaling variable $k^z t$.

One can multiply and divide by ξ^z to rewrite the scaling of τ_k :

$$\tau_k = k^{-z} \Omega^{-1}(k\xi) = \xi^z \hat{\Omega}^{-1}(k\xi), \quad (71)$$

where $\hat{\Omega}^{-1}(x) = x \Omega^{-1}(x)$. From (71) we learn the important fact that, at fixed ξ , the correlation time depends on the observation scale, more precisely that it is smaller for shorter length scales (larger k). Since τ must be finite for finite ξ , $\hat{\Omega}^{-1}(x)$ must be finite for $x \rightarrow 0$, and from the second equality we conclude that the relaxation time of global quantities (i.e. at $k = 0$) grows with growing correlation length as

$$\tau \sim \xi^z. \quad (72)$$

This important relationship highlights the fact that static, or equal-time, correlations, are nevertheless the result of a dynamic process of information exchange, and that the farther this information travels, the slower the system becomes. So static correlations should not be viewed as instantaneous correlations; in fact they contain contributions from all frequencies, since $C(k, t = 0) = \int_{-\infty}^{\infty} \tilde{C}(k, \omega) d\omega/(2\pi)$.

Note finally that in disordered systems the correlation time may grow with correlation length faster than a power law, leading to extremely slow dynamics even far from a critical point, as it happens in structural glasses (see e.g. Kob 2004). We do not consider here such systems, which require more complicated correlation functions (multi-point, or point-to-set) to detect growing correlations (Biroli *et al* 2008). A brief discussion of point-to-set correlations with emphasis on numerical simulation can be found in Cavagna *et al* (2010b);

broader theoretical reviews are Berthier and Biroli (2011), Biroli and Bouchaud (2012) and Lubchenko and Wolynes (2007).

5.2. Finite-size effects

Scaling laws as written in the previous subsection apply only to infinite systems. Although infinite systems exist only in theory, experimental systems can be considered infinite when ξ is much smaller than the system's linear size L . It is true that when the critical point is approached, ξ will eventually grow to be larger than any system size. However, in condensed matter systems samples are typically many orders of magnitude larger than microscopic lengths, and ξ can be extremely large with respect to inter-particle distances while still being smaller than L . The condition $\xi \geq L$ will only be true for a range of temperatures so thin around the critical point that it this effect can in practice be ignored.

In contrast, in experiments in biology the number of units making up the system (aminoacids, cells, organisms) is quite far from the 10^{23} or so that condensed matter systems can boast, and the effects of finite size must be properly taken into account. The same is true for numerical simulations, where in fact finite-size effects have long been used (Privman 1990) to extract information about the critical region. More recently, the effects of finite size have also been exploited in experiments on biological systems.

Finite-size effects blur the sharp transitions and singularities that distinguish phase transitions in the thermodynamics of infinite systems. If a critical system is finite, it can be scale invariant only up to its size L ; in this sense L acts as an effective correlation length if $\xi > L$. Conversely, if ξ is finite but larger than L , the system will appear nevertheless scale invariant. Thus if L is finite, scaling relations must be modified to account for the fact that correlations cannot extend beyond L . These modified relations are known as *finite-size scaling* relations (Barber 1983 and Cardy 1988).

According to the finite-size scaling ansatz (Barber 1983), when ξ is of the order or greater than L the scaling relations must be modified to include the ratio L/ξ . This ansatz can be justified with renormalisation group arguments (Amit and Martín-Mayor 2006 and Cardy 1988). For the static correlation function, instead of (66) we write

$$C(r) = \frac{1}{r^{d-2+\eta}} f(r/\xi, L/\xi), \quad \text{if } L > \xi, \quad (73)$$

with $f(x, y) \rightarrow f(x)$ for $y \rightarrow \infty$, and

$$C(r) = \frac{1}{r^{d-2+\eta}} g(r/L, L/\xi) \quad \text{if } L < \xi, \quad (74)$$

with $g(x)$ another scaling function.

At the critical point, we are effectively left with a scaling relation of the same type as (66) but with L replacing ξ (Cardy 1988),

$$C(r) = \frac{1}{r^{d-2+\eta}} g(r/L) = \frac{1}{L^{d-2+\eta}} \hat{g}(r/L) \quad (75)$$

(and a different scaling function). Correlations are truly scale-free only in the limit $L \rightarrow \infty$, where the scaling function is replaced by a constant. In a finite system, the function $g(x)$ will introduce a modulation of the power law when $r \sim L$. Scale invariance manifests instead as a correlation scale that is proportional to L : enlarging the system also enlarges the correlation scale. This, together with the fact that the space-averaged correlation function has at least one zero, can be exploited to show that a system is scale-invariant. Calling r_0 the first zero of $C_{\text{sp}}(r)$, one can show (Cavagna *et al* 2018, sections 2.3.3) that

$$r_0 \sim \xi \log(L/\xi), \quad L \gg \xi, \quad (76a)$$

$$r_0 \sim L, \quad L \ll \xi. \quad (76b)$$

We emphasise that r_0 is *not* in general a correlation length, since it behaves differently: it does not have a finite limit for $L \rightarrow \infty$ (even away from the critical point) and it does not scale like ξ with control parameters. However, if $\xi = \infty$, then r_0 is proportional to L (the only correlation scale). So, if one can show that (76b) holds for a set of systems of different sizes, but otherwise equivalent, one can argue that those systems are in fact scale-free. Such reasoning has been used e.g. in Attanasi *et al* (2014), Cavagna *et al* (2010a), Fraiman and Chialvo (2012) and Tang *et al* (2017).

To expand on this important issue, we must note that scale invariance, as evidenced by the validity of relation (76b) can arise when the system's parameters are fixed at the values corresponding to the critical point in the thermodynamic limit ($\xi = \infty$), as shown above, but in a subtler way as well (also related to criticality): if one varies the system size *and* control parameters together in such a way that the ratio L/ξ remains constant, then relation (74) still leads to (75) and (76). This is what Attanasi *et al* (2014) find for midge swarms: swarms

of different sizes form at different densities (density is the control parameter here), and yet for the set of swarms it still holds that $r_0 \sim L$, implying a scaling of the form (75). Since ξ is changing from system to system (because the density changes), what must be happening is that for each size the density is adjusted so that the ratio L/ξ is constant across all swarms, and furthermore ξ must be much larger than L , or r_0 would not be proportional to L . It is remarkable that although midges are not programmed to fly at some predefined distance from one another, they still find a way to tune themselves to near-criticality and display scale-free behaviour. How a tuning of this kind can happen is open to speculation (see Attanasi *et al* 2014 and Chialvo *et al* 2020), but it is important to note that, as Attanasi *et al* (2014) emphasise, the critical point is not strictly defined for systems of finite size, and that the maximum susceptibility will be found for size-dependent values of the control parameters. This pseudo-critical value can be noticeably different from the rigorous thermodynamic-limit critical value. We thus have to consider the possibility that some biological systems are critical not because through evolution they found optimal values of their working parameters, but because evolution has endowed them with some means of adjusting those parameters.

