Persistent homology analysis of phase transitions

Persistent homology analysis of phase transitions

Irene Donato,[*] Matteo Gori,[†] and Marco Pettini[‡]
Aix-Marseille University, CNRS Centre de Physique Théorique UMR 7332, Campus de Luminy, Case 907, 13288 Marseille Cedex 09, France

Giovanni Petri[§]
ISI Foundation, Turin, Italy

Sarah De Nigris[‖]
NaXys, Département de Mathématique, Université de Namur, 8 repart de la Vierge, 5000 Namur, Belgium

Roberto Franzosi[¶]
Qstar, Istituto Nazionale di Ottica, largo E. Fermi 6, 50125 Firenze, Italy

Francesco Vaccarino[#]
Dipartimento di Scienze Matematiche "G.L.Lagrange", Politecnico di Torino, C.so Duca degli Abruzzi 24, Italy and ISI Foundation, Turin, Italy

(Received 17 June 2015; revised manuscript received 25 March 2016; published 20 May 2016)

Persistent homology analysis, a recently developed computational method in algebraic topology, is applied to the study of the phase transitions undergone by the so-called mean-field *XY* model and by the ϕ^4 lattice model, respectively. For both models the relationship between phase transitions and the topological properties of certain submanifolds of configuration space are exactly known. It turns out that these *a priori* known facts are clearly retrieved by persistent homology analysis of dynamically sampled submanifolds of configuration space.

DOI: 10.1103/PhysRevE.93.052138

I. INTRODUCTION

Topological methods lie at the base of many successful physical theories [1], with fields of applications ranging from dynamical systems and quantum computation to the theory of phase transitions and topological field theories. In recent years, the possibility has been investigated [2] that at least for a broad class of physical systems the deep origin of phase transitions is a major topological change of some submanifolds of phase space or, equivalently, of configuration space. The central idea is that the singular energy dependence displayed by the thermodynamic observables at a phase transition is the shadow of such major topological change.

This new point of view about the deep origin of phase transitions was originally proposed for theoretical reasons, in fact, after the Yang-Lee theorem the mathematical description of phase transitions requires the thermodynamic limit ($N \to \infty$) in order to break the analyticity of thermodynamic observables. However, phase transition phenomena occur in nature as dramatic qualitative changes of some physical property also very far from thermodynamic limit. Let us think of Bose-Einstein condensation, of Dicke's superradiance in microlasers, of superconductive transitions in very small metallic objects, of the filament-globule transition in homopolymers, of the folding transition in proteins, of a microscopic snowflake melting into a droplet of liquid water, to mention just some examples.

The question was: Can we think of a different mathematical approach unifying the description of phase transitions in finite, small N systems with the standard description resorting to the thermodynamic limit dogma? At least for a broad class of physical potentials the answer was in the affirmative as can be seen in Refs. [2] and [3,4]. However, in some sense similarly to the Yang-Lee theory for which analytically finding the zeros of the grand partition function is in practice possible only for a few models (essentially given by the Lee-Yang "circle theorem"), also the topological approach suffers from computational difficulties, and analytic topological information can be obtained only for a very few models. Also the direct numerical measurement of topological properties of the configuration space of physical systems faces serious computational issues because of the high dimensionality of the associated manifolds. The idea that some of the mentioned computational obstacles could be overcome comes from the observation of the existence of new computational tools in the fields of discrete geometry and topology. These new methods have already been developed for analyzing data in high-dimensional spaces [5]. Hence, we expect that they could be useful to investigate topological changes also in physical configuration spaces by identifying their homology from random samples.

In the present paper we resort to persistent homology analysis. Persistent homology [6–8], a particular sampling-based technique from algebraic topology, was originally introduced in 2002 [9] by Edelsbrunner *et al.* with the aim of extracting coarse topological information from high-dimensional data sets [5]. In a nutshell, while homology

[*]irene.irened@gmail.com
[†]gori6matteo@gmail.com
[‡]pettini@cpt.univ-mrs.fr
[§]giovanni.petri@isi.it
[‖]denigris.sarah@gmail.com
[¶]bob.franzosi@gmail.com
[#]francesco.vaccarino@gmail.com

detects the connected components, tunnels, voids of a given topological space, persistent homology computes multiscale homological features obtained from a discrete sample of a topological space X by foliating it appropriately. Hitherto, the study of persistent homology has already proved useful in various fields such as biological and medical data analysis, neuroscience [10], sensor network coverage problems [11], to quote just a few of them.

Here persistent homology is applied to the study of equilibrium phase transitions. Two models are considered for which we rigorously know what to expect: the so-called mean-field XY model (MFXY) and the classical lattice ϕ^4 model. For the MFXY model both the thermodynamics and the configuration space topology are exactly known, whence the topological origin of phase transition is rigorously ascertained; while for the ϕ^4 model it is analytically known that the phase transition does not correspond to any topology change in configuration space at any finite N (see Sec. II B for a discussion on this model).

The benchmarking so performed gave sharp and unambiguous results in the good direction. This could open new interesting perspectives for practical applications of the above-mentioned topological theory of phase transitions.

II. PHASE TRANSITIONS AND TOPOLOGY

It is well know that the unbounded growth with N of certain thermodynamic quantities, eventually leading to singularities in the $N \to \infty$ limit, is the hallmark of an equilibrium phase transition. Apart from several studies on specific models [2], two theorems state that these unbounded growths are necessarily due to appropriate topological transitions in configuration space [3,4]. The following exact formula

$$\begin{aligned} S_N(v) &= (k_B/N)\log\left[\int_{M_v} d^N q\right] \\ &= \frac{k_B}{N}\log\left[\mathrm{vol}\left[M_v \setminus \bigcup_{i=1}^{\mathcal{N}(v)} \Gamma\left(x_c^{(i)}\right)\right] \right. \\ &\quad \left. + \sum_{i=0}^{N} w_i\, \mu_i(M_v) + \mathcal{R}\right], \end{aligned} \tag{1}$$

makes explicit the relation between thermodynamics and topology, where S is the configurational entropy, v is the potential energy per degree of freedom, and the $\mu_i(M_v)$ are the Morse indexes (in one-to-one correspondence with topology changes) of the submanifolds $\{M_v = V_N^{-1}((-\infty, v])\}_{v\in\mathbb{R}}$ of configuration space; in square brackets: the first term is the result of the excision of certain neighborhoods of the critical points of the interaction potential from M_v; the second term is a weighted sum of the Morse indexes, and the third term $\mathcal{R}$ is a smooth function of N and v. It is evident that sharp changes in the potential energy pattern of at least some of the $\mu_i(M_v)$ (thus of the way topology changes with v) affect $S(v)$ and its derivatives. It can be proved that the occurrence of phase transitions necessarily stems from this topological part of thermodynamic entropy [3,4].

