

Persistent Homology in Two-Dimensional Atomic Networks

David Ormrod Morley,¹ Philip S. Salmon,² and Mark Wilson¹

¹*Department of Chemistry, Physical and Theoretical Chemistry Laboratory,
University of Oxford, South Parks Road, Oxford OX1 3QZ, UK*

²*Department of Physics, University of Bath, Bath BA2 7AY, UK*

(Dated: February 18, 2021)

The topology of two-dimensional network materials is investigated by persistent homology analysis. The constraint of two dimensions allows for a direct comparison of key persistent homology metrics (persistence diagrams, cycles, Betti numbers) with more traditional metrics such as the ring-size distributions. Two different types of networks are employed in which the topology is manipulated systematically. In the first, comparatively rigid networks are generated for a triangle-raft model, which are representative of materials such as silica bilayers. In the second, more flexible networks are generated using a bond-switching algorithm, which are representative of materials such as graphene. Bands are identified in the persistence diagrams by reference to the length-scales associated with distorted polygons. The triangle-raft models with the largest ordering allow specific bands B_n ($n = 1, 2, 3, \dots$) to be allocated to configurations of atoms separated by n bonds. The persistence diagrams for the more disordered network models also display bands albeit less pronounced. The persistent homology method thereby provides information on n -body correlations that is not accessible from structure factors or radial distribution functions. An analysis of the persistent cycles gives the primitive ring statistics, provided the level of disorder is not too large. The method also gives information on the regularity of rings that is unavailable from a ring-statistics analysis. The utility of the persistent homology method is demonstrated by its application to experimentally-obtained configurations of silica bilayers and graphene.

I. INTRODUCTION

The development of efficient methods for identifying structure in large and complex data sets is becoming increasingly important in the age of “big data”. The field of topological data analysis has emerged to address this need, driving the development of statistical tools to analyse and understand big, and potentially noisy, real-world data [1]. One tool in particular, persistent homology, has received significant attention [2]. Persistent homology is concerned with finding the fundamental topological features of a given set of points. Here *homology* means counting the number of connected components and cycles in a system. The persistence aspect arises as these topological features can be calculated across different length scales, with those existing across multiple length scales being identified as more *persistent*. The more persistent a topological feature, the more it can be thought of as reflecting the true underlying system topology.

The application of persistent homology to atomistic materials seems natural. Atomic systems have long been thought of in terms of collections of point-like particles, and the study of the emergent structure is deep-rooted in the field. Any process that can elucidate as yet hidden structure in materials, or improve its description, naturally has potential to be extremely useful. In this vein, persistent homology has already been applied to diverse topics such as granular matter [3], porous media [4], water networks [5], fullerenes [6] and, of particular importance to this work, amorphous materials [7–11]. The latter studies claim to highlight structures that are not available using more conventional techniques, quantify the medium range order in glass, and explain phenom-

ena such as the origin of the first sharp diffraction peak in disordered materials. Persistent homology remains, however, a qualitative descriptor, with some doubt as to the “added value” from this technique.

In this paper the utility of persistent homology will be assessed using two-dimensional amorphous materials. The interpretation of the properties of these reduced dimensional systems is simpler compared to three dimensional networks, and should bring insight to the analysis of the latter. The paper begins by outlining the relevant theory to persistent homology, with the aid of small example systems. Persistent homology is then calculated for triangle rafts (a proxy for amorphous bilayers of, *e.g.*, silica), systematically generated with increasing levels of disorder. The results are contrasted with those obtained from generic random-networks as produced from bond switching. In addition, the persistence diagrams are discussed for configurations obtained directly from experiments on silica bilayers and graphene. Finally, the conclusions of these investigations are assessed in the context of more complex systems, and the potential utility of persistent homology as an analytic tool for materials characterisation is examined.