Finally, turning to Fourier space, the finite-size version of (68) can be written, for $L < \xi$,

$$\hat{C}(k; L) = k^{-2+\eta} F(kL, L/\xi) = L^{2-\eta} \hat{F}(kL, L/\xi). \quad (77)$$

The finite-size version of the dynamic scaling law (70) is

$$\tilde{C}(k, t; L) = \tilde{C}(k; L) \hat{h}(t/\tau_k, kL, L/\xi), \quad \tau_k = k^{-z} \Omega^{-1}(kL, L/\xi). \quad (78)$$

Then, at the scale-free point $\xi = \infty$ it holds that $\tilde{C}(k, t; L)$ measured for systems of different size depends on $k^z t$ if one measures each system at a k such that $kL = \text{const}$. This scaling law has been successfully applied to study the dynamical behaviour of swarms of midges in the field (Cavagna *et al* 2017). For the global relaxation time one has, instead of (72)

$$\tau = L^z \hat{\Omega}^{-1}(L/\xi). \quad (79)$$

To conclude, let us point out that from what we have said it follows that a single measurement of ξ on a finite system does not tell one anything about whether correlations are long range or not. An unknown numerical prefactor is involved in the estimation of ξ , and the estimators we have discussed tend to be less reliable when $\xi \sim L$, so that it does not make sense to discuss whether, say, $\xi = L/2$ is large or not. Instead, it is necessary to measure the trend of ξ with control parameters, and if at all possible, with L . Moreover, studying different system sizes can show whether the correlation scale grows with L and establish that the system is scale-free in situations where changing the control parameters is not feasible.

There can be situations, especially in biological experiments, where changing system size is not possible (e.g. how would one study human brains of different sizes?). In such circumstances, one may attempt to do finite-size scaling by measuring subsystems ('boxes') of size $W \ll L$ and studying how the quantities change when varying W . The idea can be traced back to Binder (1981), and has also been employed out of equilibrium (Fernández and Martín-Mayor 2015). The applicability of relations (76) with the box size W in place of L for systems of fixed size was recently demonstrated for several simulated systems by Martin *et al* (2021).

6. Space-time correlations in a neuronal network model

The finding of scale-free avalanches in cortical tissue (Beggs and Plenz 2003 and Shew *et al* 2009) and strong correlations in retinal cells (Schneidman *et al* 2006, Tkačik *et al* 2015 and 2006) are early experimental evidence of criticality in neuronal populations (see Mora and Bialek (2011) and Chialvo (2010) for reviews of these and more recent experiments). At the level of the whole brain, fMRI measurements of neuronal activity (Expert *et al* 2011, Fraiman and Chialvo 2012 and Tagliazucchi *et al* 2012) have found long-range correlation, providing support for the conjecture that the brain operates at a critical point (Bak 1996, Beggs 2008 and Chialvo and Bak 1999). Dynamic systems studies (e.g. Deco and Jirsa 2012) have also provided support for the criticality hypothesis.

Notwithstanding the importance of these works in forwarding our understanding of how complex behaviour can arise in the brain, it is fair to say that the picture of the brain as a critical system is still work in progress, and surely new experiments will be attempted in the near future. Since critical systems display distinctive dynamic as well as static characteristics, much information can potentially be gained by studying systematically static and dynamic correlations together on the same system, and in particular the scaling relations. For instance, study of static correlations in midge swarms using finite-size scaling led to the finding (Attanasi *et al* 2014) that they are in a critical regime compatible with the critical Vicsek model (Vicsek *et al* 1995). A subsequent investigation of dynamic correlations in the same system showed that they obey dynamic scaling (Cavagna *et al* 2017), but with a novel dynamic exponent, incompatible with the dynamics of the Vicsek model. The exponent turned out to be different from those of the classical dynamic models (Hohenberg

and Halperin 1977), and this spurred still-ongoing efforts to understand whether a combination of inertia and activity ingredients can explain this new exponent (Cavagna *et al* 2019).

Clearly, magnets, swarms and brains are quite different systems, and in particular the special network structure of neuronal assemblies poses specific problems (Korchinski *et al* 2021). Certainly it is not the case to take tools developed in the statistical mechanics of condensed matter and blindly apply them to study the brain. But knowledge of other critical systems suggests the relation between correlation time and length scales can hold valuable information. An optimal opportunity to attempt this kind of study may come from novel optogenetics techniques being applied to the simultaneous recording of the activity of thousands of neurons (see Emiliani *et al* 2015, for a review). In that direction we can mention the recent experimental report (Ribeiro *et al* 2020) describing the relation (76) with the box size W computed from opto-genetic data from the visual and auditory cortices of freely behaving mice. The question, of course, deserves careful experimental study, but we wish to conclude this article with an exploration of a simple excitable network model, which can serve as some illustration to the exposition of correlation functions, and hopefully stir up interest for experimental studies of length and time correlation scales in the brain.

6.1. The ZBPCC model

We choose the simple neural network model proposed by Zarepour *et al* (2019). The model is a Greenberg–Hastings cellular automaton (Greenberg and Hastings 1978) implemented on a small-world network (this model was also used by Haimovici *et al* (2013) but on a different graph). In detail, each site of the network has three possible states, $S_i = \{Q, E, R\}$, corresponding to quiescent, excited or refractory, respectively, and the transitions among them are ruled by the probabilities

$$P_{i,Q \rightarrow E} = 1 - [1 - r_1] \left[1 - \Theta \left(\sum_j W_{ij} \delta_{S_j, E} - T \right) \right], \quad (80a)$$

$$P_{i,E \rightarrow R} = 1, \quad (80b)$$

$$P_{i,R \rightarrow Q} = r_2. \quad (80c)$$

$P_{i,a \rightarrow b}$ is the probability that site i will transition from a to state b , $\Theta(x)$ is Heaviside's step function ($\Theta(x) = 1$ for $x \geq 0$ or 0 otherwise), δ_{ij} is Kronecker's delta, r_1 , r_2 and T are numeric parameters and W_{ij} is the network's connectivity matrix (see below). Thus an active site always turns refractory in the next time step, and a refractory site becomes quiescent with probability r_2 . We set $r_2 = 0.3$ as in previous work (Haimovici *et al* 2013 and Zarepour *et al* 2019), so that the site remains refractory for a few time steps. The probability for a quiescent site to become active is written as 1 minus the product of the probabilities of *not* becoming active through the two mechanisms at work: *spontaneous activation*, which occurs with a small probability r_1 (mimicking an external stimulus), or *transmitted activation*, which occurs with certainty if the sum of the weights of the links connecting i to its active (excited) neighbours exceeds a threshold T . All sites are updated simultaneously, and we chose $r_1 = 10^{-6}$ so that external excitation events are relatively rare.

The model runs on an undirected weighted graph described by the connectivity matrix elements $W_{ij} \geq 0$, where a nonzero element means a bond with the specified weight connects sites i and j , and $W_{ij} = W_{ji}$. Neither the connectivity nor the weights depend on time. The graph is a bidirectional Watts–Strogatz small-world network (Watts and Strogatz 1998) with average connectivity $\langle k \rangle$ and rewiring probability π . The network is constructed as usual (Watts and Strogatz 1998) by starting from ring of N nodes, each connected symmetrically to its $\langle k \rangle / 2$ nearest neighbours; then each link connecting a node to a clockwise neighbour is rewired to a random node with probability π , so that average connectivity is preserved. The rewiring probability is a measure of the disorder in the network; $\pi = 0$ preserves the original network topology, while $\pi = 1$ gives a completely random graph. Once the network bonds are drawn, each is assigned a random weight drawn from an exponential distribution, $p(W_{ij} = w) = \lambda e^{-\lambda w}$, with $\lambda = 12.5$ chosen to mimic the weight distribution of the human connectome (Zarepour *et al* 2019).

The simplest way to characterise the state of the network is perhaps the activity order parameter, i.e. the fraction of active sites

$$a = \frac{1}{N} A, \quad A = \sum_i \delta_{S_i, E}. \quad (81)$$

By analogy to magnetic systems, one can define an associated susceptibility, proportional to the variance of a ,

$$\chi = \frac{1}{N} [\langle A^2 \rangle - \langle A \rangle^2], \quad (82)$$

where $\langle . . . \rangle$ is a time average.

