Homological percolation and the Euler characteristic

Omer Bobrowski*
Viterbi Faculty of Electrical Engineering Technion, Israel Institute of Technology, Haifa 32000, Israel

Primoz Skraba†
School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom

(Received 29 October 2019; accepted 31 January 2020; published 9 March 2020)

In this paper we study the connection between the zeros of the expected Euler characteristic curve and the phenomenon which we refer to as homological percolation—the formation of "giant" cycles in persistent homology, which is intimately related to classical notions of percolation. We perform an experimental study that covers four different models: site percolation on the cubical and permutahedral lattices, the Poisson-Boolean model, and Gaussian random fields. All the models are generated on the flat torus $\mathbb{T}^d$ for $d = 2, 3, 4$. The simulation results strongly indicate that the zeros of the expected Euler characteristic curve approximate the critical values for homological percolation. Our results also provide some insight about the approximation error. Further study of this connection could have powerful implications both in the study of percolation theory and in the field of topological data analysis.

DOI: 10.1103/PhysRevE.101.032304

I. INTRODUCTION

This paper aims to make a connection between two seemingly unrelated mathematical topics in the context of spatial stochastic processes. The first is a large-scale phenomenon we refer to as *homological percolation*, where *giant cycles* are formed. The second is an integer-valued topological invariant, known as the *Euler characteristic* (EC).

To describe these two topics and the connection between them, we will use the language of *persistent homology* [1–4], which is the main workhorse in the field of topological data analysis (TDA) [5,6]. Persistent homology is an algebraic-topological functional that is applied to *filtrations* (nested sequences) of topological spaces. It is essentially a tool that tracks the formation and destruction of topological features such as connected components (0-cycles), holes (1-cycles), cavities (2-cycles), and their higher-dimensional analogs (nontrivial k-cycles). Persistent homology has been demonstrated to be a powerful tool in the analysis of various types of data (e.g., neuroscience [7], cosmology [8], and complex networks [9]).

In a data-analytic language, we can think of the giant cycles as the "topological signal" hidden in the filtration, as they capture information about the underlying space. The small cycles are considered as nuisance "noise" one might wish to filter, in order to reveal the signal, see Fig. 1. One of the main challenges in TDA is to identify which feature belongs to which group. Suppose that X is a "nice" topological space, and that we have a filtration $\{X_t\}_{t\in\mathbb{R}}$ of spaces such that $X_s \subset X_t \subset X$ for all $s < t$. We can classify the cycles captured by persistent homology into two groups: the group of "giant" cycles will be those that are also nontrivial cycles (holes) in X, while all other cycles (that are trivial in X) will be considered "small." These cycles could just as accurately be referred to as infinite or essential cycles.

The appearance or birth of the giant k-cycles, or in other words the intrinsic homology classes of the underlying space, is the phenomenon we refer to as *homological percolation*. Our definition is inspired by the question: *At what scale does the nontrivial k-cycles of the underlying space first appear*. These are precisely the cycles described above.

From the theoretical-probabilistic perspective, the study of this type of phenomena (in terms of intrinsic topology) is at a very early stage. However, relying on classical results in percolation theory together with recent simulations (including in this paper), it is conjectured that homological percolation occurs as a sharp phase transition. In other words, given a random filtration $\{X_t\}$ the probability for creating the giant cycles switches from zero to one as a result of an infinitesimal increase in the filtration parameter t. Moreover, for a fixed k the thresholds for all giant k-cycles coincide, and the critical values are increasing in k. In other words, if t_k^{perc} is the critical value for the emergence of the giant k-cycles, then $t_k^{\text{perc}} \leqslant t_{k'}^{\text{perc}}$ for all $k < k'$.

An important difference in our definition is the use of persistent homology to quantify "giant" rather than geometric size. While in the percolation literature, "giant" or "infinite" is often described in terms of the number of vertices or the diameter of the component, a different characterization has recently been widely used [11]: asking whether the origin is connected to the boundary of a surrounding box as the size of the box goes to infinity. This is an inherently topological definition, as it asks if such a path exists. The relationship

*omer@ee.technion.ac.il
†p.skraba@qmul.ac.uk

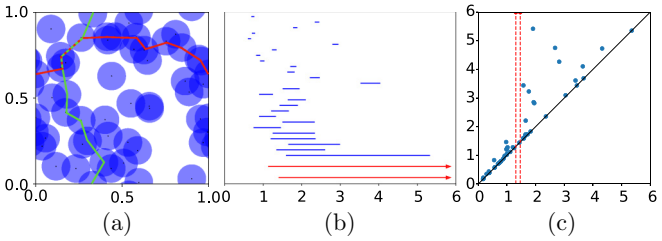

FIG. 1. Persistent homology for a point sample on the two-dimensional flat torus (a unit box with periodic boundary conditions). (a) A set of 50 points $\mathcal{X}$ sampled from the flat torus. The filtration X_t taken here is the union of balls of radius t around the points. We increase t from 0 to ∞ and calculate PH_1. (b) The barcode for PH_1. Each 1-cycle is represented by a bar, where the endpoints are the radii in which the cycle was formed and filled in (birth, death). The two red bars (bottom two with arrows) correspond to the two cycles in the torus ("the giant cycles") while the blue ones are considered as "noise". (c) The persistence diagram for PH_1. Here the (birth, death) pairs are plotted as points in the plane. Two points are far away from the diagonal (shown by the vertical red dashed line), representing the true nontrivial or giant cycles of the torus. This figure was generated using the GUDHI package [10].

between this notion and our definition is discussed in more detail in Sec. II.

There are a few other higher-dimensional definitions for percolation studied in the literature. A common definition, known as "Topological percolation" [12], studies changes in the behavior of the number of k-dimensional nontrivial cycles in X (the kth *Betti number*). In a number of models, it has been observed that for each dimension, there is a range where the corresponding Betti number dominates [13] and the percolation threshold is defined in terms of the crossover between these ranges. This phenomenon has been studied in mean-field models, such as higher-dimensional versions of the Erdős-Rényi random graph (e.g., the Linial-Meshulam model [14] or the Erdős-Rényi clique complex [13]). A related phenomenon is the appearance of the "giant" shadow, where most new simplices form new topological features, which has been observed to also occur in the vicinity of the crossover of the Betti numbers (see Ref. [15]). As the underlying space of the these mean-field models is the complete simplex, it has no intrinsic topology making our definition useless for the study of these types of phenomena. On the other hand, the definition in terms of Betti numbers only counts the number rather than the size of the cycles which appear, while our definition specifically asks when a *global structure* appears. In our experiments, we investigate the relationship between these two definitions which concludes this part of the story.

The second half of the story in this paper is about a rather different object. The *Euler characteristic* (EC) is an integer-valued topological invariant that can be assigned to a topological space. Quite remarkably, the EC can be defined in several different ways, vastly different in nature (e.g., geometric, combinatorial, topological, analytic), which are all equivalent under quite general conditions (local compactness). There are many ways to compute the EC. For example, we can compute the EC by counting the number of cells in a cell complex, counting the critical points of a Morse function, or integrating the Gaussian curvature of a manifold. For our purposes, we use the "homological" definition of the EC, i.e.,

$$\chi(X) := \sum_k (-1)^k \beta_k(X), \tag{1.1}$$

where X is a topological space and $\beta_k(X)$ denotes the kth Betti number. The EC is an intriguing mathematical object [16,17], and over the years it was also found to be very useful as a statistical tool. Two areas of applications where the EC was proven to be quite powerful are cosmology [18–20] and brain imaging [21,22]. In the random setting, somewhat surprisingly, much more is known about the distribution of the EC for a random space X (as we see in Sec. III) compared to the individual Betti numbers defining it.

Given a topological space X and a filtration $\{X_t\}_{t\in\mathbb{R}}$, one may calculate its *EC curve* $\chi(t) := \chi(X_t)$. For several random filtrations, such as the ones discussed in this paper, the expected value $\mathbb{E}\{\chi(t)\}$ has been analyzed in the past [23,24]. In all the models we discuss here, as well as many others, while the EC curves look completely different, the expected EC curve has exactly $(d-1)$ zeros (where d is the dimension of the generating model). Denote these zeros by $t_1^{\mathrm{ec}}, \ldots, t_{d-1}^{\mathrm{ec}}$, and recall that $t_1^{\mathrm{perc}}, \ldots, t_{d-1}^{\mathrm{perc}}$ are the homological-percolation thresholds. The question we wish to pursue in this paper is the then following: Is there a connection between t_k^{perc} and t_k^{ec}?

A priori, there is no obvious reason why such a connection should exist. Indeed, both homological percolation and the EC curve are related to the homology of a given filtration. However, homological percolation describes the giant cycles formed, while the EC contains information about the total number of cycles, regardless of their size. Therefore, while it could be expected that the percolation defined in terms of dominating behavior of Betti numbers [12] would be related to the EC (see discussion of related work below), the connection with the intrinsic topology of the underlying space is much more unexpected.

Our main goal in this paper is to argue that such an unexpected connection exists and is potentially universal in the sense that it occurs across vastly different stochastic models. We note that we are not aiming to provide any analytic statements here. Instead, we want to suggest that this link exists by presenting simulation results for several random systems.

In this paper we consider four percolation models: site percolation on a cubical grid, site percolation on a permutahedral grid, a continuum percolation model, and the sublevel sets of Gaussian random fields. In all these models, an explicit formula for the expected EC curve can be calculated. We simulate these models on the d-dimensional *flat torus* (i.e., a d-dimensional box with periodic boundary conditions) and compare the critical percolation values to the zeros of the expected EC curve.

A. Main results

The simulations we present in Sec. IV highly suggest a positive answer to the question above. In all models and all dimensions tested, the simulations indicate that $t_k^{\mathrm{perc}} \approx t_k^{\mathrm{ec}}$, where determining the exact meaning of "$\approx$" remains future work. Note that all the models we study depend on a

parameter n (either grid size, or number of points). Defining $\Delta_k := (t_k^{\text{perc}} - t_k^{\text{ec}})$, it would be tempting to conjecture that $\Delta_k \xrightarrow{n\to\infty} 0$. However, our simulation results indicate that while the difference converges, the limit might be nonzero. If the model is symmetric with respect to the parameter t (e.g., the permutahedral complex and a zero-mean Gaussian field), and if d is even, then our simulations as well as analytical arguments show that indeed $\Delta_{d/2} = 0$. A second-order observation we make from the simulations is that the sign of the error term Δ_k is not arbitrary. For $k < d/2$ it seems that we always have $\Delta_k < 0$, while for $k > d/2$ we have $\Delta_k > 0$. We also experimentally illustrate a connection between our definition and a definition in terms of Betti numbers which has been studied before (see below).

In addition to the interesting and surprising mathematical phenomenon that we reveal here, there are also potential applied aspects to the conjectures we make in this paper. In most models in statistical physics, the exact percolation thresholds are not known. At best there exist some theoretical bounds, or numerical approximations. However, since the zeros of the expected EC curves can be found in many cases, our hope is that these could be used them as an improved approximation for the real percolation thresholds. In addition to probability and statistical physics, we also believe that these results can have implications in TDA, for example by enhancing the detection of significant topological features in data. For example, we can infer that we have captured the k-dimensional homology of the underlying space if once the $k+1$-Betti numbers dominate.