A. Mean-field XY model

The mean-field XY model is defined by the Hamiltonian [12,13]

$$\mathcal{H}(p,\varphi) = \sum_{i=1}^{N}\frac{p_i^2}{2} + \frac{J}{2N}\sum_{i,j=1}^{N}[1-\cos(\varphi_i-\varphi_j)] - h\sum_{i=1}^{N}\cos\varphi_i. \tag{2}$$

Here $\varphi_i \in [0,2\pi]$ is the rotation angle of the ith rotator and h is an external field. Defining at each site i a classical spin vector $\mathbf{m}_i = (\cos\varphi_i, \sin\varphi_i)$, the model describes a planar (XY) Heisenberg system with interactions of equal strength among all the spins. We consider the ferromagnetic case $J = 1$. The equilibrium statistical mechanics of this system is exactly described, in the thermodynamic limit, by mean-field theory. In the limit $h \to 0$, the system has a continuous phase transition, with classical critical exponents, at the critical temperature $T_c = 1/2$, or at the critical energy density $E_c/N = 3/4$ [12].

The entire configuration space M of the model is an N-dimensional torus, parametrized by N angles. The submanifolds $M_v \subset M$ are defined by

$$\begin{aligned} M_v &= \mathcal{V}^{-1}(-\infty, v] \\ &= \{(\varphi_1,\ldots,\varphi_N) \in M : \mathcal{V}(\varphi_1,\ldots,\varphi_N) \leqslant v\}, \end{aligned} \tag{3}$$

i.e., defined by the constraint that the potential energy per particle $\mathcal{V} = V/N$ does not exceed a given value v.

Morse theory [14] states that topology changes of the M_v occur in correspondence with critical points of $\mathcal{V}$, i.e., those points where $\nabla\mathcal{V} = 0$. This implies [2] that there are no topological changes for $\mathcal{V} > 1/2 + h^2/2$, i.e., all the M_v with $\mathcal{V} > 1/2 + h^2/2$ are diffeomorphic to the whole M.

The Euler characteristic, a topological invariant of the manifolds M_v, which is exactly computed in Refs. [15,16], is defined by

$$\chi(M_v) = \sum_{k=0}^{N}(-1)^k \mu_k(M_v), \tag{4}$$

where the Morse number μ_k is the number of critical points of $\mathcal{V}$ that have index k [14].

After a monotonic growth with v, a sharp, discontinuous jump to zero of $\chi(M_v)$ is found at the phase transition point, that is, at $v_c = 1/2 + 0^+$. However, as already shown in Refs. [15,16], it is just this major topological change occurring at v_c that is related to the thermodynamic phase transition of the mean-field XY model.

B. ϕ^4 model

The lattice ϕ^4 model is defined by the Hamiltonian

$$\mathcal{H}(p,\varphi) = \sum_{\mathbf{i}\in\mathbb{Z}^d}\left[\frac{p_{\mathbf{i}}^2}{2} + \frac{J}{2}\sum_{\mu=1}^{d}(\varphi_{\mathbf{i}+\mathbf{e}_\mu}-\varphi_{\mathbf{i}})^2 - \frac{1}{2}m^2\varphi_{\mathbf{i}}^2 + \frac{\lambda}{4}\varphi_{\mathbf{i}}^4\right], \tag{5}$$

where $\mathbf{e}_\mu$ is the unit vector in the μth direction of the d-dimensional lattice. At equilibrium and for $d \geqslant 2$, this

model—representing a set of linearly coupled nonlinear oscillators—shows a second-order phase transition with nonzero critical temperature. This phase transition is due to a spontaneous breaking of the discrete $O(1)$, or $\mathbb{Z}_2$, symmetry.

Recently, this model has been proposed as a counterexample of the topological theory of phase transitions [17]. In fact, the phase transition of the $d \geqslant 2$ lattice ϕ^4 model occurs at a critical value v_c of the potential energy density, which belongs to a broad interval of v values void of critical points of the potential function. This means that the $\{\Sigma^N_{v<v_c}\}_{v\in\mathbb{R}}$ are diffeomorphic to the $\{\Sigma^N_{v>v_c}\}_{v\in\mathbb{R}}$ so that no topological change seems to correspond to the phase transition. Since then, it is commonly believed that this result undermines the value of the topological theory [18].

However, the following heuristic reasoning could suggest a different explanation of the situation. Both the microcanonical caloric curve (temperature versus energy) and the law of divergence of the specific heat of the ϕ^4 model show markedly softer transition patterns [19,20] than those shown by other models undergoing a continuous phase transition, such as the mean-field XY model and the p-trigonometric model [2]. For these latter models it has been analytically proved that major topology changes correspond to the critical energy value of the phase transition [15,21,22]. We can thus wonder whether, in some sense to be specified, a softer topology change in phase space and in configuration space of the ϕ^4 model can be at the roots of its phase transition. In the absence of critical points, and in presence of a discrete ($\mathbb{Z}_2$) symmetry breaking, one is led to think of a corresponding breaking of connectedness of phase and configuration spaces. This could be defined a soft topology change because it would be an asymptotic change of the number of phase-space connected components: from one component in the symmetric phase (vanishing magnetization) to two disjoint components in the broken-symmetry phase (of positive and negative magnetization, respectively).

This heuristic argument has inspired a detailed analysis, reported in Ref. [23], based on numerical simulations of the equations of motion derived from the Hamiltonian (5). Along the numerical trajectories in phase space one can follow the time evolution of the order parameter $\mathfrak{M}$ (magnetization)

$$\mathfrak{M}(t) = \frac{1}{N}\sum_{\mathbf{i}} \varphi_{\mathbf{i}}(t). \tag{6}$$

It is found that below the phase transition critical energy density $\varepsilon_c = E_c/N$ the phase-space trajectories are trapped for some average time τ_{tr} in a region of either positive or negative value of $\mathfrak{M}$, then move to another region of opposite sign of $\mathfrak{M}$, and so on, back and forth. The interesting result is that τ_{tr} grows with a power law of N, the number of degrees of freedom, that is: $\tau_{tr} \propto N^{\alpha(\varepsilon)}$ at $\varepsilon < \varepsilon_c$, with an exponent $\alpha = \alpha(\varepsilon)$, which increases by lowering the value of ε. To the contrary, at $\varepsilon > \varepsilon_c$ the trapping time is short (in units of the inverse of the shortest linear frequency of the system) and is a flat function of N.