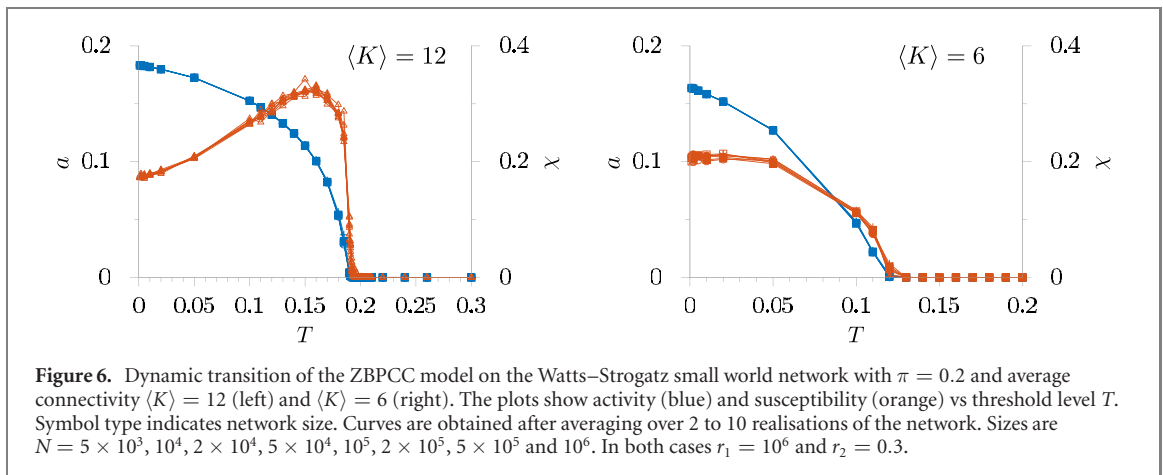


Figure 6. Dynamic transition of the ZBPCC model on the Watts–Strogatz small world network with $\pi = 0.2$ and average connectivity $\langle K \rangle = 12$ (left) and $\langle K \rangle = 6$ (right). The plots show activity (blue) and susceptibility (orange) vs threshold level T . Symbol type indicates network size. Curves are obtained after averaging over 2 to 10 realisations of the network. Sizes are $N = 5 \times 10^3, 10^4, 2 \times 10^4, 5 \times 10^4, 10^5, 2 \times 10^5, 5 \times 10^5$ and 10^6 . In both cases $r_1 = 10^6$ and $r_2 = 0.3$.

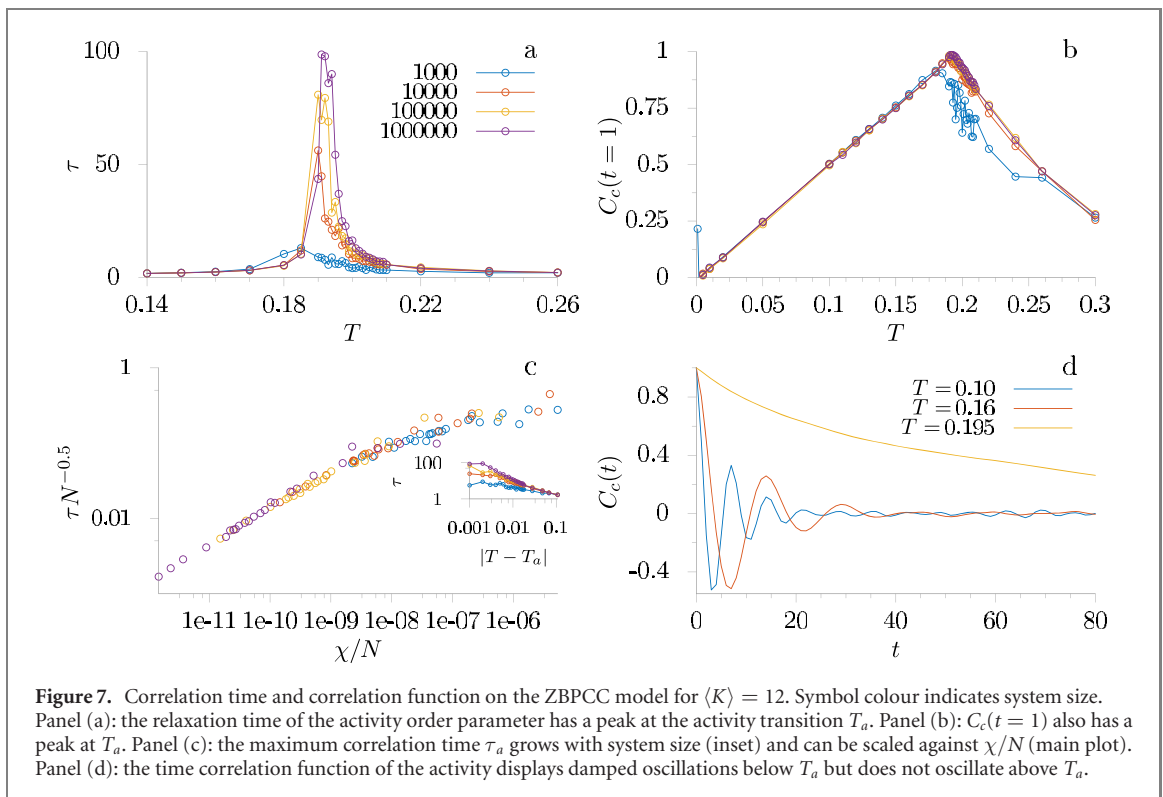
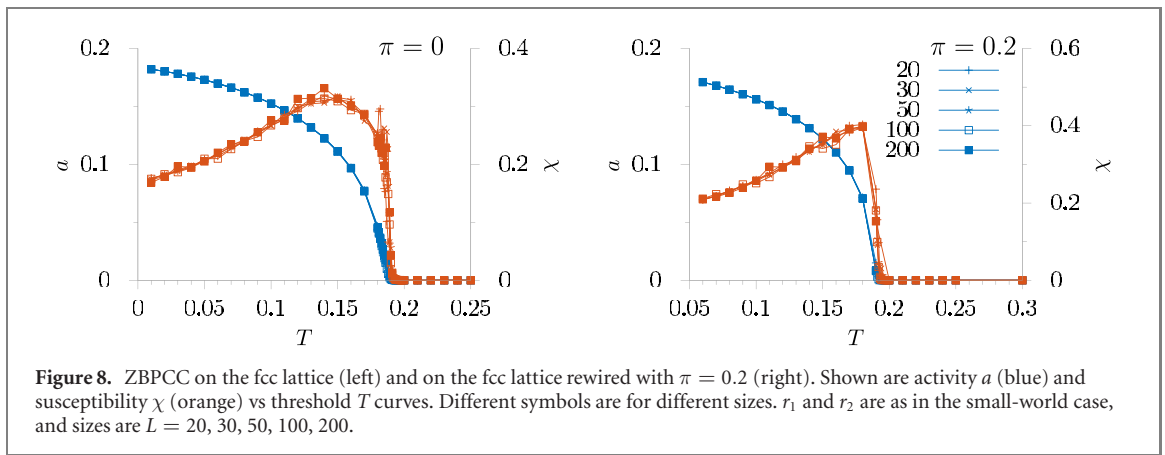


Figure 7. Correlation time and correlation function on the ZBPCC model for $\langle K \rangle = 12$. Symbol colour indicates system size. Panel (a): the relaxation time of the activity order parameter has a peak at the activity transition T_a . Panel (b): $C_c(t = 1)$ also has a peak at T_a . Panel (c): the maximum correlation time τ_a grows with system size (inset) and can be scaled against χ/N (main plot). Panel (d): the time correlation function of the activity displays damped oscillations below T_a but does not oscillate above T_a .