II. OVERVIEW OF TRADITIONAL METHODS

Amorphous network-forming materials are usually built from simple local-structural-units such as the SiO_4 tetrahedron in amorphous silica. Bond-length and bond-bending flexibility leads, however, to structural complexity as these units connect to form configurations on longer length scales [12]. This complexity affects the network

properties, but it is difficult to identify suitable metrics that capture its true nature. Typically, the ordering on a medium or intermediate length scale is associated with the formation of rings of variable size, where a ring is identified by finding a closed path of sequentially linked network nodes. In amorphous silica, for example, the Si atoms provide the network nodes and the closed path is found by following the Si-O bonds. In this paper, we will consider only primitive rings, *i.e.*, those that cannot be subdivided into two smaller rings [13]. A ring can be classified by its size, k , denoting the number of constituent nodes and the distribution of ring sizes is termed the ring statistics, p_k .

In persistent homology, cycles from the first homology group (see section III B) also involve a set of network nodes that enclose a two-dimensional hole. In this analysis, the cycles are identified from the point-cloud provided by the atomic positions without reference to the nature of the chemical bonds that link those atoms. The information gained on the network topology can therefore be different to that obtained by identifying the primitive rings found in a more “traditional” approach. In a three-dimensional system, a comparison of cycles versus rings is complicated by the difficulty in visualising where they sit within the network structure. This obstacle does not, however, manifest itself in two-dimensions where the ring structure is straightforward to visualise. It will be shown that the rings often correspond to the most persistent cycles found using persistent homology, although this is not a requirement.

III. BACKGROUND TO PERSISTENT HOMOLOGY

This section will give a broad outline of the mathematical background behind persistent homology, but with the focus on a practicable implementation for studying physical two-dimensional networks. It will begin by introducing the concept of a filtered simplicial complex, before discussing homology groups, persistent homology and the visualisation of persistence. Actual numerical persistent homology calculations were carried out with the GUDHI library [14].

A. Filtered Simplicial Complexes

The input for a persistent homology calculation is simply a point cloud, *i.e.*, a set of coordinates in Euclidean space. For materials, this corresponds to the set of atomic positions, $\mathbf{r} = \{r_1, r_2, \dots, r_N\}$, as obtained from simulation or, for instance, experimental microscopy. Using this point cloud, a filtered simplicial complex can be constructed [15]. To explain what is meant by a filtered simplicial complex, it is useful to break down this definition further:

- An m -simplex, σ , is a subset of the total points, $\sigma \subseteq \mathbf{r}$, with dimension $m = |\sigma| - 1$. Put simply, a 0-simplex corresponds to a point, a 1-simplex to a line, a 2-simplex to a triangle, a 3-simplex to a tetrahedron, and so on.
- A simplicial complex, \mathbf{K} , is then a set of simplices, $\mathbf{K} = \{\sigma_1, \sigma_2, \dots\}$.
- A filtration of a simplicial complex, is a sequence of subcomplexes: $\mathbf{K}_a \subseteq \mathbf{K}_b \subseteq \dots \subseteq \mathbf{K}$, where each subcomplex, \mathbf{K}_ϵ , occurs at an increasing filtration value, denoted ϵ (see below).

For a given point cloud, there are many ways to construct a filtered simplicial complex (Cech, Vietoris-Rips, Witness *etc.*), each method identifying simplices and calculating filtration values in a different manner [16]. In this work, for reasons that will become apparent, the *alpha complex* is used.

To illustrate these concepts, an introductory example with just three points is given in figure 1. Considering only the simplices at first: figure 1(a) contains three 0-simplices, the points A, B, C ; figures 1(b) and 1(c) three 1-simplices, the lines AB, AC, BC ; and figure 1(d) a 2-simplex, the triangle ABC . Once the simplices have been identified, a filtration value, ϵ , is computed and assigned to each. For the alpha complex, filtration values are determined from the radius of the circumcircles containing the simplices. The specific algorithm is as follows:

- 0-simplices: have a filtration value of $\epsilon = 0$.
- 1-simplices: provided the circumcircle is empty, they have a filtration value equal to its radius, which for two points i and j is equivalent to half the line length $\epsilon = r_{ij}/2$. If the circumcircle contains a 2-simplex, then the filtration value is set to the filtration value of that simplex.
- 2-simplices: have a filtration value equal to the circumradius.