Let us now denote with $\Sigma^N_E = \mathcal{H}^{-1}(E)$ an N-dimensional constant energy hypersurface of phase space, with $\Phi^H_t : \Sigma^N_E \to \Sigma^N_E$ the Hamiltonian flow, with $\mathcal{M}^+_E \subset \Sigma^N_E$ the set of all the phase-space points for which $\mathfrak{M} > 0$, with $\mathcal{M}^-_E \subset \Sigma^N_E$ the set of all the phase-space points for which $\mathfrak{M} < 0$. By inversely reading the function $\tau_{tr}(N)$ we deduce that: at $\varepsilon < \varepsilon_c$, for any chosen value of $\tau_{tr} > 0$ there is a number of degrees of freedom $N(\tau_{tr})$ such that for any larger value of this number, i.e., for $N > N(\tau_{tr})$, and for times up to $t \in [0,\tau_{tr}]$ we have $\Phi^H_t(\mathcal{M})^+_E = \mathcal{M}^+_E$ and $\Phi^H_t(\mathcal{M})^-_E = \mathcal{M}^-_E$. In other words, the sets $\mathcal{M}^\pm_E$ are invariant sets of the Hamiltonian flow on a time scale τ_{tr} with the remarkable property that $\tau_{tr} \to \infty$ in the limit $N \to \infty$. Hence, any phase-space trajectory originating in one of the two regions $\mathcal{M}^\pm_E$ will remain therein forever in the limit $N \to \infty$. This amounts to an asymptotic breaking of topological transitivity at $\varepsilon < \varepsilon_c$, and this in turn is equivalent to an asymptotic loss of connectedness of the energy level sets Σ^N_E [24]. Of course the asymptotic change of topology of the energy level sets entails also the asymptotic change of topology of the potential level sets Σ^N_v, which are submanifolds of the Σ^N_E. At variance with other models, the topological change of the ϕ^4 model corresponds to an asymptotic loss of diffeomorphicity among the $\Sigma^N_{v<v_c}$ and the $\Sigma^N_{v>v_c}$ occurring in the absence of critical points. Hence, in the opinion of the authors, the ϕ^4 model need no longer be regarded as a counterexample to the theory of [3,4].

This sheds light on the way of reformulating the theorems in Refs. [3,4]. In fact, uniform convergence of the Helmholtz free energy to a twice differentiable thermodynamic limit function (thus the absence of first- and second-order phase transitions) derived under the assumption that the equipotential level sets $\Sigma^N_v = V_N^{-1}(v)$ are diffeomorphic at any finite N must be extended to encompass also asymptotic diffeomorphicity. In so doing the ϕ^4 model need not be regarded as a counterexample to the topological theory.

Notice that, at variance with the MFXY model for which a sharp topological signature of the phase transition shows up at rather small N values, the phase transition of the $d \geqslant 2$ lattice ϕ^4 model is more akin to what is required by the Yang-Lee theory (that is, asymptoticity), even if tackled from the topological point of view. Now, taking advantage of the above-mentioned results in a reverse form, if the phase-space sampling through the Hamiltonian dynamics of the ϕ^4-model is performed at a given $\widetilde{N}$ for a sufficiently long time $t \gg \tilde{\tau}_{tr}(\widetilde{N})$, then also for $E < E_c$ the energy level sets Σ^N_E appear simply connected. In other words, in view of the application of persistent homology analysis, this model is a good candidate for a negative check against the MFXY model.

III. TOPOLOGICAL ANALYSIS

In the following we report on the topological analysis which begins by sampling the configuration space of each system at different energies. Then we apply persistent homology analysis.

A. Samples of the configuration space

We begin by constructing samples of the configuration spaces to be studied. For the MFXY model, this is done by numerically integrating the equations of motion derived from Hamiltonian (2) with the external field set to $h = 0$ for a system of N spins, with N up to 6000. The numerical integration is performed by means of a fifth-order optimal symplectic algorithm [25]. We sampled the configuration

space for the following values of the energy density $\varepsilon = E/N = 0.6, 0.75, 0.88$, that is, below, at, and above the critical energy, respectively. The system is initialized with a Gaussian distribution for both conjugated variables $\{\varphi_i, p_i\}$. The total angular momentum ($P = \sum_i p_i = 0$) is imposed to vanish. Given the initial conditions for the aforementioned energies, the system dynamics is evolved for a $T = 1.26 \times 10^7$ time steps, with an integration step of $\Delta t = 0.05$. With these integration step and the use of a fifth-order symplectic algorithm the relative energy fluctuations were kept at $\Delta E/E \simeq 10^{-9}$. Then 6000 snapshots are uniformly sampled in time after a transient dynamics to equilibrate kinetic and potential energies to their equipartition values.

For the ϕ^4 model we set $J = 1, m^2 = 2, d = 3$, and $\lambda = 0.1$ in the Hamiltonian (5). We consider a three-dimensional (3d) cubic lattice with 8^3 sites, periodic boundary conditions, and an integration time step $\Delta t = 0.05$. With these parameters, the use of a third order symplectic algorithm [26] kept the relative energy fluctuations at $\Delta E/E \simeq 10^{-9}$ (a lower-order algorithm was required in this case with respect to the MFXY model because trigonometric functions are replaced by the polynomial form of the ϕ^4 potential). Then the Hamiltonian dynamics is numerically simulated at two different values of the energy density, that is, $\varepsilon = 25$ well below the transition energy density $\varepsilon_c \simeq 31$ [20], and $\varepsilon = 35$ well above ε_c.

B. Persistent homology

The main idea of persistent homology is to build an increasing sequence of simplicial complexes, called a filtration (see Ref. [7]), from a point cloud, i.e., a set of points embedded in a metric space. We report a detailed mathematical description of persistent homology in the supplementary material and refer the interested reader to Ref. [7]. Here we streamline the topological analysis. The standard way to obtain a simplicial complex from a set of points S is to construct its ρ-Rips-Vietoris complex [7], an abstract simplicial complex that can be defined on any set of points in a given metric space $\mathcal{M}$. The n simplices of the ρ-Rips-Vietoris complex are determined by subsets of $n+1$ points $\{p_0, \ldots, p_n\}$ such that $D(p_i, \rho) \cap D(p_j, \rho) \neq \emptyset$ for all $i \neq j \in \{0, \ldots, n\}$, where $D(p, \rho)$ is ball of radius ρ centered at p. Persistent homology is a powerful instrument in that it does not select just an ρ value, but rather studies how the homology of the space, and in particular of the ρ-Rips-Vietoris complexes, changes as ρ varies. As ρ is increased, simplexes are added in the ρ-Rips-Vietoris simplicial complex. A new simplicial complex is added to the filtration only when a new simplex is born along the (continuous) parameter ρ, i.e., the ρ-Rips-Vietoris complex has changed. Thus, the filtration is discrete: it can be indexed by integers, useful to characterize the topological features of the space.