We computed network activity as a function of T for $\pi = 0.2$ and average connectivity $\langle K \rangle = 12$ and $\langle K \rangle = 6$ (figure 6). In both cases we find what appears to be a continuous transition, where activity goes smoothly to zero at $T_a \approx 0.19$ for $\langle K \rangle = 12$ and at $T_a \approx 0.12$ for $\langle K \rangle = 6$. The transition is however unusual in that χ has a (not very sharp) maximum at a different value of T , and the height of the peak in χ does not scale with size (i.e. unlike usual second order transitions, here χ never diverges).

According to Zarepour *et al* (2019), for this value of π and $\langle K \rangle = 12$ one has a continuous transition, as we find here, but $\langle K \rangle = 6$ lies in the ‘no transition’ region of $(\pi, \langle K \rangle)$ space (see their figure 4). This discrepancy is due to the fact that Zarepour *et al* (2019) chose to study the percolation of active clusters instead of the total activity. Here we prefer the simpler activity parameter, which can be measured without knowledge of network connectivity (which may be difficult to obtain experimentally). Also it turns out that the transition described in terms of the activity order parameter is consistent with the dynamical behaviour, as discussed next.

The time correlation of the activity reveals a noticeable slow-down of the dynamics near T_a . We have computed the correlation time τ_a , defined according to (62), from the time correlation function of the order parameter (see figure 7). The correlation time shows a peak that grows with system size at the activity transition T_a . Looking at the normalised time correlation function after one step, $C_c(t = 1)$, a quantity which peaks at a regular second order phase transition (Chialvo *et al* 2020), we find it displays a clear peak at T_a .



The peak of τ_a vs T for growing system size can also be appreciated in logarithmic scale (figure 7(c), inset), with curves of τ_a vs $|T - T_a|$ saturating at higher values for larger systems. Unfortunately, it is not possible to explore standard dynamic scaling in this model, because it is defined on a graph with no spatial structure, so that space correlations cannot be defined. We make an attempt using χ (which in a ferromagnetic transition would scales with a power of ξ , although here it never diverges). Somewhat surprisingly, a reasonable scaling behaviour is obtained plotting $N^{-1/2}\tau$ vs χ/N , i.e. the variance of a .

Finally, inspection of the full correlation function shows a qualitative change in its shape, with damped oscillations found at low T giving way to an overdamped-like decay at higher T (figure 7(d)).

6.2. ZBPCC on the fcc lattice

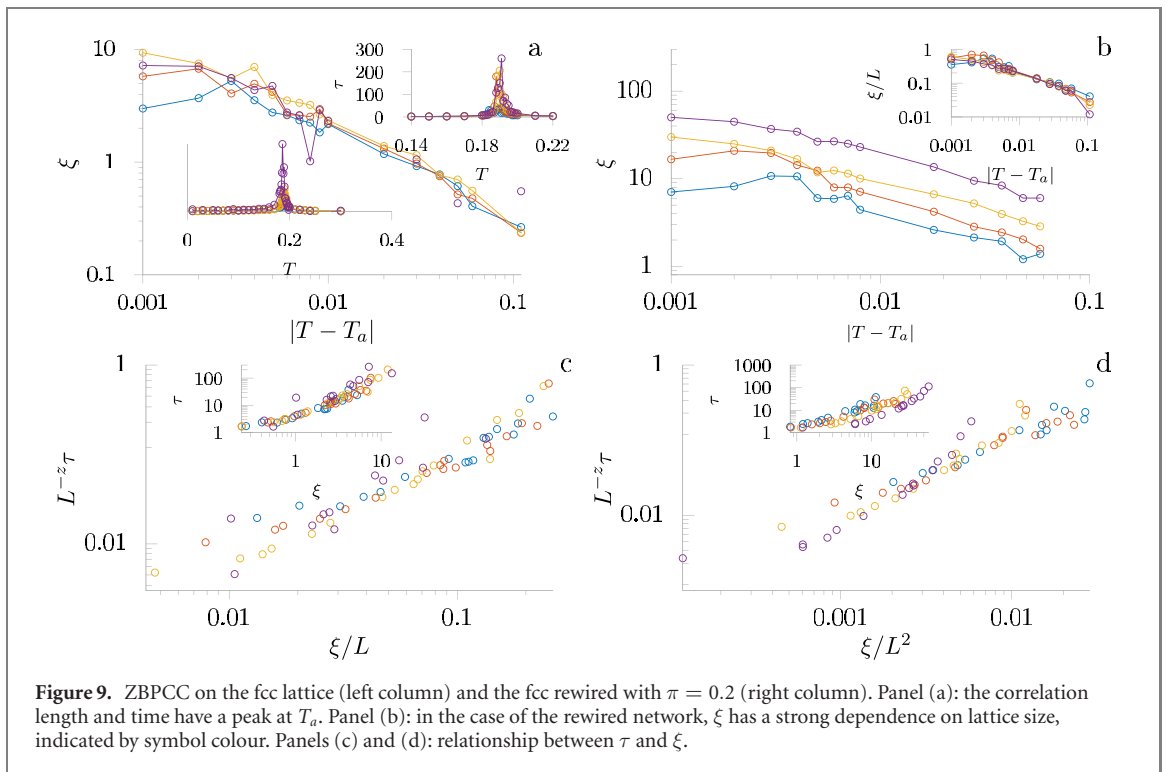
To overcome the impossibility of computing space correlations on the ZBPCC model, we explore the simplest strategy to endow the ZBPCC with a spatial structure: we build a graph connecting the nearest neighbours of a regular lattice, and then rewire it with probability π as in the small world case. We chose the face-centred cubic (fcc) lattice, which has connectivity $K = 12$ (a value in which a transition is observed both with the present approach focusing on the activity and in the approach of Zarepour *et al*).

Figure 8 shows the behaviour of the activity for the pure fcc lattice ($\pi = 0$) and for the fcc lattice rewired with $\pi = 0.2$. The fcc lattice shows a situation similar to the small world, with a transition in a but a maximum in χ shifted with respect to T_a , although for the rewired lattice the maximum is closer to T_a . In both cases $T_a \approx 0.19$. These results suggest that T_a depends mainly on the network connectivity, and is not very sensitive to the details of network topology.

In this version of the model we can use the underlying fcc lattice to define distances and compute space correlations. We have computed $C(k)$ and extracted ξ_2 from it using definition (64) and a quadratic fit at small k (figure 9). For $\pi = 0$, the largest ξ is about a tenth of the lattice size, and finite-size effects are modest. On the contrary, on the rewired fcc, there is a strong size effect. However, instead of the situation usual in a regular lattice, with ξ vs T curves superimposing at high T and saturating at different values on approaching the critical point (see figure 5), here ξ grows with L almost uniformly for all T , and the ξ vs T curves seem to scale when dividing by L (figure 9(b), inset). Given that this effect only appears with rewiring, it may be related to the fact that when a link is rewired, the new neighbour is chosen randomly over all the nodes without regards to the euclidean distance, a procedure that is not very realistic.

Finally we have computed the correlation time of the activity. This shows a size-dependent maximum at T_a in both cases, confirming the dynamical relevance of T_a . Comparing τ_a and ξ shows that both grow together, but the attempt at applying dynamic scaling (figures 9(c) and (d)) is not very satisfactory. A more detailed analysis is needed of the role played by topological details and network disorder.

This brief exploration of neuronal network models should serve to illustrate the usefulness of a combined study of space and time correlations on these systems, even though their behaviour will not necessarily turn out to be exactly like what is found in condensed matter systems. In this case, the time correlation of the activity suggests that total activity is the relevant order parameter to characterise the network state and dynamic transition. However, several questions remain unanswered at this point, such as the relationship between the activity transition and the percolation transition studied by Zarepour *et al* (2019), the reasons behind the scarce size-effects of the correlation length in the fcc case, and the strange scaling of ξ with L and of τ with ξ in the rewired fcc. The somewhat artificial way space has been introduced in the model is perhaps to blame for some of the strange features of space correlations, and it is worthwhile to consider models with realistic spatial structure and connectivity, using directly the human connectome for a graph, as in Haimovici *et al* (2013) or



Ódor (2016). However, a thorough exploration of these issues would take us too far from the aim of the present article, and are left to be pursued in future work.