Having calculated the filtration values for each simplex, a subcomplex can be generated, \mathbf{K}_ϵ , which contains only the simplices with filtration values less than or equal to ϵ . Finally, the filtered simplicial complex emerges as a sequence of subcomplexes at increasing filtration values from $\epsilon = 0 \rightarrow \infty$. Again, figure 1 shows this process for the case of three points. The first subcomplex, \mathbf{K}_0 , will consist only of a set of discrete points (figure 1(a)). As the filtration value increases, higher dimensionality simplices will be included, as lines and triangles form (figure 1(b)-(d)). The last subcomplex, \mathbf{K}_∞ , will contain all the determined simplices (figure 1(d)). Crucially, for the alpha complex, this will be equivalent to the Delaunay triangulation, which is often used to characterise two-dimensional systems [17]. This result provides the motivation for choosing the alpha complex, as the Delaunay triangulation (which is the dual of the Voronoi diagram) is well-defined in two dimensions. As such it provides

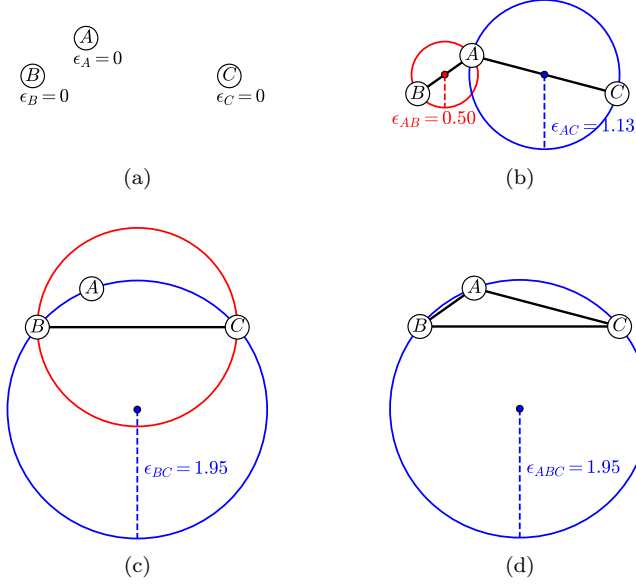


FIG. 1. Construction of an alpha simplicial complex and its filtration. Panel (a) shows three 0-simplices, $\sigma_1 = \{A\}$, $\sigma_2 = \{B\}$ and $\sigma_3 = \{C\}$, with filtration values of $\epsilon = 0$; $\mathbf{K}_0 = \{\sigma_1, \sigma_2, \sigma_3\}$. Panel (b) shows an additional two 1-simplices $\sigma_4 = \{A, B\}$ and $\sigma_5 = \{A, C\}$, with filtration values given by the radii of the respective circumcircles (dashed lines in circles); $\mathbf{K}_{1.13} = \{\sigma_1, \sigma_2, \dots, \sigma_5\}$. Panel (c) shows an additional different 1-simplex $\sigma_6 = \{B, C\}$, in which the circumcircle (red circle) contains the 2-simplex $\{A, B, C\}$, and so the filtration value is set to the value of this 2-simplex, *i.e.*, $\epsilon_{BC} = \epsilon_{ABC}$ (blue circle). Panel (d) shows the 3-simplex $\sigma_7 = \{A, B, C\}$ with a filtration value given by the circumradius; $\mathbf{K}_{1.95} = \mathbf{K}_\infty = \{\sigma_1, \sigma_2, \dots, \sigma_7\}$. These subfigures can be viewed as a series of subcomplexes with $\mathbf{K}_0 \subseteq \mathbf{K}_{1.13} \subseteq \mathbf{K}_{1.95} = \mathbf{K}_\infty$. The final complex is the Delaunay triangulation of the original point set. As depicted, panel (c) does *not* represent a subcomplex in the filtration, as it does not contain the simplices $\{A, B\}$, $\{A, C\}$.

the best opportunity to relate the results of persistent homology to the properties of well understood systems.