C. Simplicial complexes in configuration space

In most applications of persistent homology, the parameter ρ is taken to represent the Euclidean distance between points in S. In the case of physical configuration spaces we replace it by a Riemannian one. In fact, the configuration space M of a standard Hamiltonian systems (that is with quadratic kinetic energy) equipped with the Jacobi metric [2], is a complete Riemannian manifold, which means that given any two points there exists a length-minimizing geodesic connecting them (Hopf-Rinow theorem [27]). Of course this is also the case of the mean-field XY and ϕ^4 models, thus the distance among two points P_1 and P_2 in M is:

$$d(P_1, P_2) = \int_{P_1}^{P_2} \left([E - V(\varphi_1, \ldots, \varphi_N)] \sum_{k=1}^{N} (d\varphi^k)^2 \right)^{\frac{1}{2}}. \quad (7)$$

In other words, computing this distance requires solving the equations of motion with assigned initial and final conditions. In practice this is computationally very heavy. We therefore take advantage of the robustness of topological information with respect to metrical deformations and observe that the integral contains a nonconstant factor multiplying the Euclidean arc length. We then choose to approximate $d(P_1, P_2)$ by replacing the factor by its mean among the initial and final values:

$$d(p_1, p_2) = \frac{1}{2}[\sqrt{E - V(p_1)} + \sqrt{E - V(p_2)}] d_{\text{eucl}}(p_1, p_2) \quad (8)$$

$$d_{\text{eucl}}(p_1, p_2) = \sqrt{\sum_{k=1}^{N} [\varphi^k(p_2) - \varphi^k(p_1)]^2}. \quad (9)$$

An important computational issue lies in the size of the produced simplicial complexes. Indeed, already for a sample of the configuration space S with cardinality $N = 6000$ points, the set of complexes will contain a huge number of simplices hindering efficient computation, since the number of all simplices for all dimensions up to $N-1$ scales as number of subsets of N, that is 2^N. So, we first restrict ourselves to the study of the first two homology groups, H_0 and H_1, which allows us to consider only simplices up to dimension 2 and then adopt a subsampling strategy, which allows us to reduce the dimension of the problem by choosing a representative subset of points $L \subset S$ without losing important topological features of the configuration space. The subsampling is based on a suitable selection of landmark points called sequential maxmin [28,29]. In sequential maxmin, the first landmark is picked randomly from S. Inductively, if L_{i-1} is the set of the first $i-1$ landmarks, then let the ith landmark be the point of S, which maximizes the distance (8) from all the points of L_{i-1}. Since the starting node is chosen at random, the resulting L subsets will change if the algorithm is iterated. In our case, this allows us to perform a bootstraplike procedure, by repeatedly subsampling the full point clouds and then aggregating the homological signatures detected. The results we present are obtained from 20 different subsamples, each containing 300 points.

IV. RESULTS

Persistent homology computes the generators of topological features (homology groups) persisting across different scales and assigns them birth and death values related to their points of appearance and disappearance along the filtration. That is, when the radius ρ of the balls varies, for any persistent

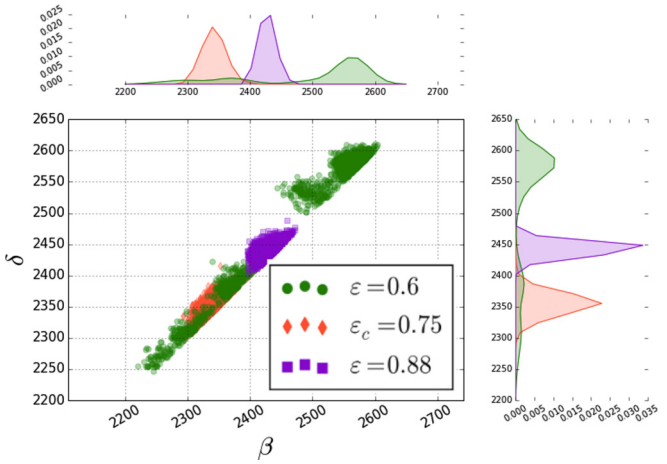

FIG. 1. Persistence diagram for the MFXY model. Persistence distributions of H_1 generators below ($\varepsilon = 0.6$), at ($\varepsilon_c = 0.75$), and above ($\varepsilon = 0.88$) the phase transition.

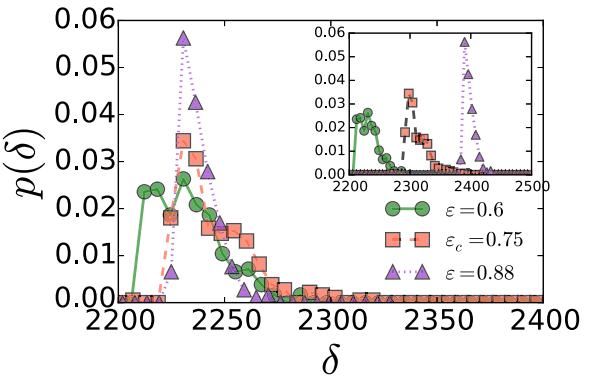

FIG. 3. Homological features of the MFXY model. Raw (inset) and rescaled (main plot) distributions of deaths for the generators of the first homology group H_0. Note that the width and shape of the distributions change across the transition, becoming more and more narrow as the energy is increased.

generator g (see the Appendix for the formal definition) we have the value of the parameter ρ of the filtration where g first appears (birth index indicated by β_g) and the value where it disappears (death index indicated by δ_g). In this way, connected components, one-dimensional cycles, three-dimensional voids, and similar higher-order structures of the topological space $\mathcal{M}$ acquire a weight proportional to the length of their persistence interval, $\pi_g = \delta_g - \beta_g$. Note that for H_0, $\pi_g = \delta_g$, because all (dis)connected components are already present at the beginning. For higher-order homology groups H_k the generators can instead appear and disappear freely along the filtration.

In Figs. 1 and 2 the basic descriptors of persistent homology, that is, persistence diagrams, are displayed for the H_1 generators of the MFXY model and of the ϕ^4 model, respectively.

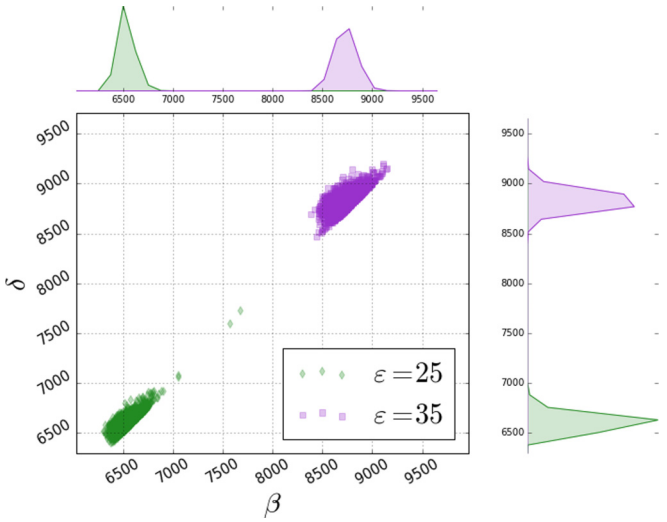

FIG. 2. Persistence diagram for the ϕ^4 model. Persistence distributions of H_1 generators below ($\varepsilon = 25$) and above ($\varepsilon = 35$) the phase transition, occurring at the critical energy density $\varepsilon_c \simeq 31$.