7. Conclusions

We have presented space (static), time (dynamic), and space-time correlation functions, and discussed how they can be defined and computed in a variety of situations. We have also considered how to define and compute characteristic correlation lengths and times, and summarised our knowledge of scaling relations and finite-size effects for systems near a critical point. Finally we have used the simple ZBPCC model as an illustration of the kind of analysis that could be carried out in more realistic neuronal models, or even in experiments on networks of neurons.

We have argued that these tools from statistical physics are extremely useful to study collective phenomena in biological systems, because they are designed to capture information about the space and time structure of fluctuations around average quantities, and in that structure lies much information on the nature of complex systems. It should be clear that correlation functions are not ‘one size fits all’. We have discussed a variety of them, and only within two-point correlations. The choice of the fluctuating quantity to consider can lead to different descriptions of the same system, as the example of the ZBPCC model shows, and reconciling seeming contradictions will only expand one’s understanding of the system under study. In this sense, more is better, and especially so when dealing with biological systems where expectations based on naive transposition of equilibrium thermodynamics ideas about phase transitions can prove wrong. In particular, we have advocated studying both space and time correlations whenever possible. Also in the case of slowly driven out-of-equilibrium systems, their study together with other indicators of criticality such as power-law distributions of relaxation events (avalanches) can be highly informative (see Jensen 2021, in this focus issue).

We hope this paper will be useful to students and researchers outside statistical physics interested in computing correlation functions from simulation and experimental data. It might also help convince experimentalists in biology to invest in the necessary effort to acquire new experimental data with enough resolution to compute space and time correlations.

Acknowledgments

I thank D Chialvo for helpful discussions on the ZBPCC model, and for bringing some references to my attention. I also thank A Cavagna and I Giardina, who over a long collaboration helped shape my understanding of

concepts and practicalities of correlation functions, especially as applied to biological systems. This work was supported in part by the BRAIN initiative Grant U19 NS107464-01, and by Grants from Universidad Nacional de La Plata and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina).

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Appendix A. Stochastic processes

A *stochastic process*, or sometimes *random field*, is a family of random variables indexed by \mathbf{r} and t . Thus there must exist a family of probability densities $P(a, \mathbf{r}, t)$ that allows to compute all moments $\langle a^n(\mathbf{r}, t) \rangle$ and in general any average $\langle f(a(\mathbf{r}, t)) \rangle$. But $P(a, \mathbf{r}, t)$ is not enough to characterise the stochastic process, because in general the variables $a(\mathbf{r}, t)$ at different places/times will not be independent. Thus $P(a, \mathbf{r}, t)$ is actually a marginal distribution of some more complicated multivariate probability density.

A complete characterisation of the stochastic process would need an infinite-dimensional probability distribution $P[a(\mathbf{r}, t)]$ that gives the joint probability for all random variables $a(\mathbf{r}, t)$. However, one can usually avoid the difficulties associated with such a distribution by considering the set of n -variable joint distributions

$$P_n(a_1, \mathbf{r}_1, t_1, a_2, \mathbf{r}_2, t_2, \dots, a_n, \mathbf{r}_n, t_n). \quad (\text{A1})$$

Most stochastic processes (Priestley 1981, section 3.2.1) can be completely specified by the set of all joint probabilities of the form (A1) for all n and all possible choices of $\mathbf{r}_1, t_1, \dots, \mathbf{r}_n, t_n$. This should include all processes of interest in physics and biology. To define two-point correlation functions we need only the set corresponding to $n = 2$ (7), which gives also the set for $n = 1$ by marginalising on the second variable.

The joint probability distributions that define the process are TTI if they obey (we omit the space variables for clarity)

$$P_n(A_1, t_1, \dots, A_n, t_n) = P(A_1, t_1 + s, \dots, A_n, t_n + s) \quad (\text{A2})$$

for all times, s and n . Stochastic processes obeying TTI are called *completely stationary processes*. A related but less restrictive notion is that of stationary processes up order M , defined by the requirement that all the joint moments up to order M exist and are TTI:

$$\langle A^{m_1}(t_1) A^{m_2}(t_2) \dots A^{m_n}(t_n) \rangle = \langle A^{m_1}(t_1 + s) A^{m_2}(t_2 + s) \dots A^{m_n}(t_n + s) \rangle \quad (\text{A3})$$

for all $s, n, \{t_1, \dots, t_n\}$ and $\{m_1, \dots, m_n\}$ such that $m_1 + m_2 + \dots + m_n \leq M$. This is less restrictive not only because of the bound on the number of random variables considered, but also because invariance is imposed only on the moments, and not on the joint distributions themselves.

Appendix B. The estimator $\hat{C}_k^{(c)}$ always has a zero

An important property of estimator (37) will always have zero, whatever the length of the sequence used to compute it. To see this, consider the quantity

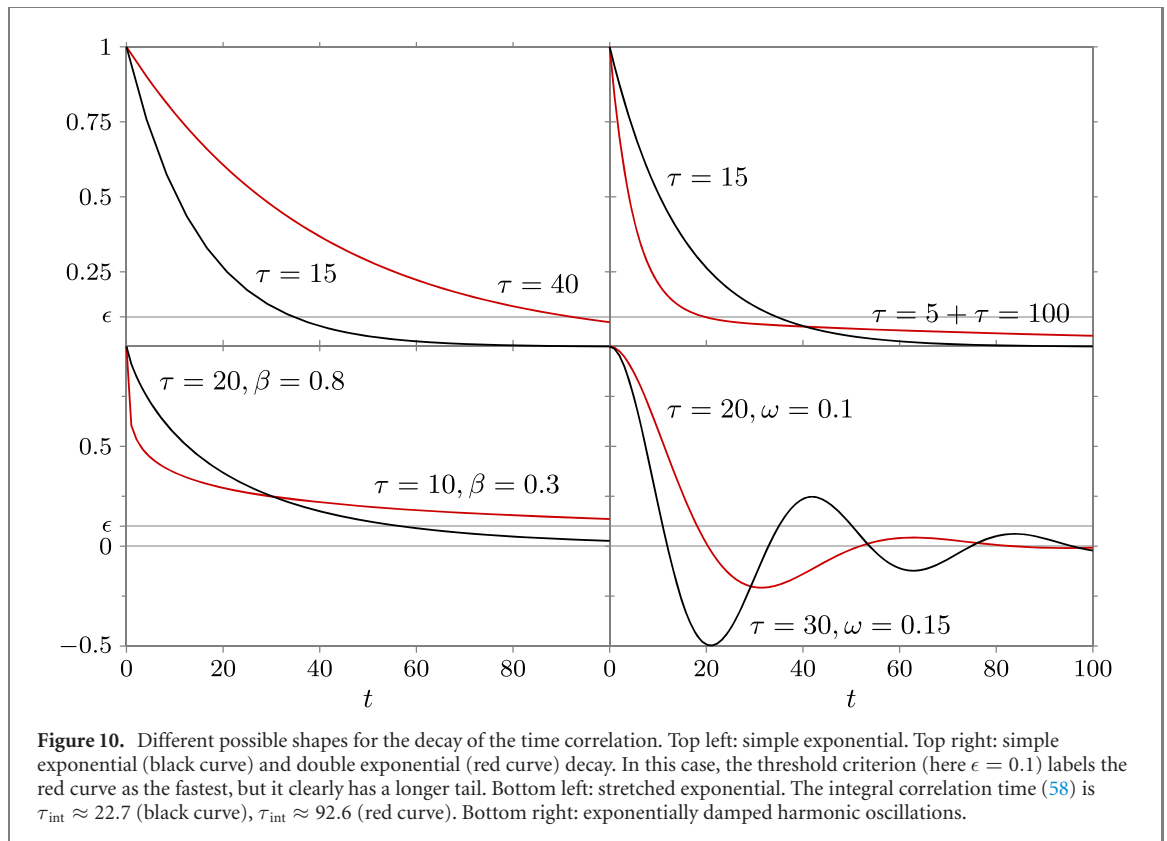
$$B_i = (N - i) \hat{C}_{c,i} = \sum_{j=1}^{N-i} \delta a_j \delta a_{j+i}, \quad (\text{B1})$$

and compute the sum

$$\sum_{i=0}^{N-1} B_i = \sum_{i=0}^{N-1} \sum_{j=1}^{N-1-i} \delta a_j \delta a_{j+i} = \sum_{i=0}^{N-1} \sum_{j=1}^N \sum_{k=1}^N \delta a_j \delta a_k \delta_{k,j+i}. \quad (\text{B2})$$