Percolation, i.e., the appearance of giant or infinite connected components, has been and continue to be the subject of intensive research in statistical physics, complex systems, and mathematics. Identifying numerical values for relevant thresholds is an important and difficult problem. So it is not surprising, given the computability of the Euler characteristic that the the connection between percolation thresholds and the EC has been previously studied in Ref. [25], and has certainly inspired the current work. There, the authors focused on classical percolation (the formation of giant connected components) which is a special case of the higher-dimensional homological notion we consider here. Thus, there is only a single threshold to consider. In addition, the models considered in Ref. [25] are two- and three-dimensional, whereas here we wish to argue that this phenomenon is more universal in that it occurs in all dimensions and across various types of percolation models. In our setting, we define percolation *in terms of persistent homology* and investigate the connection with zeros of the Euler characteristic curve. In upcoming work [26], we will make the connection between classical notions of percolation and the homological percolation we introduce here precise. Another related work is Ref. [27], where the authors studied a closely related notion on the two-dimensional torus for discrete models looking for the appearance of components which "wrap-around," which they studied through winding numbers.

As mentioned earlier, there are numerous other notions of topological phase transitions. For example, Ref. [28] studies the connection between the EC zeros and the switch between the dominant Betti number from β_k to β_{k+1}. This is shown to occur around the $(k+1)$th zero of the expected EC. While we also show experimentally that these phase transitions are related to our notion of homological percolation, these are very different types of phase transitions. The Betti numbers are quantitative descriptors, counting the total number of cycles. The majority of these cycles can be shown to be small and local. Moreover, from the definition of the EC (1.1), it is highly conceivable that the EC zeros are strongly connected to changes in the Betti numbers. On the other hand, in this paper we study the emergence of the essential cycles, which is a qualitative phenomenon that describes the formation of global structures. Here the connection to the EC is quite unexpected. In addition, the model studied in Ref. [28] is the clique complex over the Erdős-Rényi random graph, which is different than the geometric models studied here. We note that the connection between the crossover of Betti numbers and the EC has also been previously observed in several geometric models [29,30]. The connection with the statistics of persistence diagrams (i.e., the distribution of "small" cycles) has been studied in Refs. [31,32]. We refer the reader to the references in Ref. [28] for other connections between physical quantities and the zeros of the Euler characteristic curve. Finally, some aspects of persistent homology in the continuum percolation model were explored in Ref. [33] and the special case of the appearance of top-dimensional homology, which we do not consider here has been investigated in Refs. [23,34,35].

II. PRELIMINARIES

In this section we wish to provide a rather nonformal introduction to the fundamental terminology we will be using in this paper. We include references for more formal treatment of the topics discussed.

Homology is an algebraic-topological structure that describes various types of topological phenomena in topological spaces using algebraic structures. Let X be a topological space. In this paper we will use homology with field coefficients. In this case, the kth homology $H_k(X)$ is a vector space, such that its basis elements correspond to the following features. The basis of $H_0(X)$ corresponds to the connected components of X (also referred to as 0-cycles), $H_1(X)$ to "holes" in X (1-cycles), $H_2(X)$ to "voids" or "bubbles" in X (2-cycles), and, more generally, $H_k(X)$ represents k-cycles that can be thought of as shapes similar to a k-sphere. The Betti numbers are the corresponding dimensions $\beta_k(X) := \dim H_k(X)$ that count the number of nontrivial k-cycles in X. In Fig. 2 we present a few examples for spaces along with their Betti numbers. There are many excellent introductions to homology theory, for example, we refer the reader to [36]. In addition to describing the topology of a single space X, the language of homology also provides means to match k-cycles between two spaces. Let X, Y be topological spaces, and let $f : X \to Y$ be a continuous function. Then for every k there exists a corresponding linear transformation called the *induced map* $f_* : H_k(X) \to H_k(Y)$ that maps k-cycles in X into k-cycles in Y.

Persistent homology is one of the fundamental tools used in the field of *applied topology* or *topological data analysis*. The motivation for developing persistent homology was that

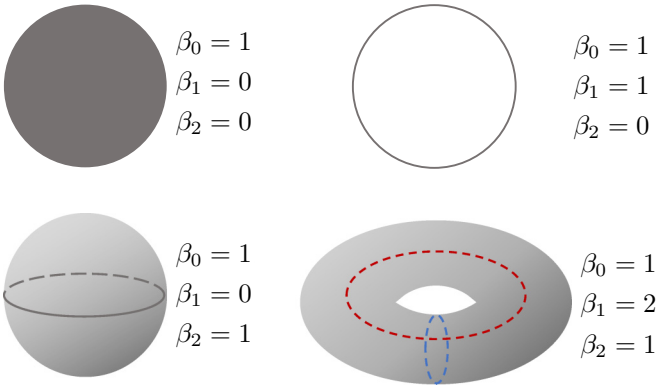

FIG. 2. Example of simple topological spaces with their Betti numbers. From left to right: a disk, a circle, a two-dimensional sphere, and a two-dimensional torus. For the torus, we marked the two 1-cycles in dashed lines.

as data-analytic features, homological properties can be quite unstable in the sense that small perturbations to the data may result in a significant change of the homological structure. The solution provided by persistent homology is that instead of extracting the homological features of a single space, we consider a sequence of spaces and extract information about homological cycles together with their evolution throughout the filtration. This can be thought of as a "multiscale" version of homology.

A bit more concretely, a filtration $\mathbf{X} = \{X_t\}_t$ is a set of topological spaces such that for all $s < t$ we have $X_s \subset X_t$. The inclusion maps $i : X_s \hookrightarrow X_t$ induce mappings between cycles $i_* : H_k(X_s) \to H_k(X_t)$. These mappings allow us to track the evolution of cycles as they form and disappear throughout the filtration. Without getting into the formal mathematical definitions, we can think of the kth persistent homology $\mathrm{PH}_k(\mathbf{X})$ as a collection (more accurately a graded-module) of k-persistent cycles. For each cycle $\gamma \in \mathrm{PH}_k(\mathbf{X})$ we can assign two values $\mathrm{bth}(\gamma)$ and $\mathrm{dth}(\gamma)$ standing for "birth" and "death" $[\mathrm{bth}(\gamma) \leqslant \mathrm{dth}(\gamma)]$, representing the times where the cycle γ was formed and later filled in. As a data-analytic tool, persistent homology provides a topological signature for data that also includes some geometric information that makes it more robust to noise than the fixed-scale homology. See Fig. 1 for an example. For more details as well as formal definitions see Refs. [3,37]. For an overview of TDA see Refs. [5,6,38].

A. Giant cycles

We are now ready to define homological percolation. Let M be a "nice" compact space, and let $\{X_t\}$ be a filtration such that $X_t \subset M$ for all t. As mentioned in the Introduction, by giant cycles we refer to those cycles that appear in the filtration for some t and represent one of the nontrivial cycles in $H_k(M)$. These are also referred to as *essential cycles*. We will make this description a bit more formal.

For each t, the inclusion map $i : X_t \hookrightarrow M$ induces a map in homology $i_{*,t} : H_k(X_t) \to H_k(M)$. The image of $i_{*,t}$ stands for all the cycles that exist in X_t and are mapped to nontrivial cycles in M. We will refer to these cycles as "giant." By the term *homological percolation* we refer to the study of how and when these giant cycles are formed. For example, the longest red bars in Fig. 1(b) represent the two holes of the torus, and therefore we consider them as giant, while the other cycles are noise.

Suppose that the filtration $\{X_t\}$ is generated at random (we will discuss specific random models in Sec. III). For each t, fixing k we can define the following events:

$$E_t := \left\{\mathrm{Im}\left(i_{*,t}\right) \neq 0\right\}, \qquad A_t := \left\{\mathrm{Im}\left(i_{*,t}\right) = H_k(M)\right\}.$$

In other words, E_t is the event that there *exists* a giant k-cycle in X_t, while A_t is the event that *all* possible giant k-cycles exist in X_t. It is conjectured that similarly to other percolation models, the appearance of the giant k-cycles follows a sharp phase transition. This means that there exists a value $t_k^{\mathrm{perc}} > 0$ such that

$$\mathbb{P}(E_t) = \mathbb{P}(A_t) = \begin{cases} 1 & t > t_k^{\mathrm{perc}} \\ 0 & t < t_k^{\mathrm{perc}} \end{cases}, \tag{2.1}$$

where in most cases we study, the stochastic model has an intrinsic parameter n and the equalities above hold in the limit when $n \to \infty$. This conjecture is supported by simulations (as the ones presented in this paper), as well as a theoretical work in progress [26] for the Boolean model discussed below. It is further conjectured that the thresholds are ordered, so that $t_1^{\mathrm{perc}} < t_2^{\mathrm{perc}} < \cdots < t_{d-1}^{\mathrm{perc}}$ where d is the maximal degree possible (dictated by the dimension of M).

For our experiments, we choose M to be the d-dimensional torus. This has the advantage that it has nontrivial homology in all dimensions. In this setting, the giant component in classical percolation is related to the one-dimensional homology of the torus. We expect that the giant component "wraps" around to form a giant 1-cycle but this has not been rigorously proven [39] and something we will address in Ref. [26].

The *Euler characteristic (EC)* is an integer-valued additive functional. Using the language of homology, one can define the EC of a topological space X as

$$\chi(X) := \sum_k (-1)^k \beta_k(X),$$

where $\beta_k(X)$ are the Betti numbers discussed above. One of the key properties of the EC is that it is a *topological invariant*, namely X and Y are two spaces that are "similar" topologically in the sense that there is a continuous transformation form one to the other (known as *homotopy equivalence*) and then $\chi(X) = \chi(Y)$. The EC shows up in various areas of mathematics (combinatorics, integral geometry, topology, analysis, etc.) and can be defined in various different ways.

In most stochastic models, evaluating quantities related to the distribution of homology or persistent homology is between difficult to impossible. Surprisingly, however, this is not the case for the EC. For example, the expected values of the Betti numbers are unknown in almost all stochastic models studied to date, while in almost all the models an explicit formula for the expected EC exists (see Sec. III). Therefore, for probabilistic and statistical analysis, the EC is much favorable, and indeed several interesting applications in statistics and data science were developed based on EC calculations [20,21,40,41].

Studying a filtration $\{X_t\}$ as in persistent homology, we can define the EC curve $\chi(t) := \chi(X_t)$, that tracks the evolution of

the EC in time. In the random setting we study in this paper, we will mainly focus on the expected EC curve

$$\bar{\chi}(t) := \mathbb{E}\{\chi(t)\}.$$

III. RANDOM PERCOLATION MODELS

We focus on three different types of stochastic models to establish our conjectures about the connection between the Euler characteristic and homological percolation. In this section we provide the basic definitions for these models, as well as the formulas we use to calculate the expected EC curve. One of the main reasons for choosing these models is that in all of them, we can derive an explicit formula for the expected EC curve $\bar{\chi}(t)$ (where $t = p, \lambda,$ or α depending on the model below).

All the models we discuss generate random subsets of the d-dimensional flat torus $\mathbb{T}^d$. By "flat torus" we refer to the quotient of the box $[0,1]^d$ with the relation $\{0 \sim 1\}$ (i.e., opposite faces are "glued" together). The flat torus is a good model to study topological phenomena as (a) the metric on it is locally Euclidean, (b) it is a manifold with no boundary, and (c) it has nontrivial homology in all degrees $k = 0, \ldots, d$. More precisely, $\beta_k(\mathbb{T}^d) = \binom{d}{k}$. While the models discussed below may be defined for general Riemannian manifolds (e.g., with nonzero curvature), we do not investigate this further in this paper. This investigation would go beyond the scope of this paper as even generating Poisson samples for arbitrary manifolds is a difficult problem in its own right. We note that while the EC should generically have $(d-1)$ zeros, not all manifolds have nontrivial homology for all k. Thus, it is possible that there will be a zero without a corresponding giant cycle. Nevertheless, we conjecture that it is still true that the kth zero of the EC approximates the birth time of the giant k-cycles, assuming the latter exist. Showing this is left as a future direction of research.