Usually one considers important topological features to be those associated with generators of H_n such that their π_g is large with respect to some meaningful length. In our case we do not have a given reference scale. We can however compare the results obtained at energies below and above the transition energy in order to look for topological signatures of a phase transition. We show the distributions of δ_g for the H_0 generators of the MFXY model (Fig. 3) and of the ϕ^4 model (Fig. 4). In the former case, as the energy is increased, the peak of the distribution δ_g of H_0 becomes progressively narrower and centered at larger ρ values. To the contrary, in the latter case the peak of the distribution shifts to larger ρ values at higher energies, but it does not broaden.

In order to show that this behavior is genuinely due to topological features and not due to the different geometrical sizes of the point clouds, we take the point cloud at the lowest energy and affinely rescale the point clouds at higher energies

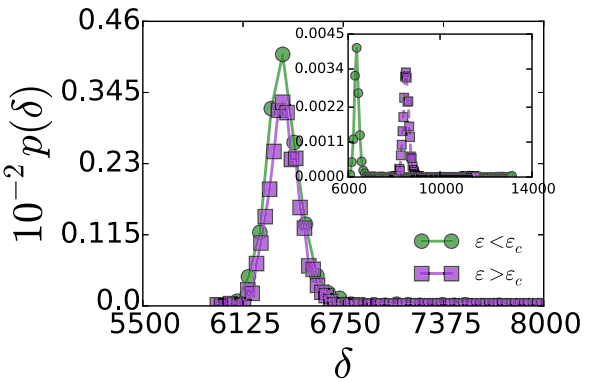

FIG. 4. Homological features of the ϕ^4 model. Raw (inset) and rescaled (main plot) distributions of deaths for the generators of the first homology group H_0. At variance with the MFXY model, here there is no appreciable change in the width and shape of the distributions across the transition. The green points refer to $\varepsilon = 25 < \varepsilon_c$. The velvet points refer to $\varepsilon = 35 > \varepsilon_c$.

as to make them comparable i.e.

$$d_{ij}(\varepsilon) \to \frac{\langle d(\varepsilon_0)\rangle}{\langle d(\varepsilon)\rangle} d_{ij}(\varepsilon), \tag{10}$$

where $d_{ij}(\varepsilon)$ is the distance between points i and j for the point cloud at energy density ε. In this way, we can meaningfully compare the persistences of generators belonging to clouds of different size. Below the transition of the MFXY model, the distribution of the H_0 persistences of configuration space covers more scales than it does at and above the transition energy, respectively. This broader distribution means that the corresponding point cloud is heterogeneously distributed in the embedding space $\mathcal{M}$ compared to the distributions, definitely more homogeneous, in the other two cases. No variation of the peak widths of the H_0 persistence distributions is observed in the case of the ϕ^4 model.

Figures 3 and 4 display the raw (inset) and rescaled (main plot) distributions of deaths for the generators of the first homology group H_0. The rescaling is necessary to make the point clouds, sampled at different energies, comparable. In fact, the death and birth indexes are the values of the radius of the balls where the generators appear and disappear. Thus, without the rescaling, β_g and π_g would reflect the size of the underlying manifold. Note that for the MFXY model the width and shape of the distributions change across the transition, becoming more and more narrow as the energy is increased, while there is no appreciable change in the ϕ^4 case. The different topological signatures highlight the presence of a topological change in the case of the MFXY model, that is absent in the ϕ^4 model. In Figs. 5 and 6 the distributions of persistences for the generators of the homology group H_1 confirm what is found for H_0. In this case the difference in functional forms for the H_1 persistence distribution below and above the MFXY transition is even clearer, while, again, we find no differences for the ϕ^4 model.

Now let us comment about the hollowness detected by the H_1 homology group. For what concerns the MFXY model, below the phase transition energy, the H_1 persistence distribution displays a long tail, which disappears at and above the transition (Fig. 5). We observe that the three sets of points superpose for values of π less then approximately 25. This

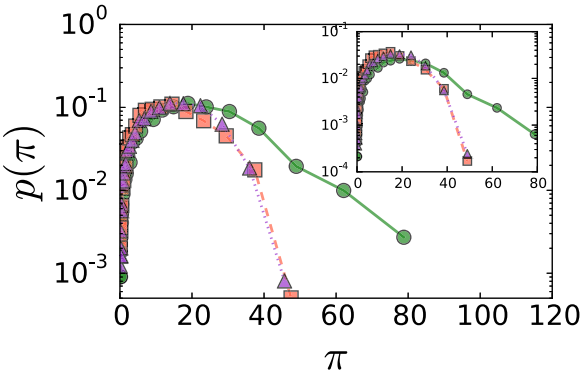

FIG. 5. Distributions of persistences for the generators of the homology group H_1 in the case of the MFXY model. In this case the difference in functional forms for the H_1 persistence distribution below and above the transition is even clearer.

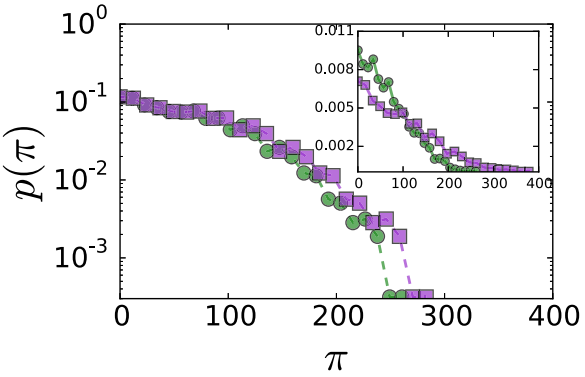

FIG. 6. Distributions of persistences for the generators of the homology group H_1 in the case of the ϕ^4 model. In this case no difference is found in functional form for the H_1 persistence distributions below and above the transition.

range of π values, in the present context, can be attributed to what is commonly referred to as noise, whereas larger π values are usually considered as bringing about meaningful topological information. Thus, the stronger persistence of meaningful cycles, which corresponds to the long tail observed below the phase transition point of the MFXY model, certainly probes a change of shape of configuration space. Additionally, this change of shape can be interpreted as the signature of a change of the dimension of high-order homology groups.