But $\sum_{i=0}^{N-1} \delta_{k,j+i}$ equals 1 if $k \geq j$ and 0, so

$$\begin{aligned} \sum_{i=0}^{N-1} B_i &= \sum_{j=1}^N \sum_{k=j}^N \delta a_j \delta a_k = \frac{1}{2} \sum_{j \neq k}^N \delta a_j \delta a_k + \sum_{j=1}^N (\delta a_j)^2 \\ &= \frac{1}{2} \left[\sum_{j=1}^N \delta a_j \right]^2 + \frac{1}{2} \sum_{j=1}^N (\delta a_j)^2 = \frac{1}{2} \sum_{j=1}^N (\delta a_j)^2 > 0, \end{aligned} \quad (\text{B3})$$



where the last equality follows because $\sum_j \delta a_j = 0$. Now we can easily do the same sum starting from $i = 1$:

$$\sum_{i=1}^{N-1} B_i = -B_0 + \sum_{i=0}^{N-1} B_i = -\frac{1}{2} \sum_{j=1}^N (\delta a_j)^2 < 0. \tag{B4}$$

This shows that at least some of the B_i must be negative. But since $B_0 > 0$, the conclusion is that B_k , and hence $\hat{C}_{c,k}$, which differs from it by a positive factor, must change sign at least once for $k \geq 1$.

Appendix C. The shape of the correlation decay

Figure 10 illustrates several (simple) possible shapes of a decaying function. It should make it clear why a threshold to define a correlation scale is misleading unless the curves to be compared are all the same shape. Apart from the simple exponential $\rho(t) = e^{-t/\tau}$ and the sum of exponentials, the figure illustrates two interesting cases.

One is the Kolrausch–William–Watts functions (lower left panel),

$$\rho_K(t) = e^{-(t/\tau_K)^\beta}, \tag{C1}$$

a function widely used to describe non-exponential decays, employing only three parameters. Here τ_K is a time scale and β is the stretching exponent: the KWW function gives a stretched exponential for $\beta < 1$, and a compressed exponential for $\beta > 1$. Definition (57) gives 0 in the compressed case, and ∞ in the stretched case. This can be understood considering the decay as a superposition of exponential processes,

$$\rho_K(t) = \int_0^\infty w(\tau) e^{-t/\tau} d\tau, \tag{C2}$$

which defines the correlation time distribution function $w(\tau)$. It can be seen that the support of $w(\tau)$ extends to infinity for $\beta < 1$ (Lindsey and Patterson 1980). The integral correlation time is instead more useful here,

$$\tau_{\text{int}} = \frac{\tau_K}{\beta} \Gamma\left(\frac{1}{\beta}\right), \tag{C3}$$

where $\Gamma(x)$ is Euler’s gamma function. So τ_{int} is proportional to τ_K , but τ_{int} incorporates information about the shape of the tail of the decay, making it better suited to compare two decays with different β .

Algorithm 1. Compute the connected correlation of sequence a (of length N) using the direct $O(N^2)$ method. The connected correlation is returned in vector C .

```

1: function TIMECORR( $a, N$ )
2:    $\mu \leftarrow 0$  ▷ Compute average
3:   for  $i = 1, \dots, N$  do
4:      $\mu \leftarrow \mu + a_i$ 
5:    $\mu \leftarrow \mu/N$ 
6:   for  $i = 1, \dots, N$  do ▷ Clear  $C$  vector
7:      $C_i \leftarrow 0$ 
8:   for  $i = 1, \dots, N$  do ▷ Correlation loop
9:      $d \leftarrow a_i - \mu$ 
10:    for  $k = 0, \dots, N - i$  do
11:       $C_{k+1} \leftarrow C_{k+1} + d^k (a_{i+k} - \mu)$ 
12:    for  $i = 1, \dots, N$  do ▷ Normalize and return
13:       $C_i \leftarrow C_i / (N - i - 1)$ 
14:    return  $C$ 

```

The other case is that of damped oscillations, that appear in systems with important inertial effects. In the simplest harmonic form we have

$$\rho_{\text{osc}}(t) = e^{t/\tau} \cos(\omega_0 t). \quad (\text{C4})$$

The integral (58) and spectrum (62) scales are

$$\tau_{\text{int}} = \frac{\tau}{1 + (\omega_0 \tau)^2}, \quad \tau_0 = \frac{\tau}{\sqrt{1 + (\omega_0 \tau)^2}}. \quad (\text{C5})$$

In the overdamped case, $\omega_0 \tau \ll 1$ both scales tend to τ , but in the opposite limit $\omega_0 \tau \gg 1$, where the function oscillates many times before significant damping is observed, $\tau_{\text{int}} \sim 0$ while $\tau_0 \sim 1/\omega_0$.

Appendix D. Algorithms to compute time correlations

The estimators can be computed numerically by straightforward implementation of equations (35) or (37), although in the stationary case it is much more efficient to compute the connected correlation through relation (31) using a fast Fourier transform (FFT) algorithm. Let us focus on the stationary case and examine in some detail these algorithms.

Algorithm 1 presents the direct method. It is straightforward to translate the pseudo-code to an actual language of your choice. Apart from some missing variable declarations, the only thing to consider is that it is probably inconvenient (or even illegal, as in classic C) to return a large array, and it is better to define C as an output argument, using a pointer or reference (as e.g. FORTRAN or C do by default) to avoid copying large blocks of data. The advantages of this algorithm are that it is self-contained and simple to understand and implement. Its main disadvantage is that, due to the double loop of lines 8–11, it runs in a time that grows as N^2 . For N up to about 10^5 , algorithm 1 is perfectly fine: a good implementation in a compiled language should run in a few seconds in a modern computer. But this time grows quickly; in the author's computer $N = 5 \times 10^5$ takes 35 s, for $N = 10^6$ the time is two and a half minutes. In contrast, the algorithm with FFT takes 1 s for $N = 10^6$ and 11 s for $N = 10^7$.

If the correlation of really long sequences is needed, the FFT-based algorithm, though more difficult to get running, pays off with huge savings in CPU time at essentially the same numerical precision. The idea of the algorithm is to compute the Fourier transform of the signal, use (30) to obtain the Fourier transform of the connected correlation, then transform back to obtain $C_c(t)$. This is faster than algorithm 1 because the clever FFT algorithm can compute the Fourier transform in a time that is $O(N \log N)$.