A. Site percolation models

We start with simple discrete models for random subsets of $\mathbb{T}^d$. As opposed to the continuous models we discuss later, discrete percolation models are well studied and often more tractable, both from a theoretical and simulation perspective. We will examine two types of structures, discussed next.

1. Cubical complex

A cubical complex Q is a collection of cubical faces (i.e., vertices, edges, squares, cubes, etc.), that is closed under the boundary operation. We denote by Q_n^d the cubical complex obtained by taking the flat torus $\mathbb{T}^d = [0,1]^d \backslash \{0 \sim 1\}$ and splitting it into n equal-size boxes, where we assume that $n = m^d$, for some $m \in \mathbb{N}$. Note that every d-dimensional box is in Q_n^d together with all its k-dimensional faces ($k = 0, \ldots, d-1$).

We will consider cubical complexes Q that are subsets of Q_n^d. Each such complex is homeomorphic to a subset of $\mathbb{T}^d$ via the natural embedding. Thus, we will interchangeably refer to Q as either a sub complex of Q_n^d or as a closed subset of $\mathbb{T}^d$. Denote by $F_k(Q)$ the number of k-dimensional faces of Q. Then the EC of Q can be calculated by (cf. Ref. [36])

$$\chi(Q) = \sum_{k=0}^{d} (-1)^k F_k(Q). \tag{3.1}$$

The *random* cubical complex we study here, denoted $Q(n,p)$, is generated by taking Q_n^d and declaring each d-dimensional face (or a site) as either *open* with probability p or *closed* with probability $1-p$, independently between the faces. We then define $Q(n,p)$ as the union of all open boxes (together with their lower-dimensional faces). See an example in Fig. 3(a) and the persistence diagram in Fig. 3(e). Calculating the expected EC using (3.1) (see Appendix A) yields

$$\begin{aligned}\bar{\chi}_Q(p) :&= \mathbb{E}\{\chi(Q(n,p))\} \\ &= n \sum_{k=0}^{d} (-1)^k \binom{d}{k} [1 - (1-p)^{2^{d-k}}]. \end{aligned} \tag{3.2}$$

Remark. Notice that the definition of $Q(n,p)$ does not imply any connection between $Q(n,p_1)$ and $Q(n,p_2)$ for $p_1 \neq p_2$. However, in order to discuss percolation phenomena as well as persistent homology for the cubical complex, we want the sequence $\{Q(n,p)\}_{p=0}^1$ to be a filtration, so that $Q(n,p_1) \subset Q(n,p_2)$ for all $p_1 < p_2$. The simplest way to establish that is the following. Let $U_1, \ldots, U_n$ be n independent and identically distributed random variables, so that $U_i \sim U[0,1]$. Then, for any fixed p, we say that site i is open in $Q(n,p)$ if $U_i \leqslant p$. This way for all $p \in [0,1]$ the complex $Q(n,p)$ has the distribution we desire, and indeed $Q(n,p_1) \subset Q(n,p_2)$ for all $p_1 < p_2$.

2. Permutahedral complex

We introduce an alternative discrete model to address some inherent shortcomings of the cubical model. In two dimensions, a hexagonal tiling is often used in percolation theory instead of the $\mathbb{Z}^2$ grid. Here we will use a higher-dimensional notion, we refer to as a *permutahedral* tessellation, where the basic building block is a *permutahedron*—the generalization of a hexagon to arbitrary dimensions. Notice that taking all 3! permutations of the coordinates (1,2,3) yields six vertices in $\mathbb{R}^3$ that form a hexagon. Similarly, a d-dimensional permutahedron is the polytope obtained by taking the convex hull of all $(d+1)!$ permutations of $(1, \ldots, d+1)$ in $\mathbb{R}^{d+1}$.

Next we construct the *permutahedral lattice*. The definitions here follow in Ref. [42] (Section 6.6) and Ref. [43]. Define

$$\widehat{\mathbb{R}}^d := \left\{ (x_0, x_1 \ldots, x_d) \in \mathbb{R}^{d+1} \,\middle|\, \sum_{i=0}^{d} x_i = 0 \right\},$$

i.e., $\widehat{\mathbb{R}}^d$ is a d-dimensional plane in $\mathbb{R}^{d+1}$. Define the A_d lattice as $A_d := \mathbb{Z}^{d+1} \cap \widehat{\mathbb{R}}^d$, and its dual A_d^* as

$$A_d^* = \left\{ x \in \widehat{\mathbb{R}}^d \mid \forall y \in A_d : x \cdot y \in \mathbb{Z} \right\}.$$

Taking $\pi : \widehat{\mathbb{R}}^d \to \mathbb{R}^d$ to be the natural isometry, then the Voronoi cells of $\pi(A_d^*)$ form a permutahedral tessellation of $\mathbb{R}^d$, and the set of centers of these cells $\pi(A_d^*)$ is called a *permutahedral lattice*.

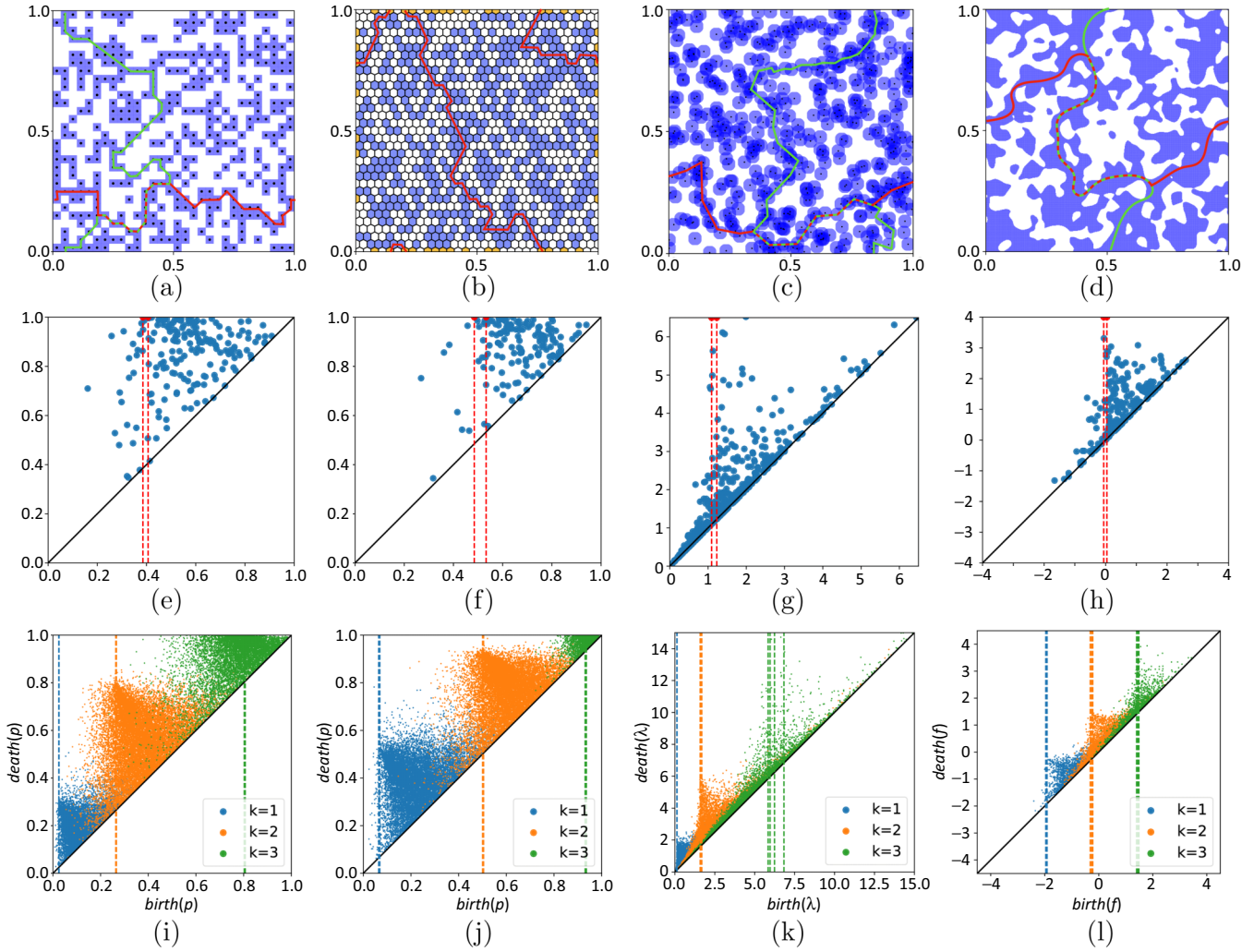

FIG. 3. The emergence of the giant 1-cycles in two dimensions for (a) the cubical site model $Q(n, p)$ with ($n = 2500$, $p = 0.43$), (b) the permutahedral site model $P(n, p)$ with ($n = 1024$, $p = \frac{1}{2}$), (c) Poisson-Boolean model $B(n, \lambda)$ with ($n = 500$), and (d) the GRF model $G(\alpha)$ (computed on a 512×512 grid). The red and green lines mark the two giant 1-cycles (recall the period boundary gluing). Note in $P(n, p)$ that only one giant cycle has appeared and that it is the sum of the two "obvious" giant cycles. [(d)–(g)] PH_1 for the corresponding models with the red vertical lines mark the birth times of the giant 1-cycles. [(h)–(l)] The persistence diagrams for the corresponding models for $d = 4$. In this case we have homology in degrees $k = 1, 2, 3$ (which alternatively dominate from left to right in increasing order of dimension). The vertical lines mark the birth times of the giant k-cycles. The relevant scale parameter is $n = 65536$ for $Q(n, p)$, $P(n, p)$, as well as the grid size for the GRF model. For $B(n, \lambda)$, $n = 5000$.

A *site* in this model is the closure of a Voronoi cell of the a point in $\pi(A_d^*)$. Each site has the structure of a d-dimensional permutahedron, hence the name of the model. Note that the points of $\pi(A_d)$ form the vertices of this polytope (see also Ref. [44]).

To go from a tessellation of $\mathbb{R}^d$ to a tessellation of the the torus $\mathbb{T}^d$, we use an analogous method as in the cubical case, gluing "opposite" faces together. In practice, this is done by identifying points of $\pi(A_d^*)$ (similarly to the periodic Delaunay complex [45]). As in the cubical case, this limits our choice of grid size (n), depending on the dimension. We denote the corresponding tessellation of $\mathbb{T}^d$ as P_n^d.