Let us remark that the performed samplings of configuration space submanifolds are definitely sparse and they could not be other then sparse had we taken billions of points. Not to speak of the huge total number of simplexes, growing as 2^N with N the number of sample points. This notwithstanding, the results shown in Fig. 5 clearly tell us that the MFXY phase transition corresponds to a change of the topology of the configuration space submanifolds, in perfect agreement with the available theoretical knowledge. The same concordance is found in the case of the ϕ^4 model where we see that the difference in H_1 persistences disappears, in perfect agreement with *a priori* known absence of topological changes – at finite N – of the

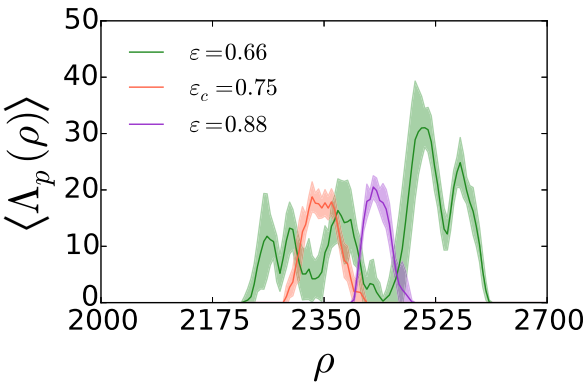

FIG. 7. Average persistence landscape of the H_1 homology for the MFXY model. Λ_p is the average function (see text) reported as a function of the radius ρ of the balls used to construct the Rips-Vietoris simplicial complex. The shadows around solid lines are 95% confidence band.

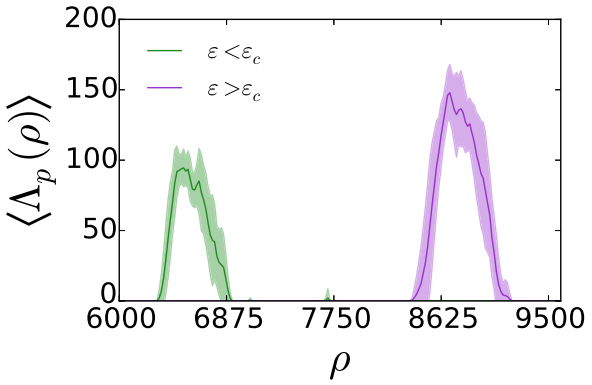

FIG. 8. Average persistence landscape of the H_1 homology for the ϕ^4 model. Λ_p is the average function (see text) reported as a function of the radius ρ of the balls used to construct the Rips-Vietoris simplicial complex. The shadows around solid lines are 95% confidence band.

underlying configuration space in correspondence with the phase transition.

Finally, in Figs. 7 and 8 we show the outcomes of a different method of getting insight to the shape of data obtained by sampling the configuration space of the MFXY and ϕ^4 models, respectively. This is the so-called persistence landscape, which combines the main tool of persistent homology method, that is, persistence diagram, with statistics [30]. With respect to the barcode or persistence diagram this descriptor has the technical advantage of being a function, thus allowing the use of the vector space structure of its underlying function space to apply the theory of random variables with values in this space. Theory and details of this method can be found in Refs. [30] and [31]. In practice, one proceeds by computing the H_1 homology for a subsample of the original data set, then one associates to each generator a symmetric tent-shaped function peaking in the middle of the persistence interval of the corresponding generators and finally one considers the envelope of the functions defined in this way over all the generators. Informally, one can think of the persistence landscape as the envelope of the $\pi/4$ clockwise rotated persistence diagram (operation that can be given a proper mathematical definition) thus associating a curve $\Lambda_p(\rho)$ to each persistence diagram. In our case, we iterated this procedure for the different subsamples, in our case 20 subsamples, obtaining the curve $\langle \Lambda_p(\rho) \rangle$ averaged over the samples. Each curve reported in Fig. 7 reports the results for different energy values: below, at, and above the phase transition point. A marked difference is again obtained above and below the phase transition in the case of the MFXY model, and no relevant difference between the patterns below and above the phase transition in the case of the ϕ^4 model, apart from a meaningless translation.

V. CONCLUDING REMARKS

The results reported for each model in the figures shown in the preceding section, and especially the comparison with those reported in Figs. 5, 6, 7, and 8 are strongly supportive of the validity of the application of persistent homology to probe major topological changes in the configuration spaces of physical systems undergoing phase transitions.

Let us remark that the formulation of the topological theory of phase transitions stems from the combined effect of the investigation of the Hamiltonian dynamical counterpart of phase transitions on the one hand, and of the geometrization of Hamiltonian flows seen as geodesic flows on suitably defined Riemannian manifolds on the other hand [2]. In fact, it has been observed that the peculiar dynamical changes occurring at a phase transition correspond to special geometrical changes of the mechanical manifolds. Then it turned out that these special geometrical changes had to be due to more fundamental changes of topological kind. In other words, this theory has deep roots and rather compelling motivations [2]. Moreover, developing this unconventional viewpoint on phase transitions was of prospective interest to tackle phase transition phenomena in finite and small N systems (mesoscopic and nanoscopic systems), in the microcanonical ensemble (especially when this is not equivalent to the canonical ensemble), in the absence of order parameters (for example in gauge models, i.e., with local symmetries), in amorphous and disordered materials, in polymers and proteins, in biophysical systems, in strongly inhomogeneous systems. However, as mentioned in Sec. I, computational difficulties have frustrated these expectations.

Now the results reported in the present work show that persistent homology, by providing handy computational tools (which are presently available as open access software packages), can lend new credit to the prospective practical interest of the topological theory of phase transitions. Additionally, especially, since improvements of the numerical algorithms are continuously underway.

Moreover, this opens many fascinating and challenging questions related with the mentioned necessarily sparse sampling of high-dimensional manifolds. It is not out of place to mention that this situation is reminiscent of Monte Carlo methods, which typically allow efficient estimates of multiple integrals in high-dimensional spaces with very sparse samplings. Monte Carlo methods owe their efficacy to the so-called importance sampling technique, suggesting that further developments in the proposed application of the persistent homology could be found in a somewhat similar direction.

ACKNOWLEDGMENT

This work was supported by the Seventh Framework Programme for Research of the European Commission under FET-Open grant TOPDRIM (Grant No. FP7-ICT-318121).

APPENDIX: SIMPLICIAL COMPLEXES

We can see a simplicial complex X as a set of polyhedrons (convex hulls of linearly independent points: points, lines, triangles, tetrahedra, and higher-dimensional equivalents) in $\mathbb{R}^N$ attached in a good way, i.e., the intersection of two polyhedrons is empty or a face of the two and all the faces of a polyhedron of X is also a polyhedron of X. We can also think of simplicial complexes as abstract sets, with the definition:

Definition 1. An (abstract) simplicial complex is a nonempty family X of finite subsets, called faces, of a vertex set V such that $\sigma \subset \tau \in X$ implies that $\sigma \in X$.

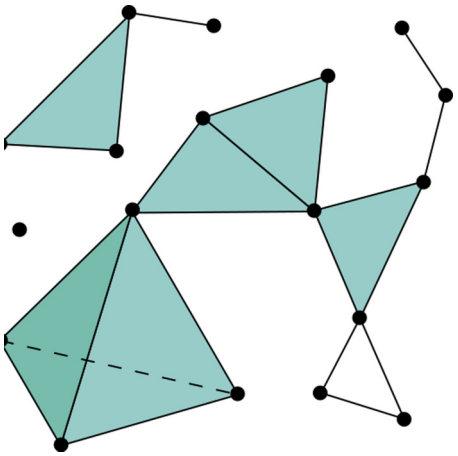

FIG. 9. A graphic representation of a simplicial complex.