B. Homology and Persistent Homology

Having introduced the notion of a filtered simplicial complex, the importance of homology and persistent homology can now be discussed. To facilitate this, two more examples will be used. Figures 2(a)-2(d) and 2(e)-2(h) show filtrations of alpha complexes for crystalline and amorphous atomic configurations, respectively, which will be referred to throughout this section.

Homology is generally concerned with quantifying the number of n -dimensional topological features in a simplicial complex. For an alpha complex, these can either be 0- or 1-dimensional. The 0-dimensional features correspond to the number of connected components, meaning the number of distinct groups of atoms. The 1-

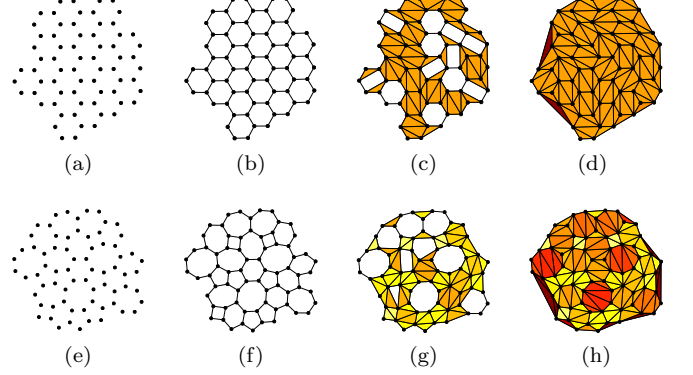


FIG. 2. Alpha complex at increasing filtration values (left to right) for crystalline (top row) and amorphous (bottom row) configurations. Initially there is a point cloud from the atomic coordinates in (a) \mathbf{K}_0 , $\beta_0 = 68$, $\beta_1 = 0$ and (e) \mathbf{K}_0 , $\beta_0 = 70$, $\beta_1 = 0$. Then lines appear in (b) $\mathbf{K}_{0.55}$, $\beta_0 = 1$, $\beta_1 = 24$ and (f) $\mathbf{K}_{0.55}$, $\beta_0 = 1$, $\beta_1 = 24$, which can be identified with nearest-neighbour bonds. Triangles from next nearest-neighbour distances follow in (c) $\mathbf{K}_{0.99}$, $\beta_0 = 1$, $\beta_1 = 12$ and (g) $\mathbf{K}_{0.99}$, $\beta_0 = 1$, $\beta_1 = 10$. Finally the Delaunay triangulation is formed in (d) \mathbf{K}_∞ , $\beta_0 = 1$, $\beta_1 = 0$ and (h) \mathbf{K}_∞ , $\beta_0 = 1$, $\beta_1 = 0$. Triangular simplices are coloured by filtration value, increasing from yellow to purple via orange and red. In the crystalline case all same dimensionality simplices form at very similar filtration values, whereas in the amorphous case the process occurs over a wider range.