Similarly to the random cubical model, we define a random permutahedral complex. Let $P(n, p)$ be a random subset of P_n^d where each site is open with probability p, and closed with probability $1 - p$. As in the cubical case, we can calculate the EC by counting faces in different dimensions. This leads to (see Appendix A),

$$\begin{aligned}\bar{\chi}_P(p) :&= \mathbb{E}\{\chi[P(n, p)]\} \\ &= n\sum_{k=0}^{d}(-1)^{d-k}\left[1-(1-p)^{k+1}\right] \\ &\quad\times\sum_{j=0}^{k+1}(-1)^{k+1-j}\binom{k}{j}j^{d+1}. \end{aligned} \tag{3.3}$$

The main reason for our interest in the permutahedral complex is that, in contrast to the cubical model, it exhibits a powerful duality property. Let $P \subset P_n^d$ be a sub complex and recall that we define a giant as an element in the image

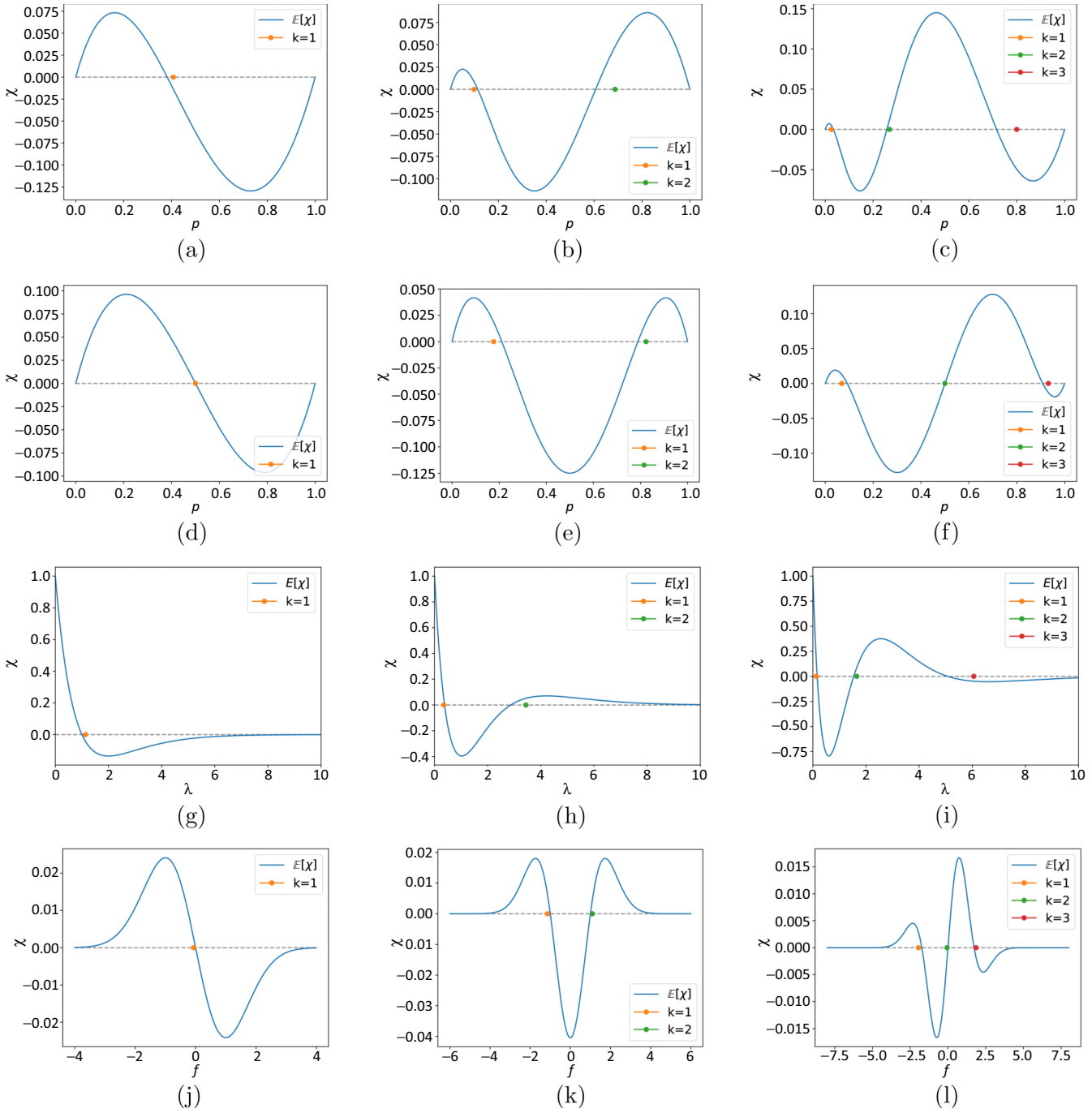

FIG. 4. The expected EC curve and the giant cycles. In each plot we draw the expected EC curve (solid line), along with the birth time of the first giant k-cycle for $k = 1, \ldots, d-1$ (dots). We simulated all the models on the d-dimensional torus, for $d = 2, 3, 4$ (from left to right). [(a)–(c)] The random cubical complex. [(d)–(f)] The random permutahedral complex. [(g)–(i)] The Boolean model. [(j)–(l)] The Gaussian random field.

$i_* : H_k(P) \to H_k(\mathbb{T}^d)$. Define,

$$\mathcal{B}_k(P) := \dim\{i_*[H_k(P)]\},$$

i.e., $\mathcal{B}_k(P)$ is the number of giant k-cycles in P. Next, let $P^c = \mathrm{cl}(P_n^d \backslash P)$, the closure of the complement. The following lemma (in fact a stronger version of it) is proved in Appendix C.

Lemma 3.1. For $0 \leqslant k \leqslant d$,

$$\mathcal{B}_k(P) + \mathcal{B}_{d-k}(P^c) = \beta_k(\mathbb{T}^d).$$

The lemma implies that whenever a giant k-cycle emerges in P, a giant $(d-k)$-giant cycle disappears in P^c and vice versa. This lemma can be viewed as a "homological" version of the duality argument used for percolation in $\mathbb{Z}^2$ (see, e.g.,

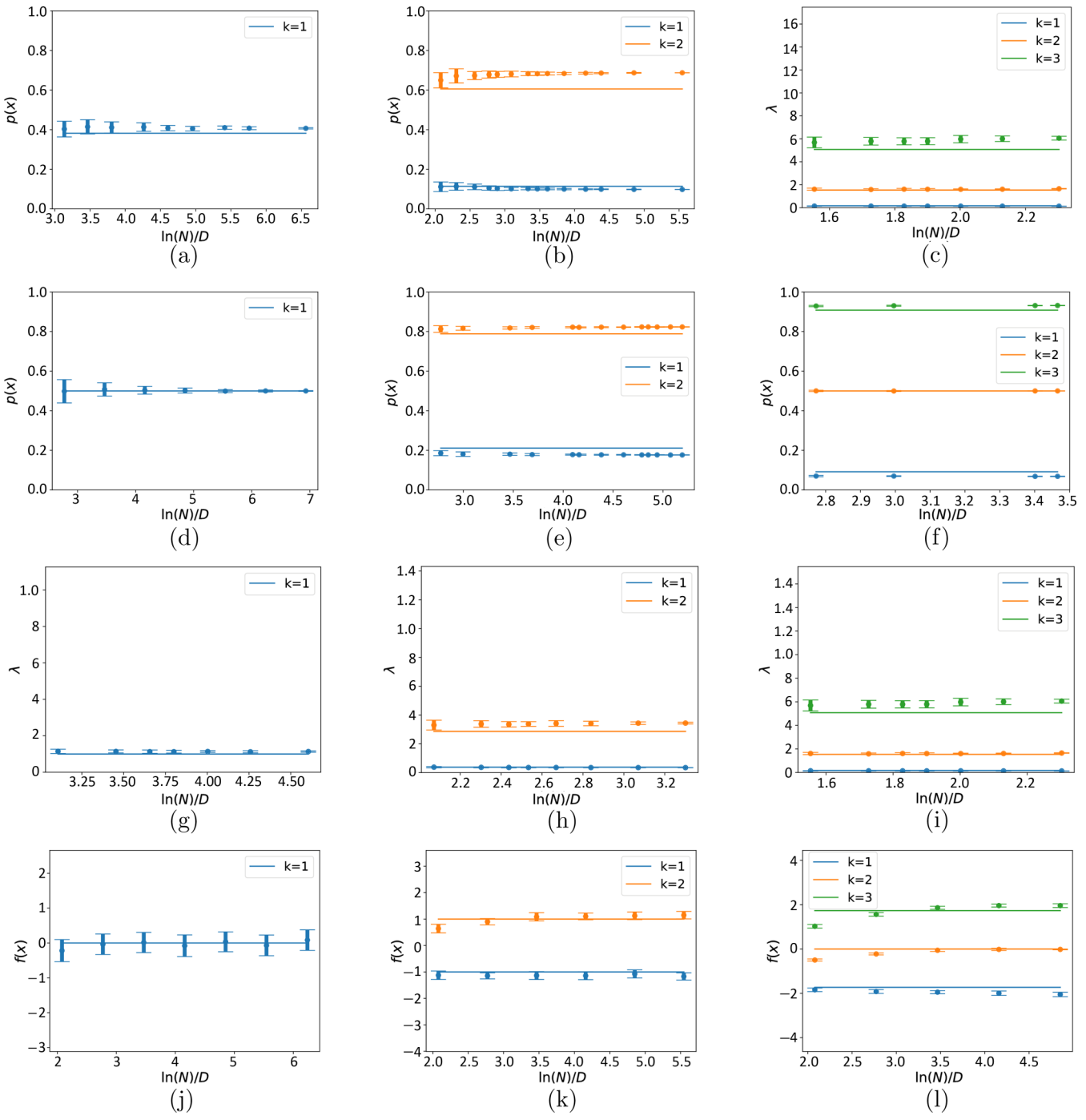

FIG. 5. Statistics for the birth time of the giant k-cycles. For each model we repeated the simulations in order to estimate the mean and variance of the birth time of the first giant k-cycle. In each figure the x axis is $\ln n$, and the y axis represent the parameter value (p, λ, or α). The dots represent the mean value estimate, and the bars around them follows the standard deviation. The horizontal lines mark the corresponding zeros of the EC curve. [(a)–(c)] The random cubical complex. [(d)–(f)] The random permutahedral complex. [(g)–(i)] The Boolean model. [(j)–(l)] The Gaussian random field.

Ref. [46]), that is, the fact that a horizontal crossing in one grid prevents a vertical crossing in the dual grid.

Notice that by the definition of $P(n, p)$, we have that $P^c(n, p)$ has the same distribution as $P(n, 1-p)$. Recall from the Introduction that we conjecture the existence of an increasing sequence of sharp thresholds denoted $p_1^{\text{perc}} < p_2^{\text{perc}} < \cdots < p_{d-1}^{\text{perc}}$ where the the giant k-cycles appear. Therefore, together with Lemma 3.1, we can show that if the sharp thresholds exist, then in the permutahedral complex case we have

$$p_k^{\text{perc}} = 1 - p_{d-k}^{\text{perc}}.$$

In other words, the appearance of the k-cycles in this model is in symmetry with the appearance of the $(d-k)$-cycles (as can be seen in the simulations later). Notice that if d is even, then we have that $p_{d/2} = 1/2$. This is a generalization of

the phenomenon known for the two-dimensional hexagonal lattice, where the (classic) percolation threshold is exactly 1/2 (see, e.g., Ref. [46]). The symmetry between $P^c(n,p)$ and $P(n,1-p)$ also implies a symmetry for the expected EC curve, so we have (see Appendix C),

$$\bar{\chi}_P(p) = (-1)^{d-1}\bar{\chi}_P(1-p).$$

While the symmetry of the EC curve is not obvious in Eq. (3.3), it is quite apparent in the simulations we present later.