We assume that the vertex set is finite and totally ordered. A face of $n+1$ vertices is called n face, denoted by $[p_0,\dots,p_n]$, and n is its dimension. We set, as usual, the dimension of the empty set as -1. The dimension of a simplicial complex is the highest dimension of the faces in the complex.

In Fig. 9 the vertices (full circles) represent 0-simplices, segments joining two vertices represent 1-simplices, triangles represent 2-simplices, and the tetrahedron represents a 3-simplex.

1. Simplicial homology

Let us fix a field k. In the following, by vector space we intend k-vector space. Given a simplicial complex X of dimension d, for any n such that $0 \leqslant n \leqslant d$ consider the vector space $C_n := C_n(X)$ of all the linear combinations of n faces of X with coefficients in k. Elements in C_n are called n chains.

The boundary operators are the linear maps sending an n face to the alternate sum of its $(n-1)$ faces, i.e.,

$$\partial_n : C_n \to C_{n-1} \tag{A1}$$

$$[p_0,\dots,p_n] \to \sum_{j=0}^{n} (-1)^j [p_0,\dots,p_{j-1},p_{j+1},\dots,p_n]. \tag{A2}$$

They share the property $\partial_{n-1}\circ\partial_n = 0$. The null space $\ker\partial_n = \{c \in C_n | \partial_n c = 0\}$, that is, the subspace of C_n containing only the n chains without boundary, is called the vector space of n cycles and denoted by $Z_n := Z_n(X)$, with by convention $Z_0 = \emptyset$. The subspace $\mathrm{Im}\,\partial_{n+1}$ of C_n, is called the vector space of n boundaries and denoted by $B_n := B_n(X)$, with by convention $B_d = \emptyset$. The property $\partial_{n-1}\circ\partial_n = 0$ is then equivalent to $B_n \subseteq Z_n$ for all n.

Definition 2. For $0 \leqslant n \leqslant d$, the nth simplicial homology space of X, with coefficients in k, is the vector space $H_n := H_n(X) := Z_n/B_n$. We denote by $\beta_n := \beta_n(X)$ the dimension of H_n, which is usually called the nth Betti number of X.

Let us see two examples. First, let us consider the simplicial complex X consisting of a triangle $[p_1p_2p_3]$ and all its edges and vertices (i.e., $X = \{[p_1p_2p_3],[p_1p_2],[p_1p_3],[p_2p_3],[p_1],[p_2],[p_3]\}$). The boundary of the 2-simplex $[p_1p_2p_3]$ is

$$\partial_2([p_1p_2p_3]) = [p_2p_3] - [p_1p_3] + [p_1p_2], \tag{A3}$$

which is a one-chain whose boundary is

$$\partial_1([p_2p_3] - [p_1p_3] + [p_1p_2]) = [p_3] - [p_2] + [p_1] - [p_3] + [p_2] - [p_1] = 0. \tag{A4}$$

Therefore, $Z_1 = B_1$ is the vector space generated by $[p_2p_3] - [p_1p_3] + [p_1p_2]$, so $H_1 = \emptyset$ and $\beta_1 = 0$.

After let us consider the simplicial complex X' consisting of all the edges and vertices of the triangle but without the face $[p_1p_2p_3]$ (i.e., $X' = X/[p_1p_2p_3]$). Therefore, Z_1' is generated by $[p_2p_3] - [p_1p_3] + [p_1p_2]$ whereas $B_1' = \emptyset$. So $H_1' = Z_1'$ and $\beta_1' = 1$. Comparing the two examples, we see that by eliminating the two-face from X (roughly speaking, punching hole in the triangle) a generator of H_1 is created. In conclusion, the homology spaces characterize the presence of holes in simplicial complexes. Indeed, the zeroth Betti number is the number of connected components of X, the first Betti number is the number of generators of two-dimensional (polygonal) holes, the third Betti number is the number of generator of three-dimensional holes (convex polyhedron), etc.

2. Persistent homology

The starting point in persistent homology is a filtration. As in Ref. [7], we call a simplicial complex X filtered if we are given a family of subspaces $\{X_v\}$ parametrized by N, such that $X_v \subseteq X_w$ whenever $v \leqslant w$ and X_v is a simplicial complex. The family $\{X_v\}$ is called a filtration.

There are many ways to construct a filtration from a point cloud or a network. The most popular filtration for data analysis is the Rips-Vietoris filtration [7].

The Rips-Vietoris complex is a simplicial complex associated to a set of points in a metric space in the following way: every point p is the center of a radius ρ ball $D(p,\rho)$ and $n+1$ points $\{p_0,\dots,p_n\}$ determine an n face in the Rips-Vietoris complex if the corresponding radius ρ balls intersect two by two, i.e., $D(p_i,\rho)\cap D(p_j,\rho) \neq \emptyset$ for all $i \neq j \in \{0\dots n\}$. Clearly, the Rips-Vietoris complex depends on the parameter ρ and if $\rho_1 < \rho_2$ the complex with ρ_1 radius balls is contained in

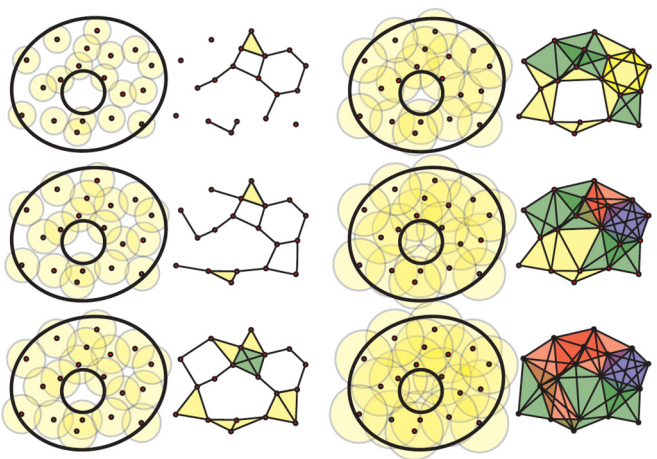

FIG. 10. Rips complex filtration. Reproduced with owner's permission from Ref. [6].

the complex with ρ_2 radius balls. To the growth of ρ we obtain an increasing sequence of simplicial complexes, a filtration, the Rips-Vietoris filtration. In this context persistent topological features of the filtration are considered as features of the point cloud. Figure 10 pictorially represents a Rips-Vietoris filtration: given a point cloud, it is shown that simplicial complexes of increasing complexity are found by increasing the radii of the balls centered at the points of the cloud.

The following basic properties of the algebraic structure of persistent homology hold.

Proposition 1. Let X and Y be two simplicial complexes, a simplicial map $f : X \to Y$ is a map sending vertices of X to vertices of Y and faces of X to faces of Y. Then f determines a linear map between the homology groups $H_i(f) : H_i(X) \to H_i(Y)$ for all i. From which, the following makes sense.