190-dimensional features are the number of “cycles” (or
191 “holes”) in the structure. These are termed as such to
192 differentiate from “rings” used elsewhere in this work, but
193 often there will be significant overlap between the two.
194 Any alpha complex has the homology groups, $H_n(\mathbf{K}_\epsilon)$,
195 which contain all the associated n -dimensional features.
196 A zeroth homology group will contain 0-dimensional fea-
197 tures, a first homology group will contain 1-dimensional
198 features *etc.*. The rank of these groups are termed the
199 Betti numbers, β_n [18]. To make this less abstract, one
200 can see how this fits with the examples in figure 2:

- Figures 2(a) and 2(e) have $\mathcal{N} = 68$ and $\mathcal{N} = 70$ individual points respectively, and so have Betti numbers of $\beta_0 = \mathcal{N}$ and $\beta_1 = 0$.
- Figures 2(b) and 2(f) have all the atoms connected leading to the formation of 24 cycles in both cases, and hence $\beta_0 = 1$ and $\beta_1 = 24$.
- Figures 2(c) and 2(g) both have a reduced number of cycles owing to the filtration values of triangular simplices being met. The Betti numbers in figures 2(c) and 2(g) are $\beta_0 = 1$, $\beta_1 = 12$ and $\beta_0 = 1$, $\beta_1 = 10$, respectively. For the crystal in figure 2(c), the cycles do not all disappear for the same filtration value because of small variations of the atomic coordinates from their ideal values, which originates from the numerical precision.

- Figures 2(d) and 2(h) have just one large connected component and no cycles, such that $\beta_0 = 1$ and $\beta_1 = 0$.

These examples illustrate a more general principle, that the “starting point”, \mathbf{K}_0 , will always have $\beta_0 = \mathcal{N}$ and $\beta_1 = 0$, and the “end point”, \mathbf{K}_∞ , will always have $\beta_0 = 1$ and $\beta_1 = 0$, *i.e.*, trivial homology. It is the filtration values in between that lead to richer behaviour.

These considerations lead onto the notion of persistent homology. Between any two subcomplexes at different filtration values, a selection of topological features will be common to both, whilst others will appear or disappear when moving from one to the other. In other words, some features will *persist* over the range of filtration values, whilst others will not. To characterise this behaviour in the vocabulary of persistent homology, each feature is said to be “born” at a given filtration value, b , and “die” at a later value, d . The lifetime of the feature is then quantified via $l = d - b$. Finding and measuring the lifetimes of topological features is therefore the crux of persistent homology. In the first instance, it is the longest lived features that are normally of most interest, as these are considered to be representative of the true system topology. In the case of two-dimensional atomic materials, this ought to be reflective of the ring structure. However, other more fleeting (*i.e.*, shorter lifetime) features can also be useful, as these intermediate features act as signatures for some medium range ordering [7, 9].

C. Visualising Persistence

After running a persistent homology calculation by generating a filtered simplicial complex, finding the topological features and determining the lifetimes of each, the results have to be presented in a way that highlights the fundamental features and facilitates extraction of the key topological properties of the system. There are multiple ways to do this, some of which are more suitable for small systems and others for large aggregated datasets. The representations are outlined below, with examples given in figure 3, in reference to the small crystalline and amorphous configurations presented in figure 2.

- **Persistence barcode:** represents the lifetime of each topological feature as a bar, starting at the birth value and terminating at the death value (see figures 3(a), 3(b)). The barcode therefore contains all information about each feature, and whilst useful for small samples, it becomes difficult to interpret when there are a large number of features. In addition, features with small persistence values, that originate from short-term fluctuations in the atomic coordinates, are difficult to identify.
- **Evolution in Betti numbers:** plots the total number of each n -dimensional feature for each filtration value (see figures 3(c), 3(d)), and so provides more coarse-grained information than the

barcode. It is equivalent to counting the number of bars at a specific filtration value.

- **Persistence diagram:** plots the birth and death value pairs, (b, d) , of each feature as a scatter diagram (see figures 3(e), 3(f)). This gives a more holistic view of the data as a whole. For large data sets, histograms can be constructed, with points coloured by their relative multiplicities. The persistence diagram is therefore suitable for visualising large amounts of aggregate information.
- **Lifetime distribution function:** presents a histogram of the lifetimes of each feature. This is useful for analysing a large number of features, but loses information on the exact birth and death values. However, for atomic systems it will be seen that persistent cycles have similar birth values and so lifetime is strongly correlated with the death value.