B. Boolean model

Let $X_1, X_2, \ldots,$ be a sequence of independent and identically distributed random variables uniformly distributed on $\mathbb{T}^d$. Let $N \sim \text{Poisson}(n)$ be a Poisson random variable, independent of $\{X_i\}$. Then the process $\mathcal{P}_n := \{X_1, \ldots, X_N\}$ is called a *spatial Poisson process* on $\mathbb{T}^d$. The simple Boolean model we consider here is merely the union of balls

$$B_r(\mathcal{P}_n) := \bigcup_{p\in\mathcal{P}_n} B_r(p),$$

where $B_r(p)$ is a closed ball of radius $r > 0$ around p. In this model, it is known [47] that percolation occurs when $\omega_d n r^d = \lambda$ (or $r = (\lambda/\omega_d n)^{1/d}$) for some fixed value $\lambda > 0$, where ω_d is the volume of the unit ball in $\mathbb{R}^d$. Consequently, the percolation model we study is

$$B(n,\lambda) := B_{(\lambda/\omega_d n)^{1/d}}(\mathcal{P}_n). \tag{3.4}$$

In Ref. [23] a formula for the expected EC of $B(n,\lambda)$ was proved. The main idea was to consider the distance function to the point process $\mathcal{P}_n$ and evaluate the expected number of its critical points. A mathematical framework known as *Morse theory* provides a formula similar to (3.1), where the number of k faces is replaced by the number of critical points of index k. The result in Ref. [23] is the following formula:

$$\bar{\chi}_B(\lambda) := \mathbb{E}\{\chi(B(n,\lambda)\} = ne^{-\lambda}\left(1 + \sum_{k=1}^{d-1} A_{d,k}\lambda^k\right), \tag{3.5}$$

where $A_{d,k}$ are defined in Ref. [23] via some geometric integrals. For $d = 2$ we can show that this results in

$$\bar{\chi}_B(\lambda) = ne^{-\lambda}(1-\lambda).$$

For $d = 3$, the calculations in Ref. [48] show that

$$\bar{\chi}_B(\lambda) = ne^{-\lambda}\left(1 - 3\lambda + \frac{3}{32}\pi^2\lambda^2\right).$$

For higher dimensions, the integral formulas in Ref. [23] are difficult to calculate explicitly. In our simulations for $d = 4$, we use numerical methods to approximate the coefficients $A_{d,k}$.

C. Gaussian random fields

The last model we study is of a completely different nature than the previous ones. A real-valued Gaussian field on the torus, is a random function $f : \mathbb{T}^d \to \mathbb{R}$ such that for every k and every collection of points $x_1, \ldots, x_k \in \mathbb{T}^d$, the random variables $f(x_1), \ldots, f(x_k)$ have a multivariate Gaussian (normal) distribution. It is known that the entire distribution of the random field f is determined by its expectation function $\mu : \mathbb{T}^d \to \mathbb{R}$ and covariance function $C : \mathbb{T}^d \times \mathbb{T}^d \to \mathbb{R}$, defined as

$$\begin{aligned}\mu(x) &:= \mathbb{E}\{f(x)\},\\ C(x,y) &:= \text{Cov}(f(x), f(y))\\ &= \mathbb{E}\{[f(x)-\mu(x)][f(y)-\mu(y)]\}\end{aligned}$$

for all $x, y \in \mathbb{T}^d$. In this paper we will consider f with $\mu \equiv 0$ and with a covariance function

$$C(x,y) = \exp\left(-\frac{||x-y||^2}{\sigma^2}\right). \tag{3.6}$$

For a given Gaussian field f, we will study the percolation phenomena as well as the EC for the sublevel sets, defined as

$$G(\alpha) := \left\{x \in \mathbb{T}^d : f(x) \leqslant \alpha\right\}.$$

Notice that by definition $\{G(\alpha)\}_{\alpha=-\infty}^{\infty}$ is a filtration. This, in particular, implies that we can define the notions of persistent homology and homological percolation for this model as well.

The evaluation of the expected EC for $G(\alpha)$ is the most complicated of all the models in this paper. This was done in Ref. [24] via a formula known as the *Gaussian kinematic formula* (GKF). The fundamental idea behind the GKF is to use Morse theory in a similar way to that of the Boolean model. With some assumptions on the mean $\mu(\cdot)$ and the covariance function $C(\cdot,\cdot)$, one can show that the random function f is a Morse function with probability 1. Briefly, Morse functions are differentiable, and have at most a single critical point at each level α. Evaluating the expected number of critical points then leads to the expected EC of $G(\alpha)$, as in the Boolean model.

The GKF as presented in Ref. [24] covers general Gaussian fields defined on general Riemannian manifolds. In the special case that we examine in this paper, i.e., a Gaussian field on $\mathbb{T}^d$ with the covariance function given in (3.6), the GKF yields the following formula:

$$\bar{\chi}_G(\alpha) := \mathbb{E}\{\chi(G(\alpha)\} = \frac{2}{\omega_d}(2\pi)^{-\frac{d+1}{2}}\mathcal{H}_{d-1}(-\alpha)e^{-\alpha^2/2}, \tag{3.7}$$

where ω_d is the volume of a unit ball in $\mathbb{R}^d$ and $\mathcal{H}_n$ is the Hermite polynomial, given by

$$\mathcal{H}_n(x) = n!\sum_{j=0}^{\lfloor n/2\rfloor}\frac{(-1)^j x^{n-2j}}{j!(n-2j)!2^j}. \tag{3.8}$$

Finally, note that when simulating a Gaussian random field we have to take a discretized grid. The size of this grid will bed noted by n.

IV. RESULTS

In this section we present simulation results for the four models described above, for dimensions $d = 2, 3, 4$. The computations were done using the GUDHI [10] library. The technical details about the simulations can be found in Appendix D.

Remark. Notice that t_k^{perc} was defined in (2.1) as the (nonrandom) critical value for the *probability* of homological percolation to switch form 0 to 1. For the models we are

studying, these phase transition is always defined in the limit as $n \to \infty$. Since we do not have an access to the limit, we use the (random) birth time of the first giant k-cycle, as an approximation to t_k^{perc}.

In Fig. 4 we show the theoretical expected Euler curves, together with the mean appearance of the giant cycles. This figure demonstrates several noteworthy phenomena. Across all models and dimensions, the appearances of the giant cycles align with the zeros of the EC curve. In particular, for the permutahedral complex in the even dimensions [Figs. 4(d) and 4(f)], the giant cycle in the middle dimension aligns perfectly with the corresponding zero of the EC curve. This is a direct consequence of the symmetry between the complex and its complement discussed in Sec. III A 2. Note that in the odd dimension (3) the EC curve for the permutahedral complex is still symmetric, but there is no middle dimension for the giant cycles. Finally, for dimensions 3 and 4, the giant cycles appear before the corresponding zero in the lower dimensions and after the zero in the higher dimensions. Smaller examples in dimension 5 also follow this pattern. This leads to the following conjecture.

Conjecture 4.1. For $d \geqslant 3$ and every $k < d/2$ we have $t_k^{\text{perc}} \leqslant t_k^{\text{ec}}$, while for every $k > d/2$ we have $t_k^{\text{perc}} \geqslant t_k^{\text{ec}}$.

For the middle dimension ($d/2$ when d is even), if the model is symmetric with respect to the parameter t, then the middle giant cycle should align perfectly with the corresponding zero in the EC curve, as in the permutahedral complex. Notice that the GRF model we take is also symmetric (zero mean). We suspect that the tiny difference between $t_{d/2}^{\text{perc}}$ and $t_{d/2}^{\text{ec}}$ [Figs. 4(j) and 4(l)] is due to the fact that we are using a discretized sample of a continuous field. This is supported by the statistics presented below. For asymmetric models (e.g., the cubical and Boolean), it is not clear what should happen for $d/2$. In Figs. 4(a), 4(c), 4(g), and 4(i), we consistently see $t_{d/2}^{\text{ec}} < t_{d/2}^{\text{perc}}$; however, this behavior would have flipped if we were to take the complement objects.

Conjecture 4.1 may have significant implications. For example, in Gaussian random fields, it is not known (for $d \geqslant 3$) whether the percolation thresholds for the super- and sublevel sets are separated, i.e., whether there exists a regime where both the sub- and superlevel sets have a giant component simultaneously. In Ref. [26], we show that t_1^{perc} coincides with the percolation threshold for the sublevel sets, while t_{d-1}^{perc} coincides with the threshold for the superlevel sets. Therefore, given that the expected EC curve is known in a closed form, bounds on the relationship between the zeros of the expected EC curve and the giant cycles would imply separation. The experiments imply a refinment of the above conjecture.

Conjecture 4.2. There is an interlacing of percolation thresholds and zeros of the Euler curve. That is depending on k, either $t_{k-1}^{\text{ec}} \leqslant t_k^{\text{perc}} \leqslant t_{k+1}^{\text{ec}}$ or $t_k^{\text{ec}} \leqslant t_k^{\text{perc}} \leqslant t_{k+1}^{\text{ec}}$.

This type of result would provide rigorous upper and lower bounds on the percolation thresholds. Though the conjectures yield useful results, ultimately, the goal would be to bound the difference between t_k^{perc} and t_k^{ec}. So we next investigate the error terms $\Delta_k = (t_k^{\text{perc}} - t_k^{\text{ec}})$. In Fig. 5 we show some statistics for the examples in Fig. 4. For each of the models we repeated the simulation and estimated the mean and variance for the first birth time of a giant k-cycle. We observe that Δ_k converges rather quickly in the number of points (though

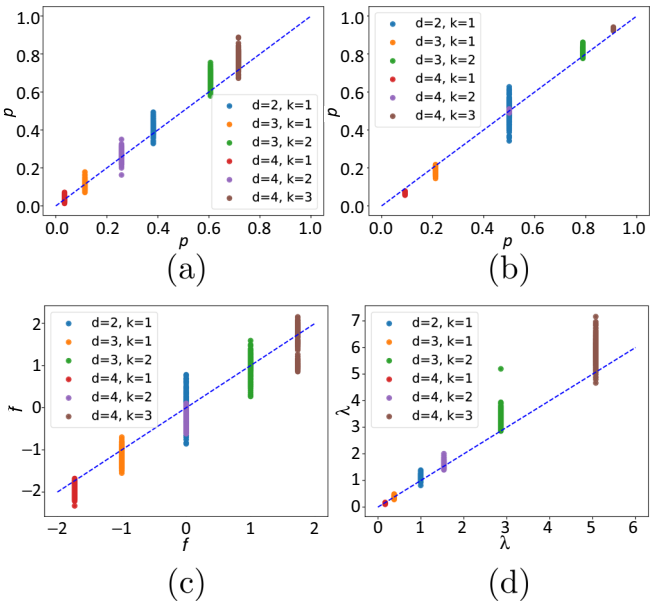

FIG. 6. Appearance of giant cycles vs. zeros of the expected EC curve. The x coordinate of each point is the corresponding zero of the expected EC curve, and the y coordinate is the value when a giant k-cycle appears. (a) Cubical site model; (b) permutahedral site model; (c) Gaussian random field; (d) Poisson-Boolean.

as noted above not necessarily to zero), and the variance becomes small very quickly—note that the x axis is logarithmic (in base e). Furthermore, with the exception of the Gaussian random field, we observe that Δ_k generally remains either always positive or always negative in line with Conjecture IV.1. Note that the standard deviation in the Gaussian random field model remains roughly constant and centered around zero for all grid sizes. This is because the covariance function of the GRF is held constant, indicating that the correlation effects are the key driver of variability in the appearance of the giant cycles.

To give an alternative view on the error term Δ_k in Fig. 6 we show a scatter plot of zeros of the EC curve and the individual appearance of the giant cycles for each of the four models. Note that each plot includes many different values of n, and the large spread in values is due to smaller values of n.

Another quantity that is interesting to consider are the Betti curves (i.e., the evolution of β_k over time). It has been observed in the past that these curves exhibit a "separation" phenomenon, where for each range of parameters a single Betti number dominates all the others (see Fig. 7). The connection of this phenomena and the zeros of the EC have been previously studied as described in Sec. II. This experimentation yields further evidence that these phenomena are related. Therefore, if we consider $\beta_0, \ldots, \beta_{d-1}$, then we can define

$$t_k^{\text{betti}} := \inf \{t : \beta_{k-1}(t) = \beta_k(t)\}.$$

Our simulations show that t_k^{betti} is tightly connected to t_k^{perc} and t_k^{ec}. Further, Figs. 7 and 8 suggest the following conclusions:

(i) The giant k-cycles appear after the peak in β_{k-1} and before the peak in β_k.