Definition 3. The persistent homology module of a filtration is given by the direct sum of the homology groups of the simplicial complexes $H_n(X_v)$ and the linear maps $i_{v,w} : H_n(X_v) \to H_n(X_w)$ induced in homology by the inclusions $X_v \hookrightarrow X_w$ for all $v \leqslant w$.

Following [7], this system is called a module because the direct sum of vector spaces $H_n = \oplus_v H_n(X_v)$ has a $k[x]$-module structure via an algebraic action given by $x \times m := i_{v,v+1}(m)$ for $m \in H_n(X_v)$. The linear maps $i_{v,v+1}$ are not always injective. A persistent homology generator is a generator of H_n as $k[x]$ module, i.e., an element $g \in H_n(X_v)$ such that there is no $h \in H_n(X_w)$ for $w < v$ with the property that $x^{v-w}h = g$. By the structure theorem on modules over principal ideal domains, the isomorphism class of a $k[x]$ module is completely determined by the degree of each generator g (birth of the generator β_g) and the degree in which the generator is annihilated by the module action (death of the generator δ_g). The persistence (lifetime) of a generator is measured by $p_g := \delta_g - \beta_g$.

Persistent homology modules can be computed using libraries like JAVAPLEX (Java) or DIONYSUS (C++), which are both available from the Stanford's CompTop group [32].

[1] M. Nakahara, *Geometry, Topology and Physics* (Adam Hilger, Bristol, 1991).

[2] M. Pettini, *Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics*, IAM Series, No. 33 (Springer-Verlag, New York, 2007).

[3] R. Franzosi and M. Pettini, Topology and phase transitions ii. Theorem on a necessary relation, Nucl. Phys. B **782**, 219 (2007).

[4] R. Franzosi, M. Pettini, and L. Spinelli, Topology and phase transitions i. Preliminary results, Nucl. Phys. B **782**, 189 (2007).

[5] P. Niyogi, S. Smale, and S. Weinberger, Finding the homology of submanifolds with high confidence from random samples, Discrete Comput. Geom. **39**, 419 (2008).

[6] R. Ghrist, Barcodes: The persistent topology of data, B. Am. Math. Soc. **45**, 61 (2008).

[7] G. Carlsson and A. Zomorodian, Persistent homology - A survey, Discrete Comput. Geom. **33**, 249 (2005).

[8] G. Carlsson, Topology and data, B. Am. Math. Soc. **46**, 255 (2009).

[9] H. Edelsbrunner, D. Letscher, and A. Zomorodian, Topological persistence and simplification, Discrete Comput. Geom. **28**, 511 (2002).

[10] G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P. J. Hellyer, and F. Vaccarino, Homological scaffolds of brain functional networks, J. Roy. Soc. Interface **11**, 20140873 (2014).

[11] V. De Silva and R. Ghrist, Coverage in sensor networks via persistent homology, Algeb. Geom. Topol. **7**, 339 (2007).

[12] M. Antoni and S. Ruffo, Clustering and relaxation in hamiltonian long-range dynamics, Phys. Rev. E **52**, 2361 (1995).

[13] A. Campa, T. Dauxois, and S. Ruffo, Statistical mechanics and dynamics of solvable models with long-range interactions, Phys. Rep. **480**, 57 (2009).

[14] J. Milnor, *Morse Theory*, Ann. Math. Studies 51 (Princeton University Press, Princeton, 1963).

[15] L. Casetti, M. Pettini, and E. G. D. Cohen, Phase transitions and topology changes in configuration space, J. Stat. Phys. **111**, 1091 (2003).

[16] L. Casetti, M. Pettini, and E. G. D. Cohen, Geometric approach to hamiltonian dynamics and statistical mechanics, Phys. Rep. **337**, 237 (2000).

[17] M. Kastner and D. Mehta, Phase Transitions Detached from Stationary Points of the Energy Landscape, Phys. Rev. Lett. **107**, 160602 (2011).

[18] Let us remark that in mathematics a counterexample to a theorem—in the absence of flaws in its proof—does not mean that the theorem is false and has to be discarded. On the contrary, a counterexample can stimulate a refinement of a theory. There is a famous counterexample that Milnor gave against De Rham's cohomology theory: the two manifolds $M = \mathbb{S}^2 \times \mathbb{S}^4$ (product of two spheres), and $N = \mathbb{C}P^3$ (complex-projective space), are neither diffeomorphic nor homeomorphic yet have the same cohomology groups. Apparently this is a lethal argument against De Rham's cohomology. But the introduction of the so-called cup product fixed the problem and saved the theory making it more powerful.

[19] L. Caiani, L. Casetti, and M. Pettini, Hamiltonian dynamics of the two-dimensional lattice ϕ^4 model, J. Phys. A: Math. Gen. **31**, 3357 (1998).

[20] L. Caiani, L. Casetti, C. Clementi, G. Pettini, M. Pettini, and R. Gatto, Geometry of dynamics and phase transitions in classical lattice ϕ^4 theories, Phys. Rev. E **57**, 3886 (1998).

[21] M. Pettini L. Casetti, and E. G. D. Cohen, Topological Origin of the Phase Transition in a Mean-Field Model, Phys. Rev. Lett. **82**, 4160 (1999).

[22] L. Angelani, L. Casetti, M. Pettini, G. Ruocco, and F. Zamponi, Topology and phase transitions: From an exactly solvable model to a relation between topology and thermodynamics, Phys. Rev. E **71**, 036152 (2005).

[23] M. Gori, R. Franzosi, and M. Pettini, arXiv:1602.01240 [cond-mat.stat-mech].

[24] J. M. Alongi and G. S. Nelson, *Recurrence and Topology*, Vol. 85 (American Mathematical Society, Providence, 2007).

[25] R. I. McLachlan and P. Atela, The accuracy of symplectic integrators, Nonlinearity **5**, 541 (1992).

[26] L. Casetti, Efficient symplectic algorithms for numerical simulations of hamiltonian flows, Phys. Scr. **51**, 29 (1995).

[27] M. P. Do Carmo, *Riemannian Geometry* (Birkhäuser, Boston, 1992).

[28] J. Silva, J. S. Marques, and J. M. Lemos, Sparse multidimensional scaling using landmark points, Conference: Neural Information Processing Systems-NIPS, 2005.

[29] J. Gamble and G. Heo, Exploring uses of persistent homology for statistical analysis of landmark-based shape data, J. Multivar. Anal. **101**, 2184 (2010).

[30] P. Bubenik, Statistical topological data analysis using persistence landscapes, J. Mach. Learn. Res. (JMLR) **16**, 77 (2015).

[31] F. Chazal, B. T. Fasy, F. Lecci, B. Michel, A. Rinaldo, and L. Wasserman, arXiv:1406.1901v1 [math.AT].

[32] http://comptop.stanford.edu/