For all these visualisation methods, it is worth emphasising that the filtration value, ϵ , has units of length, and for atomic networks is normally quoted in terms of the equilibrium bond length, r_0 .

Examining the results for the two example systems in figure 3, one can see that the crystalline and amorphous systems have both similarities and differences. To begin with, the 0-dimensional features appear very similar across all plots, with $b = 0$ and $d = r_0/2$. This finding originates from the fact that, in atomic systems, the average bond length is highly restricted, so all atoms become connected in a limited range of filtration values around half the mean bond length (as the filtration value measures the circumradius of the 1-simplex). As such, the 0-dimensional features provide little insight for atomic systems, and will be neglected in the analysis.

On the other hand, the 1-dimensional features show significant differences between the crystalline and amorphous configurations. Both systems have 24 persistent bars in their barcode (figures 3(a), 3(b)), which are born at $b = \epsilon = r_0/2$, but the lifetimes vary considerably. For the crystalline system, all persistent cycles (corresponding to hexagons) terminate at almost the same value of $\epsilon = r_0$ (there is a small variation in the bond length), whereas in the amorphous case there is a much broader distribution of values. These values will be discussed in detail in section IV A 2 and are related to the radius of the circumcircle into which each polygon is inscribed.

This effect is mirrored in the evolution of the first Betti numbers (figures 3(c), 3(d)). In both crystalline and amorphous configurations, β_1 first becomes 24 at $\epsilon = r_0/2$, as persistent 1-cycles initially form at half of the (well-defined) bond length. For the crystalline case all persistent cycles die at the same value, leading β_1 to behave as a step function at $\epsilon = r_0$. For the amorphous case, the variation in death values, resulting from the variation in persistent cycle sizes, leads to a slower decay. It is not only the persistent cycles that show variation be-

tween the crystalline and amorphous systems. The persistence diagrams, figures 3(e) and 3(f), show that the short lived features, which lie close to the $b = d$ diagonal, show greater variation in the amorphous case. This variation will be explored in greater depth in section IV.

The persistent homology analysis of the networks in figure 2 shows that the method has promise for capturing the disorder in amorphous materials. The question is whether these visualisations can be systematically interpreted to quantify the disorder, and obtain information that is not available via alternative methods.