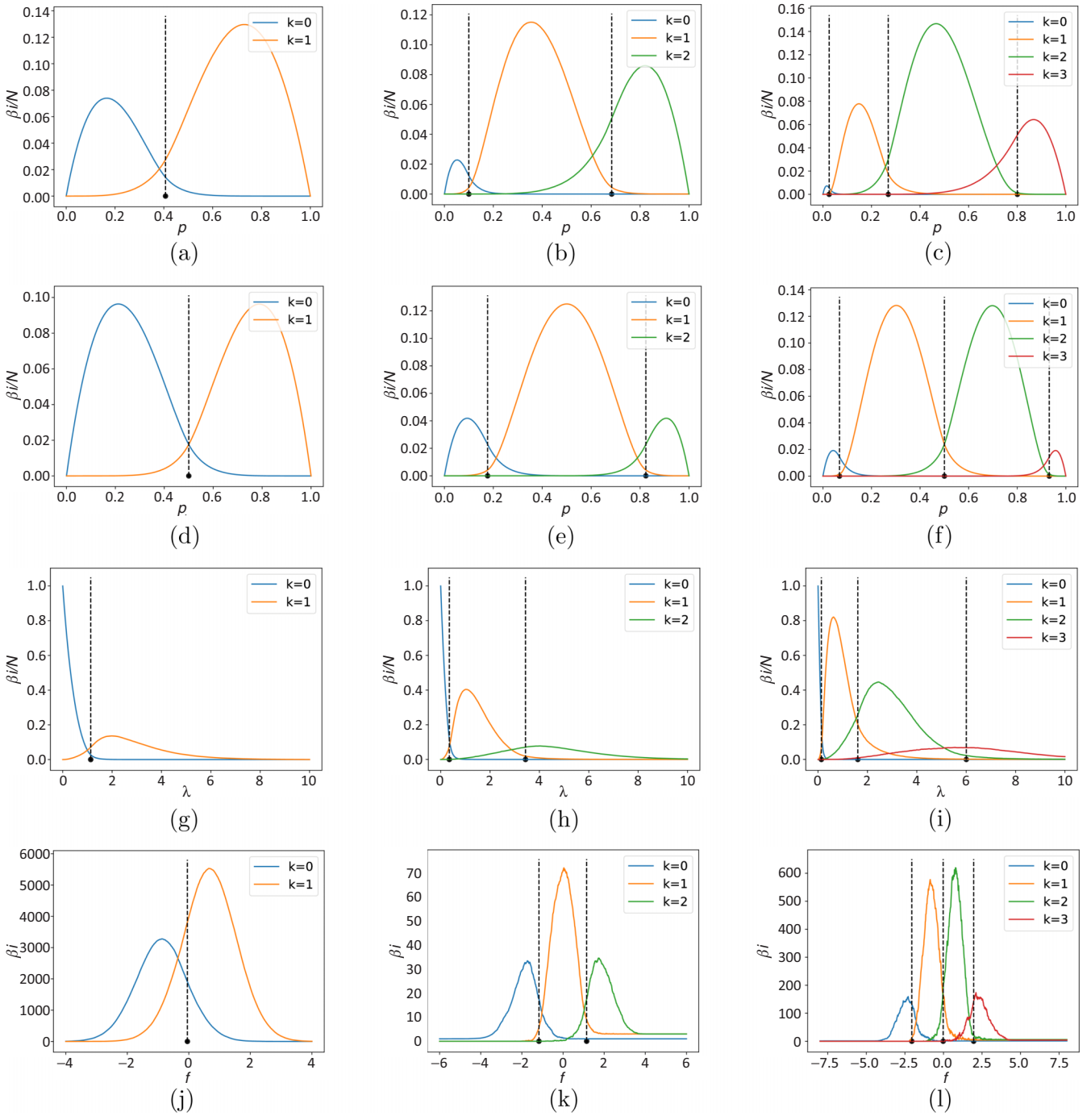

FIG. 7. The empirical (average) Betti curves and the giant cycles. In each plot, we draw the Betti curves (solid line), along with the birth time of the first giant k-cycle for $k = 1, \ldots, d-1$. We simulated all the models on the d-dimensional torus, for $d = 2, 3, 4$ (from left to right). [(a)–(c)] The random cubical complex. [(d)–(f)] The random permutahedral complex. [(g)–(i)] The Boolean model. [(j)–(l)] The Gaussian random field.

(ii) Expanding on previous work, the relationship between the zeros of the EC curve as a good approximations for the t_k^{betti} hold in a wide range of geometric models across multiple dimensions.

If the above could be shown, then one potential path to understanding the difference Δ_k could be through investigating what percentage of $(k-1)$-cycles must be filled before most k simplicies create k-cycles (rather than destroying $(k-1)$-cycles), which is when we expect giant cycles to appear. This is related to the recent study of phase transitions in nongeometric models [15], and an object known as the "giant shadow."

We observe that the simulation results also support the validity of open conjectures about the Betti curves being unimodal [49] in certain models. If the above holds over a large enough set of parameters, the shape of the Euler curve may be used to show unimodality of the Betti curves. We conclude this section by noting that the relationship between

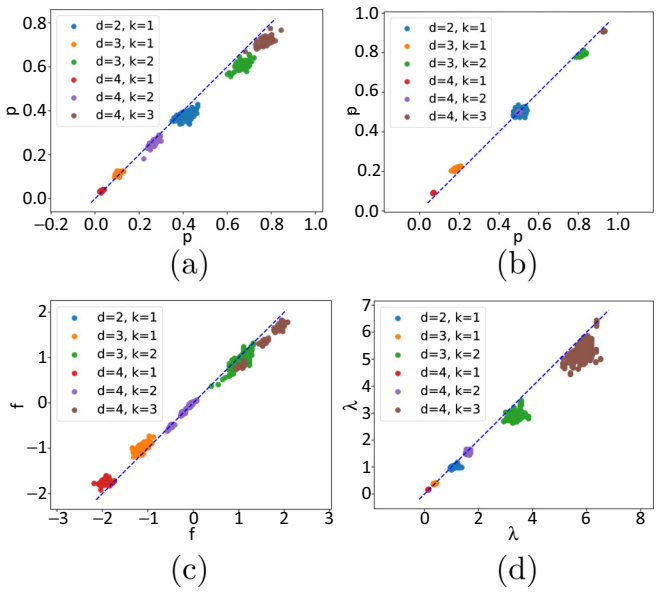

FIG. 8. Appearance of giant cycles vs. emerging of the Betti numbers. The x coordinate of each point corresponds to the appearance time of the kth giant cycle, while the y axis is the time at which the equality $\beta_i = \beta_{i-1}$ occurs. The plots provide strong experimental evidence that the appearance of higher-dimensional cycles occurs in the vicinity of this equality. (a) Cubical site model; (b) permutahedral site model; (c) Gaussian random field; (d) Poisson-Boolean.

appearance of the giant cycles and the zeros of the EC in random models where the Betti curves are known not to be unimodal [29] remains an open question which we plan to address in future work.

V. CONCLUSION

In this paper we defined a new type of percolation phenomena we call "homological percolation" where giant k-cycles appear in the homology of a random structure. Our results suggest a strong connection between the percolation thresholds t_k^{perc} and the zeros of the expected EC curve t_k^{ec}, as well as expanding settings where the latter is connected with t_k^{betti}. This connection is demonstrated in four types of random percolation models across multiple dimensions. It remains an open question as to the how far-reaching the connection between "giant" cycles and the zeros of the EC curve is.

The results in this paper are purely experimental and should serve as the basis for a deep theoretical study to prove the conjectures we made in this paper. Aside from the mathematical challenge of proving these conjectures, they can have significant implication in various fields. For example, for most models in percolation theory the exact thresholds are not known. Therefore, proving an explicit rigorous link between the expected EC curve and the percolation thresholds, will allow us to approximate the thresholds in various models, and perhaps even find their exact values.

Another application is in the field of TDA. A significant effort in TDA is to identify significant topological features in data. If we consider the giant cycles to be significant (as they represent a feature of the true underlying shape), then our conjectures suggest that in order to locate these significant features, we may calculate the EC curve and search for cycles that appear around the corresponding zero of the EC. This heuristic would avoid the need to determine certain constants which are often impossible to compute in practice and should be further developed, once any of the conjectures is proved.

This work raises several interesting research directions. We have studied three distinct phenomena: the emergence of giant cycles, the zeros of the EC, and the crossover of Betti numbers, which all seem to occur in the same vicinity. The giant cycles have a strong dependence on the underlying space, e.g., a giant cycle in a certain dimension may not exist; however, we speculate that the same relationship should hold, i.e., the kth zero should correpond to the k-dimensional giant cycle. A deeper understanding of the precise relationship between these concepts is still at a very early stage. For example, to the best of our knowledge, the giant shadow [15] remains completely unexplored for geometric models. Understanding these phenomena is undoubtedly a fundamental question which will require substantial further research.

ACKNOWLEDGMENTS

The authors are grateful to Yogeshwaran Dhandapani for suggesting the idea of checking for a connection between homological percolation and the Euler characteristic and to Stephen Muirhead for discussions on percolation in Gaussian random field models. O.B. was supported by the Israeli Science Foundation (Grant No. 1965/19). P.S. was supported by SSHRC Canada (NFRFE-2018-00431) and the Alan Turing Institute - Defense and Security Programme (D015).

APPENDIX A: CALCULATING THE EC FOR SITE-PERCOLATION MODELS

Here we present the details of the calculation of the expected Euler characteristic curve. Recall, that for the site percolation models the filtration parameter is $t = p$. To establish a formula for $\bar{\chi}(p)$, we use Eq. (3.1) and the linearity of expectation. Thus, we need to evaluate the expected number of k faces in each of the site models.

Beginning with the cubical model, we observe that each d-dimensional cube has $2^{d-k}\binom{d}{k}$ k faces on its boundary. In addition, since we consider Q_n^d (the discretization of $\mathbb{T}^d$ into n boxes), each k face is on the boundary of precisely 2^{d-k} d-dimensional boxes. Therefore, the total number of k faces in Q_n^d is exactly $n\binom{d}{k}$.

Now, for any k face, if it is included in $Q(n, p)$, then at least one of d-dimensional boxes that contains it must be open. Therefore, the probability of a k faces to be in $Q(n, p)$ is $(1 - (1 - p)^{2^{d-k}})$. Putting everything together, we have that

$$\mathbb{E}\{F_k(Q(n, p))\} = n\binom{d}{k}[1 - (1 - p)^{2^{d-k}}],$$

which then yields (3.2),

$$\begin{aligned}\bar{\chi}_Q(p) &= \sum_{k=0}^{d}(-1)^k\mathbb{E}\{F_k[Q(n, p)]\} \\ &= n\sum_{k=0}^{d}(-1)^k\binom{d}{k}[1 - (1 - p)^{2^{d-k}}].\end{aligned}$$

The calculation for the permutahedral complex $P(n,p)$ is similar, where the only difference is the face counting. From Ref. [44], each k face in P_n^d corresponds to a partition of the set $\{0,\dots,d\}$ into $d+1-k$ nonempty parts. Therefore, the number of k faces for each cell is given by a Stirling number of the second kind [50],

$$\begin{aligned} F(P_1^d) &= (d+1-k)!S(d+1,d+1-k) \\ &= \sum_{i=0}^{d+1-k}(-1)^i\binom{d+1-k}{i}(d+1-k-i)^{d+1} \\ &= \sum_{j=0}^{d+1-k}(-1)^{d+1-k-j}\binom{d+1-k}{j}j^{d+1} \end{aligned}$$

where the second equality follows from using the substitution $j=d+1-k-i$.