Actually, we need discrete versions of the Fourier transform formulas (as we remarked before, the Fourier transform of the continuous time signal does not exist). The *discrete Fourier transform* (DFT) and its inverse operation are defined (Press *et al* 1992, section 12.1) as (it is convenient to let the subindex of a_i run from 0 to $N - 1$ to write the following two equations),

$$\tilde{a}_k = \sum_{j=0}^{N-1} e^{2\pi i j k / N} a_j, \quad a_j = \frac{1}{N} \sum_{k=0}^{N-1} e^{-2\pi i j k / N} \tilde{a}_k, \quad (\text{D1})$$

Algorithm 2. Compute the connected correlation of sequence a (of length N) using a FFT. This algorithm is $O(N \log N)$.

```

1: function TIMECORR( $a, N$ )
2:    $\mu \leftarrow 0$  ▷ Compute average
3:   for  $i = 1, \dots, N$  do
4:      $\mu \leftarrow \mu + a_i$ 
5:    $\mu \leftarrow \mu / N$ 
6:   for  $i = 0, \dots, N$  do ▷ Subtract the average from signal
7:      $a_i \leftarrow a_i - \mu$ 
8:   for  $i = N, \dots, 2N$  do ▷ Pad with 0s at the end
9:      $a_i \leftarrow 0$ 
10:   $b \leftarrow \text{FFT}(a, 2N)$  ▷ Compute the FFT of  $a$  as a vector of length  $2N$ 
11:  for  $i = 1, \dots, 2N$  do ▷ Compute squared modulus of  $b$ 
12:     $b_i \leftarrow |b_i|^2$  ▷ Note that the Fourier transform is complex
13:   $C \leftarrow \text{IFFT}(b, 2N)$  ▷ Inverse FFT
14:   $C \leftarrow \text{RESIZE}(C, N)$  ▷ Discard the last  $N$  elements of  $C$ 
15:  for  $i = 1, \dots, N$  do ▷ Normalize and return
16:     $C_i \leftarrow C_i / (N - i - 1)$ 
17:  return  $C$ 

```

where we note that the inverse DFT effectively extends the sequence periodically (with period N). The discrete version of (30) is (Press *et al* 1992, section 13.2)

$$\tilde{D}_k = |\tilde{a}_k|^2, \quad \text{where } D_j = \sum_{k=0}^{N-1} a_k a_{k+j}, \quad (\text{D2})$$

and where the definition of D_j makes use of the (assumed) periodicity of a_i . D_j is almost our estimate (37): we only need to take care of the normalization and of the fact that due to the assumed periodicity of a_i some past times are regarded as future, e.g. for $k = 10$, in the sum there appear the terms $a_0 a_{10}$ up to $a_{N-11} a_{N-1}$ (which are fine), but also $a_{N-10} a_0$ through $a_{N-1} a_9$, which we do not want included. This is fixed by padding the original signal with N zeros at the end, i.e. setting $a_k = 0$ for $k = N, \dots, 2N - 1$ and ignoring the values of D_j for $j \geq N$.

In summary, to compute the connected correlation using FFT the steps are (i) estimate the mean and subtract from the a_i , (ii) add N zeros to the end of the sequence, (iii) compute the DFT of the sequence, (iv) compute the squared modulus of the transform, (v) compute the inverse DFT of the squared modulus, (vi) multiply by the $1/(N - i)$ prefactor. Pseudocode for this algorithm is presented as algorithm 2.

To translate this into an actual programming language the comments made for algorithm 1 apply, and in addition some extra work is needed for lines 10–13. First, one needs to choose an FFT routine. The reader curious about the FFT algorithm can read for example Press *et al* (1992) (chapter 12) or Duhamel and Vetterli (1990), but writing an FFT routine is not easy, and implementing a state-of-the-art FFT is stuff for professionals. Excellent free-software implementations of the FFT can be found on Internet. FFTW (Frigo and Johnson 2005), at <http://fftw.org> deserves mention as particularly efficient, although it is a large library and a bit complex to use. Note that some simpler implementations require that N be a power of two, failing or using a slow $O(N^2)$ algorithm if the requirement is not fulfilled. Also pay attention to (i) the difference between ‘inverse’ and ‘backward’ DFTs (the latter lacks the $1/N$ factor), (ii) how the routine expects the data to be placed in the input array, (iii) how it is returned, and (iv) whether the transform is done ‘in place’ (i.e. overwriting the original data) or not. If the routine is a ‘complex FFT’ it will expect complex input data (so that for real sequences you will have to set to zero the imaginary part of the a_i), while if it is a ‘real FFT’ routine it will typically arrange (‘pack’) the output data in some way, making use of the discrete equivalent of the $\tilde{A}(-\omega) = A^*(\omega)$ symmetry so as to return N real numbers instead of $2N$. This affects the way one must compute the squared modulus (lines 11–12). For example, for the packing used by the FFTW real routines, lines 11–12 translate to (in C)

```

b[0]*=b[0];
for (int i=1; i<N; i++) {
  b[i] = b[i]*b[i] + b[2*N-i]*b[2*N-i];
  b[2*N-i] = 0;
}
b[N]*=b[N];