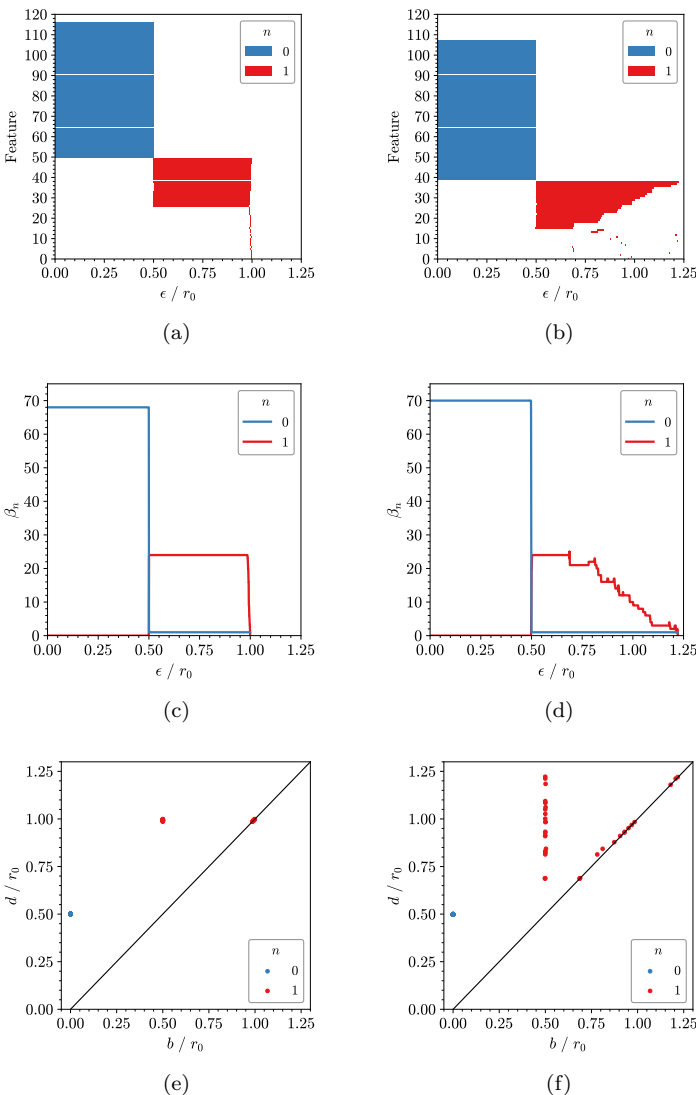


FIG. 3. The (a)–(b) barcode, (c)–(d) Betti numbers and (e)–(f) persistence diagram for crystalline (left column) and amorphous (right column) configurations.

IV. RESULTS

The use of persistent homology is investigated for two-dimensional materials generated computationally via two different methods. In the first, networks of triangle rafts are constructed in the manner described in [19]. In the second, continuous random networks are formed using the bond-switching algorithm described in [20]. Both methods invoke a “temperature” parameter, which allows for control over the degree of disorder in the generated configurations. However, in both cases this parameter does not have direct physical meaning, so any thermodynamic phase boundaries cannot be identified. In consequence, a liquid cannot be distinguished from a glass, so any network with a significant degree of disorder is classified as amorphous.

In two-dimensional materials of this type, the disorder is often quantified through the proportion of rings that are hexagons, denoted p_6 . From Euler’s formula for three-coordinate networks, the mean ring-size must be equal to six, $\langle k \rangle = 6$, as evident for the crystalline hexagonal lattice where $p_6 = 1$ [21]. As the network becomes more disordered with increasing temperature, the distribution of ring sizes broadens, and p_6 decreases. Furthermore, as the ring distribution is generally well-defined, following a lognormal or maximum entropy distribution [22, 23], p_6 alone is sufficient to quantify the overall disorder in the ring statistics. For example, p_6 is naturally coupled to the variance of the ring size distribution, $\mu_2 = \langle k^2 \rangle - \langle k \rangle^2$, termed a Lemaître law. Figure 4 shows the Lemaître solution obtained from maximum entropy, along with the data points that correspond to the network configurations considered in this paper.

The primary difference between the models created by the computational methods is, therefore, the degree of structural flexibility. Triangle rafts produce configurations with highly regular rings, whereas bond-switching allows for more irregular rings owing to a greater variability in the bond length and angles. As will be shown, this variability has consequences in the analysis of the metrics obtained from persistent homology. To complement the findings from the computational studies, a persistent homology analysis is also performed for experimental configurations of amorphous silica bilayers and graphene [24, 25].

For the triangle rafts the configurations are aperiodic. In consequence, there is the possibility of finite size effects in the results that are obtained. As illustrated in figure 2, some cycles arise from the close proximity of atoms at the periphery of the model (*i.e.*, edge simplices in the Delaunay triangulation). The number and nature of these cycles will depend on the shape of the boundary. These cycles are not expected, however, to materially affect the results because they diminish in relative number as the sample size increases and, more importantly, they do not contribute towards the initial set of persistent cycles formed near $\epsilon = r_0/2$. For the periodic configurations generated by the bond-switching algorithm, the