Now every k face belongs to $(d+1-k)$ d cells. This follows from the genericity of the corresponding Voronoi cells (see the proof of Lemma C.1 in the Appendix C). Hence, the total number of k cells is

$$\begin{aligned} F_k(P_n^d) &= \tfrac{n}{d+1-k}\sum_{j=0}^{d+1-k}(-1)^{d+1-k-j}\tbinom{d+1-k}{j}j^{d+1} \\ &= n\sum_{j=0}^{d+1-k}(-1)^{d+1-k-j}\tbinom{d-k}{j}j^{d}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{E}\{\chi_P(n,p)\} = n\sum_{k=0}^{d}(-1)^k\left[1-(1-p)^{d+1-k}\right] \\ \times\sum_{j=0}^{d+1-k}(-1)^{d+1-k-j}\binom{d-k}{j}j^{d+1}. \end{aligned}$$

Exchanging between k and $(d-k)$ then yields (3.3).

APPENDIX B: THE EXPECTED EC CURVE FOR THE GAUSSIAN RANDOM FIELD

As stated in Section III C, the expected EC is calculated via the Gaussian Kinematic Formula, developed in Ref. [24]. Suppose that M is a d-dimensional manifold, and let $f: M\to\mathbb{R}$ be a Gaussian random field with zero mean and unit variance (with some further smoothness conditions detailed in Ref. [24]). Let $D_u=[u,\infty)$, then $f^{-1}(D_u)$ is a superlevel set of f. The GKF (Theorem 4.1 in Ref. [24]) then states that

$$\mathbb{E}\left\{\chi[f^{-1}(D_u)]\right\} = \sum_{j=0}^{d}(2\pi)^{-j/2}\mathcal{L}_j(M)\mathcal{M}_j(D_u),$$

where $\mathcal{L}_j(M)$ are geometric functionals of M known as the *Lipschitz-Killing curvatures* and M_j is slightly different object known is the *Gaussian-Minkowski functional*. For the special case where $M=\mathbb{T}^d$ it can be shown that $\mathcal{L}_j(\mathbb{T}^d)=0$ for all $j<d$, and $\mathcal{L}_d(\mathbb{T}^d)=2/\omega_d$. In addition, in Ref. [24] it is shown that $M_j(D_u)=(2\pi)^{-1/2}\mathcal{H}_{j-1}(u)e^{-u^2/2}$, where $\mathcal{H}_n(u)$ are the Hermite polynomial in (3.8). Thus, we have

$$\mathbb{E}\left\{\chi[f^{-1}(D_u)]\right\} = \frac{2}{\omega_d}(2\pi)^{-\frac{d+1}{2}}\mathcal{H}_{d-1}(u)e^{-u^2/2}.$$

Finally, recall that we defined $G(\alpha)$ as the sublevel sets of f. In addition, since f is a zero-mean Gaussian field, we have that $f(x)$ and $\tilde{f}(x):=-f(x)$ have the same distribution. Since the sublevel sets of f are the superlevel sets of $\tilde{f}$, we have

$$\bar{\chi}_G(\alpha) := \mathbb{E}\{\chi[\tilde{f}^{-1}(D_{-\alpha})]\} = \mathbb{E}\{\chi[f^{-1}(D_{-\alpha})]\},$$

and therefore

$$\bar{\chi}_G(\alpha) := \mathbb{E}\{\chi(G(\alpha)\} = \frac{2}{\omega_d}(2\pi)^{-\frac{d+1}{2}}\mathcal{H}_{d-1}(-\alpha)e^{-\alpha^2/2}.$$

APPENDIX C: SYMMETRY AND DUALITY FOR THE PERMUTAHEDRAL COMPLEX

In this section we provide formal proofs for the statements on symmetry that are discussed in Sec. III A 2. For the case of the site percolation on a hexagonal grid, the symmetry around $p=1/2$ is well known. Here we extend it to arbitrary dimension, but note that the proofs assume some familiarity with algebraic topology.

The idea behind the proofs is to relate a subspace of a manifold (in this case, d torus), with its complement. Informally, the topology of the manifold and a subspace determine the topology of the complement. The most well-known example of this is Alexander duality, which relates the kth homology of a subspace with the $d-k$ cohomology of the complement. In our setting, we consider the Betti numbers so there is no distinction between homology and cohomology. Before getting to the duality, there is a technical obstacle to overcome. In the site models we consider, taking the complement of the open sites is not the same as considering the union of the closed sites, but rather it is equivalent to the *closure* of the complement. This difference can change the topology as in the case of the cubical complex, as can be seen in Fig. 9. Hence, we first prove that the complement and the closure of the complement are equivalent. As in Sec. III A 2, let $P\subseteq P_n^d$ and $P^c=\mathrm{cl}(P_n^d\backslash P)$.

Lemma C.1. For all $0\leqslant k\leqslant d$,

$$H_k(P^c)\cong H_k(P_n^d\backslash P).$$

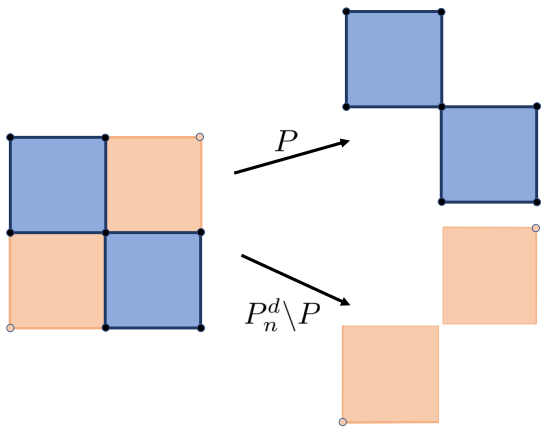

FIG. 9. An example of a failure of symmetry for the cubical complex. The dark squares indicate open sites. The complement of the open sites is different (bottom right) than the union closed sites. In particular, the latter is connected (via the point in the middle), while the former is not.

Proof. To prove this lemma, we prove a stronger statement, namely that the P^c and $P_n^d\backslash P$ are homotopy equivalent. First, consider the open cover induced by the sites in P^c, denoted by $\mathcal{U}$. That is, each element in the cover is an open neighborhood of each site. Since the sites are convex, it follows $\mathcal{U}$ is a good cover and hence P^c is homotopy equivalent to the nerve of the cover, $\mathcal{NU}$.

As noted in Sec. III A 2, each site is a permutahedron of order $d+1$. The interior of each site corresponds to the top-dimensional cell of the permutahedra which are the same for $P_n^d\backslash P$ and P^c. The two differ in that $P_n^d\backslash P$ is does not have lower-dimensional faces (of the sites) which are adjacent to sites both in P^c and in P. Taking the same open cover as above, but on $P_n^d\backslash P$, denoted by $\mathcal{U}'$. We show that this is a good cover *and* that the nerves are the same, which implies the result.

First, we note that there is a one-to-one correspondence between the $(d-k)$ faces of permutahedron and k simplices of the nerve [51]. That is, each intersection of $(k+1)$ cells corresponds to a $(d-k)$ face of the permutahedron. For any $(d-k)$ face τ in $P_n^d\backslash P$, it must be adjacent to $(k+1)$ sites in $P_n^d\backslash P$ and so cannot be adjacent to any sites in P. Note that lower-dimensional faces of τ (which are in the closure of τ) may be missing from $P_n^d\backslash P$ and so it is not convex. It, however, remains star shaped and hence the $(k+1)$ intersection of cover elements is contractible, implying that $\mathcal{U}'$ is a good cover.

The same argument also shows that any face which is in P^c but not $P_n^d\backslash P$, does not affect the nerve as the corresponding interesction remains nonempty. Note that in the $2D$ cubical complex a pairwise intersection may correspond to a vertex rather than an edge which breaks the argument above.

Hence the $\mathcal{NU} = \mathcal{NU}'$ completing the proof.

The above lemma allows us to use P^c and $P_n^d\backslash P$ interchangeably. We can now prove Lemma 3.1.

Lemma C.2 (3.1). For $0 \leqslant k \leqslant d$,

$$\mathcal{B}_k(P) + \mathcal{B}_{d-k}(P^c) = \beta_k(\mathbb{T}^d).$$

Proof. In this proof, we use P^c in place of the complement of P as they are equivalent by Lemma C.1. There exists a commutative diagram where the rows are exact, due to the long exact sequence for relative (co) homology.

$$\begin{array}{ccccccc} \mathrm{H}_k(P) & \xrightarrow{i_*} & \mathrm{H}_k(P_n^d) & \longrightarrow & \mathrm{H}_k(P_n^d, P) & \xrightarrow{\delta_k} & \mathrm{H}_{k-1}(P) \\ \cong\uparrow & & \cong\uparrow & & \cong\uparrow & & \cong\uparrow \\ \mathrm{H}^{d-k}(P_n^d, P^c) & \longrightarrow & \mathrm{H}^{d-k}(P_n^d) & \xrightarrow{j^*} & \mathrm{H}^{d-k}(P^c) & \xrightarrow{\delta^{d-k}} & \mathrm{H}^{d-k+1}(P_n^d, P^c) \end{array} \tag{C1}$$

The leftmost isomorphism follows from the Lefschetz duality, the second from Poincare duality, and the third from the Five Lemma. Note that a detailed proof can be found in Ref. [36]. We can decompose the full space as

$$H_k(P_n^d) \cong \mathrm{Im}\, i_* \oplus \mathrm{coker}\, i_*, \tag{C2}$$

and by exactness and a diagram chase, we have that $\mathrm{coker}\, i_* \cong \mathrm{Im}\, j^*$. We observe that

$$\mathcal{B}_k(P) = \dim(i_*),$$
$$\mathcal{B}_{d-k}(P^c) = \dim(j^*),$$

where the second equality follows from the equivalence for ranks of homology and cohomology over fields. Substituting into Eq. (C2), we obtain the result

$$\beta_k(\mathbb{T}^d) = \beta_k(P_n^d) = \mathcal{B}_k(P) + \mathcal{B}_{d-k}(P^c).$$

We conclude by proving the symmetry of the Euler curve.