```

ORCID iDs

Tomás S Grigera  <https://orcid.org/0000-0002-3165-4838>

References

- Allen M P and Tildesley D J 1987 *Computer Simulation of Liquids* (Oxford: Clarendon)
- Amit D J and Martin-Mayor V 2006 *Field Theory, the Renormalization Group, and Critical Phenomena: Graphs to Computers* 3rd edn (Singapore: World Scientific)
- Ashcroft N W and Mermin N D 1976 *Solid State Physics* (Fort Worth: College Publishers)
- Attanasi A *et al* 2014 *Phys. Rev. Lett.* **113** 238102
- Bak P 1996 *How Nature Works: The Science of Self-Organized Criticality* (New York: Springer)
- Barber M N 1983 *Phase Transitions and Critical Phenomena* vol 8 ed C Domb and J L Lebowitz (London: Academic) pp 145–266
- Beggs J M 2008 *Phil. Trans. R. Soc. A* **366** 329
- Beggs J M and Plenz D 2003 *J. Neurosci.* **23** 11167
- Berthier L and Biroli G 2011 *Rev. Mod. Phys.* **83** 587
- Binder K 1981 *Z. Phys. B* **43** 119
- Binney J J, Dowrick N J, Fisher A J and Newman M E J 1992 *The Theory of Critical Phenomena: An Introduction to the Renormalization Group* (Oxford: Clarendon)
- Biroli G, Bouchaud J-P, Cavagna A, Grigera T S and Verrocchio P 2008 *Nat. Phys.* **4** 771
- Biroli G and Bouchaud J-P 2012 *Structural Glasses and Supercooled Liquids: Theory, Experiments and Applications* ed P G Wolynes and V Lubchenko (New York: Wiley) pp 31–113
- Caracciolo S, Edwards R G, Pelissetto A and Sokal A D 1993 *Nucl. Phys. B* **403** 475
- Cardy J (ed) 1988 *Finite-Size Scaling* (Amsterdam: North-Holland) number 2 in Current Physics Sources and Comments
- Cardy J 1996 *Scaling and Renormalization in Statistical Physics* (Cambridge: Cambridge University Press)
- Cavagna A, Cimarelli A, Giardina I, Parisi G, Santagati R, Stefanini F and Viale M 2010a *Proc. Natl Acad. Sci.* **107** 11865
- Cavagna A, Conti D, Creato C, Del Castello L, Giardina I, Grigera T S, Melillo S, Parisi L and Viale M 2017 *Nat. Phys.* **13** 914
- Cavagna A, Di Carlo L, Giardina I, Grandinetti L, Grigera T S and Pisegna G 2019 *Phys. Rev. Lett.* **123** 268001
- Cavagna A, Giardina I and Grigera T S 2018 *Phys. Rep.* **728** 1–72
- Cavagna A, Giardina I, Orlandi A, Parisi G and Procaccini A 2008 *Anim. Behav.* **76** 237
- Cavagna A, Grigera T S and Verrocchio P 2010b *J. Stat. Mech.* **P10001**
- Cavagna A, Grigera T S and Verrocchio P 2012 *J. Chem. Phys.* **136** 204502
- Chialvo D R 2010 *Nat. Phys.* **6** 744
- Chialvo D R 2018 *Acta Phys. Pol. B* **49** 1955
- Chialvo D R and Bak P 1999 *Neuroscience* **90** 1137
- Chialvo D R, Cannas S A, Grigera T S, Martin D A and Plenz D 2020 *Sci. Rep.* **10** 1
- Cooper F, Freedman B and Preston D 1982 *Nucl. Phys. B* **210** 210
- Cugliandolo L F 2004 *Slow Relaxations and Nonequilibrium Dynamics in Condensed Matter* ed J-L Barrat, M Feigelman and J Kurchan (Berlin: Springer) number LXXVII in Les Houches Summer School p 367
- De Palo G, Yi D and Endres R G 2017 *PLoS Biol.* **15** e1002602
- Deco G and Jirsa V K 2012 *J. Neurosci.* **32** 3366
- Dombrowski C, Cisneros L, Chatkaew S, Goldstein R E and Kessler J O 2004 *Phys. Rev. Lett.* **93** 098103
- Duhamel P and Vetterli M 1990 *Signal Process.* **19** 259
- Dyson F 2004 *Nature* **427** 297
- Edelsbrunner H and Mücke E P 1994 *ACM Trans. Graph.* **13** 43
- Emiliani V, Cohen A E, Deisseroth K and Häusser M 2015 *J. Neurosci.* **35** 13917
- Expert P, Lambiotte R, Chialvo D R, Christensen K, Jensen H J, Sharp D J and Turkheimer F 2011 *J. R. Soc., Interface* **8** 472
- Fernández L A and Martín-Mayor V 2015 *Phys. Rev. B* **91** 174202
- Fraiman D and Chialvo D R 2012 *Front. Physiol.* **3** 307
- Frijo M and Johnson S G 2005 *Proc. IEEE* **93** 216
- Ginelli F, Peruani F, Pillot M-H, Chaté H, Theraulaz G and Bon R 2015 *Proc. Natl Acad. Sci. USA* **112** 12729
- Goldenfeld N 1992 *Lectures on Phase Transitions and the Renormalization Group* (Reading, Massachusetts: Perseus Books)
- Greenberg J M and Hastings S P 1978 *SIAM J. Appl. Math.* **34** 515
- Haimovici A, Tagliazucchi E, Balenzuela P and Chialvo D R 2013 *Phys. Rev. Lett.* **110** 178101
- Halperin B I and Hohenberg P C 1969 *Phys. Rev.* **177** 952
- Hanisch K H 1984 *Ser. Stat.* **15** 409
- Hansen J-P and McDonald I R 2005 *Theory of Simple Liquids* 3rd edn (New York: Academic)
- Hohenberg P C and Halperin B I 1977 *Rev. Mod. Phys.* **49** 435
- Honerkamp-Smith A R, Veatch S L and Keller S L 2009 *Biochim. Biophys. Acta Biomembr.* **1788** 53
- Itzykson C and Drouffe J-M 1989a *Statistical Field Theory* 1st edn vol 2 (Cambridge: Cambridge University Press)
- Itzykson C and Drouffe J-M 1989b *Statistical Field Theory* 1st edn vol 1 (Cambridge: Cambridge University Press)
- Jensen H J 2021 *J. Phys. Complex.* **2** 032002
- Kaneko K 2006 *Life: An Introduction to Complex Systems Biology* (Berlin: Springer)
- Kob W 2004 *Slow Relaxations and Nonequilibrium Dynamics in Condensed Matter* ed J-L Barrat, M Feigelman and J Kurchan (Berlin: Springer) number LXXVII in Les Houches Summer School
- Korchinski D J, Orlandi J G, Son S-W and Davidsen J 2021 *Phys. Rev. X* **11** 021059
- Kubo R, Toda M and Hashitsume N 1998 *Statistical Physics II* (Berlin: Springer)
- Le Bellac M 1991 *Quantum and Statistical Field Theory* (Oxford: Clarendon)
- Lindsey C P and Patterson G D 1980 *J. Chem. Phys.* **73** 3348
- Lubchenko V and Wolynes P G 2007 *Annu. Rev. Phys. Chem.* **58** 235
- Martin D A, Ribeiro T L, Cannas S A, Grigera T S, Plenz D and Chialvo D R 2021 *Sci. Rep.* **11** 15937
- Mora T and Bialek W 2011 *J. Stat. Phys.* **144** 268

- Mora T, Walczak A M, Bialek W and Callan C G 2010 *Proc. Natl Acad. Sci.* **107** 5405
- Münel C, Heermann D W, Adler J, Gofman M and Stauffer D 1993 *Physica A* **193** 540
- Muñoz M A 2018 *Rev. Mod. Phys.* **90** 031001
- Newman M E J and Barkema G 1999 *Monte Carlo Methods in Statistical Physics* (Oxford: Oxford University Press)
- Ódor G 2016 *Phys. Rev. E* **94** 062411
- Press W H, Flannery B P, Teukolsky S A and Vetterling W T 1992 *Numerical Recipes in C: The Art of Scientific Computing* 2nd edn (Cambridge: Cambridge University Press)
- Priestley M B 1981 *Spectral Analysis and Time Series* (New York: Academic)
- Privman V (ed) 1990 *Finite Size Scaling And Numerical Simulation of Statistical Systems* (Singapore: World Scientific)
- Ribeiro T L, Yu S, Martin D A, Winkowski D, Kanold P, Chialvo D R and Plenz D 2020 bioRxiv Preprint <https://doi.org/10.1101/2020.07.01.182014>
- Schneidman E, Berry M J, Segev R and Bialek W 2006 *Nature* **440** 1007
- Shew W L, Yang H, Petermann T, Roy R and Plenz D 2009 *J. Neurosci.* **29** 15595
- Sokal A D 1997 *Functional Integration: Basics and Applications (1996 Cargèse School)* ed C DeWitt-Morette, P Cartier and A Folacci (New York: Plenum)
- Sornette D 2006 *Critical Phenomena in the Natural Sciences* 2nd edn (Berlin: Springer)
- Tagliazucchi E, Balenzuela P, Fraiman D and Chialvo D R 2012 *Front. Physiol.* **3** 15
- Tang Q-Y, Zhang Y-Y, Wang J, Wang W and Chialvo D R 2017 *Phys. Rev. Lett.* **118** 088102
- Tauber U C 2014 *Critical Dynamics* (Cambridge: Cambridge University Press)
- Tkačik G, Mora T, Marre O, Amodei D, Palmer S E, Berry M J and Bialek W 2015 *Proc. Natl Acad. Sci.* **112** 11508
- Tkačik G, Schneidman E, Berry M J II and Bialek W 2006 arXiv:q-Bio0611072
- Urbina C E V, Cannas S A and Gleiser P M 2020 arXiv:2011.09057
- van Kampen N G 2007 *Stochastic Processes in Physics and Chemistry* 3rd edn (Amsterdam: Elsevier)
- Veitch S L, Soubias O, Keller S L and Gawrisch K 2007 *Proc. Natl Acad. Sci.* **104** 17650
- Vicsek T, Czirók A, Ben-Jacob E, Cohen I and Shochet O 1995 *Phys. Rev. Lett.* **75** 1226
- Villegas P, Cavagna A, Cencini M, Fort H and Grigera T S 2021 *R. Soc. Open Sci.* **8** 202200
- Watts D J and Strogatz S H 1998 *Nature* **393** 440
- Wilson K and Kogut J 1974 *Phys. Rep.* **12** 75
- Zarepour M, Perotti J I, Billoni O V, Chialvo D R and Cannas S A 2019 *Phys. Rev. E* **100** 052138
- Zhang H P, Be'er A, Florin E L and Swinney H L 2010 *Proc. Natl Acad. Sci.* **107** 13626