Lemma C.3. For the permutahedral complex we have the following symmetry:

$$\chi_P(p) = (-1)^d \chi_P(1-p).$$

Proof. Since $P^c(n, 1-p) \sim P(n, p)$, it suffices to show that $\chi(P) = (-1)^d \chi(P^c)$ for some P. Consider the diagram (C1). By exactness,

$$\beta_k(P) = \dim(\mathrm{Im}\, i_*(k)) + \dim(\mathrm{Im}\, \delta_{k+1}),$$
$$\beta_{d-k}(P^c) = \dim[\mathrm{Im}\, j^*(d-k)] + \dim(\mathrm{Im}\, \delta^{d-k}).$$

Note that we have added the dimension to the notation of the corresponding morphisms, i.e., $i_*(k) : H_k(P) \to H_k(P_n^d)$. Furthermore, the diagram implies $\dim(\mathrm{Im}\, \delta^{d-k}) = \dim(\mathrm{Im}\, \delta_k)$. Computing the Euler characteristic yields

$$\begin{aligned} \chi(P) &= \sum_{k=0}^{d} (-1)^k \beta_k(P) \\ &= \sum_{k=0}^{d} (-1)^k \big[\dim(\mathrm{Im}\, i_*(k)) + \dim(\mathrm{Im}\, \delta_{k+1})\big] \\ &= \sum_{k=0}^{d} (-1)^k \left\{ \binom{d}{d-k} - \dim[\mathrm{Im}\, j^*(d-k)] \right. \\ &\quad \left. + \dim(\mathrm{Im}\, \delta_{d-k-1}) \right\} \\ &= \sum_{k=0}^{d} (-1)^k \big\{\dim(\mathrm{Im}\, \delta^{d-k-1}) - \dim[\mathrm{Im}\, j^*(d-k)]\big\} \\ &= \sum_{k=0}^{d} (-1)^k [\dim(\mathrm{Im}\, \delta^{d-k-1}) - \beta_{d-k}(P^c) \\ &\quad + \dim(\mathrm{Im}\, \delta^{d-k})] \\ &= (-1)^d \sum_{k=0}^{d} \beta_k(P^c) + \sum_{k=0}^{d} (-1)^k [\dim(\mathrm{Im}\, \delta^{d-k-1}) \\ &\quad + \dim(\mathrm{Im}\, \delta^{d-k})] \\ &= (-1)^d \chi(P^c) + \dim(\mathrm{Im}\, \delta^d) - \dim(\mathrm{Im}\, \delta^{-1}) \\ &= (-1)^d \chi(P^c), \end{aligned}$$

where for the last inequality, we use that $\dim(\mathrm{Im}\,\delta^d) = \dim(\mathrm{Im}\,\delta^{-1}) = 0$.

APPENDIX D: SIMULATION DETAILS

Here we present some implementational details of the simulations. The persistence diagrams and hence Betti curves and appearance of the giant cycles were computed with GUDHI [10]. For the cubical site model, a n^d grid with periodic connectivity was used for varying values of n. After generating a uniform random function taking values in [0,1], with the values assigned to the top-dimensional cells. Persistence was then computed directly on the resulting cubical complex. We note that the Gaussian random field was also approximated on a cubical grid according to the method described in Ref. [52]. The resulting GRF was always generated on the unit torus, with $\sigma^2 = 10^{-3}$.

The permutahedral complex was built by constructing a set of points in the A_d^* grid embedded in $\mathbb{R}^{d+1}$. The 1-skeleton was then built by choosing an appropriate radius, so that all the neighbors were connected. Note that this can be thought of as the 1-skeleton of the Delaunay complex of the pointset. The points, along wiht their adjacent edges, were then identified with the outgoing edges appropriately identified. This embeds the 1-skeleton in $\mathbb{T}^d$. The full complex was then computed using clique completion to higher dimensions. Again a random function was assigned to the sites, which correspond to the vertices of the resulting complex. Persistent homology of the *lower-star filtration* on this complex was then computed.

Finally for the Boolean model, in dimensions 2 and 3, the α filtration [53] on the torus was used, whereas for dimension 4, for each sample point set, the connectivity threshold was first computed and was then used to compute the threshold for the construction of the Čech filtration.

[1] H. Edelsbrunner and J. Harer, Persistent homology—A survey, Contemp. Math. **453**, 257 (2008).

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

[3] H. Edelsbrunner and D. Morozov, Persistent homology: Theory and practice, in *Proceedings of the European Congress of Mathematics* (European Mathematical Society, Berlin, Germany, 2014), pp. 31–50.

[4] A. Zomorodian and G. Carlsson, Computing persistent homology, Discr. Comput. Geom. **33**, 249 (2005).

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

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

[7] G. Singh, F. Memoli, T. Ishkhanov, G. Sapiro, G. Carlsson, and D. L. Ringach, Topological analysis of population activity in visual cortex, J. Vision **8**, 11 (2008).

[8] R. J. Adler, S. Agami, and P. Pranav, Modeling and replicating statistical topology and evidence for CMB nonhomogeneity, Proc. Natl. Acad. Sci. U.S.A. **114**, 11878 (2017).

[9] D. Horak, S. Maletić, and M. Rajković, Persistent homology of complex networks, J. Stat. Mech.: Theory Exp. (2009) P03034.

[10] C. Maria, J.-D. Boissonnat, M. Glisse, and M. Yvinec, The gudhi library: Simplicial complexes and persistent homology, in *Proceedings of the International Congress on Mathematical Software* (Springer, Berlin, 2014), pp. 167–174.

[11] H. Duminil-Copin, Sixty years of percolation, in *International Congress of Mathematicians 2018, Rio de Janeiro, Brazil, 1–9 August 2018*, edited by B. Sirakov, P. N. de Souza, and M. Viana (World Scientific, Singapore, 2018), p. 5396.

[12] G. Bianconi and R. M. Ziff, Topological percolation on hyperbolic simplicial complexes, Phys. Rev. E **98**, 052308 (2018).

[13] M. Kahle and E. Meckes, Limit the theorems for Betti numbers of random simplicial complexes, Homol. Homotopy Appl. **15**, 343 (2013).

[14] N. Linial and R. Meshulam, Homological connectivity of random 2-complexes, Combinatorica **26**, 475 (2006).

[15] N. Linial and Y. Peled, On the phase transition in random simplicial complexes, Ann. Math. **184**, 745 (2016).

[16] R. J. Adler and J. E. Taylor, *Random Fields and Geometry*, Vol. 115 (Springer, Berlin, 2007).

[17] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics (John Wiley & Sons Ltd., Chichester, 1987).

[18] W. N. Colley, J.-R. Gott, III, and C. Park, Topology of cobe microwave background fluctuations, Mon. Not. R. Astron. Soc. **281**, L82 (1996).

[19] A. Kogut, A. J. Banday, C. L. Bennett, K. M. Gorski, G. Hinshaw, G. F. Smoot, and E. L. Wright, Tests for non-gaussian statistics in the dmr four-year sky maps, Astrophys. J. Lett. **464**, L29 (1996).

[20] K. J. Worsley, Boundary corrections for the expected Euler characteristic of excursion sets of random fields, with an application to astrophysics, in *Advances in Applied Probability* (Cambridge University Press, Cambridge, UK, 1995), pp. 943–959.

[21] J. E. Taylor and K. J. Worsley, Detecting sparse signals in random fields, with an application to brain mapping, J. Am. Stat. Assoc. **102**, 913 (2007).

[22] K. J. Worsley, Testing for signals with unknown location and scale in a χ 2 random field, with an application to fMRI, in *Advances in Applied Probability* (Cambridge University Press, Cambridge, UK, 2001), pp. 773–793.

[23] O. Bobrowski and S. Weinberger, On the vanishing of homology in random čech complexes, Rand. Struct. Alg. **51**, 14 (2017).

[24] J. E. Taylor and R. J. Adler, Gaussian processes, kinematic formulas and Poincaré's limit, Ann. Probab. **37**, 1459 (2009).

[25] R. A. Neher, K. Mecke, and H. Wagner, Topological estimation of percolation thresholds, J. Stat. Mech.: Theory Exp. (2008) P01011.

[26] O. Bobrowski and P. Skraba, Homological percolation in the poisson boolean model (unpublished).

[27] S. Mertens and R. M. Ziff, Percolation in finite matching lattices, Phys. Rev. E **94**, 062152 (2016).

[28] F. A. N. Santos, E. P. Raposo, M. D. Coutinho-Filho, M. Copelli, C. J. Stam, and L. Douw, Topological phase transitions in functional brain networks, Phys. Rev. E **100**, 032414 (2019).

[29] P. Pranav, H. Edelsbrunner, R. van de Weygaert, G. Vegter, M. Kerber, B. J. T. Jones, and M. Wintraecken, The topology of the cosmic web in terms of persistent Betti numbers, Mon. Not. Roy. Astron. Soc. **465**, 4281 (2017).

[30] P. Pranav, R. Van de Weygaert, G. Vegter, B. J. T. Jones, R. J. Adler, J. Feldbrugge, C. Park, T. Buchert, and M. Kerber, Topology and geometry of gaussian random fields i: On betti numbers, euler characteristic, and minkowski functionals, Mon. Not. R. Astron. Soc. **485**, 4167 (2019).

[31] V. Robins, M. Saadatfar, O. Delgado-Friedrichs, and A. P. Sheppard, Percolating length scales from topological persistence analysis of micro-ct images of porous materials, Water Resour. Res. **52**, 315 (2016).

[32] J. Feldbrugge, M. van Engelen, R. van de Weygaert, P. Pranav, and G. Vegter, Stochastic homology of gaussian vs. non-gaussian random fields: Graphs towards betti numbers and persistence diagrams, J. Cosmol. Astropart. Phys. 09 (2019) 052.

[33] L. Speidel, H. A. Harrington, S. J. Chapman, and M. A. Porter, Topological data analysis of continuum percolation with disks, Phys. Rev. E **98**, 012318 (2018).

[34] O. Bobrowski and G. Oliveira, Random čech complexes on riemannian manifolds, Rand. Struct. Alg. **54**, 373 (2019).

[35] O. Bobrowski, Homological connectivity in Random čech complexes, arXiv:1906.04861 (2019).

[36] A. Hatcher, *Algebraic Topology* (Cambridge University Press, Cambridge, 2002).

[37] H. Edelsbrunner and J. L. Harer, *Computational Topology: An Introduction* (American Mathematical Society, Providence, 2010).

[38] L. Wasserman, Topological data analysis, Annu. Rev. Appl. **5**, 501 (2018).

[39] H. Duminil-Copin, D. Ioffe, Y. Velenik *et al.*, A quantitative burton–keane estimate under strong fkg condition, Ann. Probab. **44**, 3335 (2016).

[40] E. Richardson and M. Werman, Efficient classification using the Euler characteristic, Pattern Recogn. Lett. **49**, 99 (2014).

[41] K. J. Worsley, Estimating the number of peaks in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images, in *The Annals of Statistics* (Institute of Mathematical Statistics, Hayward, 1995), pp. 640–669.

[42] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, Vol. 290 (Springer Science & Business Media, Berlin, 2013).

[43] J. Baek and A. Adams, Some useful properties of the permutohedral lattice for gaussian filtering, *In other words*, 10(1):0, (2009), https://graphics.stanford.edu/papers/permutohedral/permutohedral_techreport.pdf.

[44] G. M. Ziegler, *Lectures on Polytopes*, Vol. 152 (Springer Science & Business Media, Berlin, 2012).

[45] M. Bogdanov, M. Caroli, and M. Teillaud, Computing periodic triangulations, in *Shape Up—Exercises in Materials Geometry and Topology* (2015), pp. 60–61, https://hal.inria.fr/hal-01224549/document.

[46] G. R. Grimmett, *Percolation*, Grundlehren der mathematischen Wissenschaften, Vol. 321 (Springer-Verlag, Berlin, Heidelberg, 1999), 2nd ed.

[47] M. Penrose, *Random Geometric Graphs*, Vol. 5 (Oxford University Press, Oxford, 2003).

[48] O. Bobrowski and S. Mukherjee, The topology of probability distributions on manifolds, Probab. Theory Relat. Fields **161**, 651 (2014).

[49] M. Kahle, Random geometric complexes, Discr. Comput. Geom. **45**, 553 (2011).

[50] R. L. Graham, D. E. Knuth, O. Patashnik, and S. Liu, Concrete mathematics: A foundation for computer science, Comput. Phys. **3**, 106 (1989).

[51] A. Choudhary, M. Kerber, and S. Raghvendra, Polynomial-sized topological approximations using the permutahedron, Discr. Comput. Geom. **61**, 42 (2019).

[52] A. T. A. Wood and G. Chan, Simulation of stationary gaussian processes in [0, 1] d, J. Comput. Graph. Stat. **3**, 409 (1994).

[53] H. Edelsbrunner, Alpha shapes—A survey, Tessell. Sci. **27**, 1 (2010).