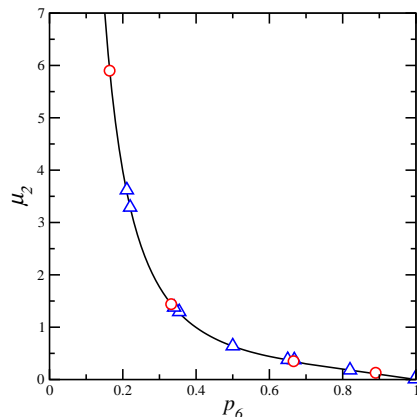


FIG. 4. The maximum entropy relationship between the proportion of six-membered rings p_6 and the variance of the ring-size distribution μ_2 , known as a Lemaître law. The blue triangles and red circles highlight the configurations generated from the triangle-raft versus bond-switching algorithms, respectively. For a given model, the value of p_6 decreases with increasing temperature.

systems are sufficiently large to allow for convergence of the key structural properties (see, for example, figure S2 in Ref. [26]).

A. Persistent Homology with Triangle Rafts

Triangle rafts are a model for two-dimensional silica bilayers that represent corner-sharing SiO_4 tetrahedral units projected onto a plane as equilateral triangles of SiO_3 (figure 5) [27]. They can be used to model amorphous phases with varying levels of disorder via the use of a Monte Carlo algorithm [19]. In this algorithm, a raft is built-up by the sequential addition of triangular subunits. The disorder in the system, as measured by the ring statistics, is controlled by a “temperature” parameter, T . At the lowest temperature the raft will be crystalline, consisting solely of hexagons. Increasing the temperature leads, however, to increasing variance of the ring-size distribution, enabling the formation of amorphous configurations. Triangle rafts can achieve diverse ring-size distributions, owing to the flexibility afforded to the structure by the oxygen linkages. In addition, the nearest-neighbour Si-Si distances remain in a relatively tight distribution (table I), as a result of the rigidity imposed by the triangular subunits.

The ability to tune the ring statistics in triangle rafts is beneficial as it allows for a systematic evaluation of the results of persistent homology calculations, using configurations with a continuous evolution in structure. In this work, rafts of $N = 1000$ rings, with ring sizes in the range $k = 4 \rightarrow 10$ and ring statistics with $p_6 = 1 \rightarrow 0.2$

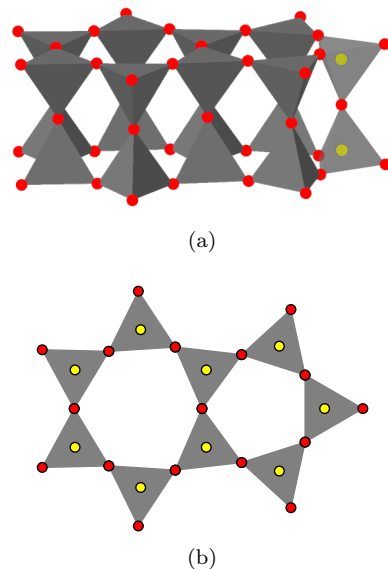


FIG. 5. Triangle rafts are a simplified two-dimensional representation of silica bilayers. A small sample of silica bilayer comprising corner-sharing tetrahedra is shown in (a) and the corresponding triangle raft in (b). Silicon and oxygen atoms are coloured yellow and red, respectively. Only the positions of the silicon atoms were used in the persistent homology calculations.

TABLE I. Relative Si-Si distances within different ring sizes in a triangle raft, assuming ring regularity. k is the number of vertices (Si atoms) in each primitive ring, *e.g.*, $k = 6$ for a hexagon, and the distances are given in units of the ideal nearest-neighbour Si-Si distance.

k	4	5	6	7	8	9	10
Si-Si	0.966	0.995	1.000	0.997	0.991	0.985	0.978

were generated and analysed using the GUDHI library [14]. The ring statistics give a measure of the structural disorder, with $p_6 = 1$ for a crystalline lattice. In these analyses, only the silicon positions were included, as the oxygen positions do not affect the network topology. The results of the persistent homology calculations will be discussed primarily in terms of the persistence diagrams, where the features are rationalised in comparison to those expected for regular versus distorted polygons. It will also be shown how the ring statistics compare to those inferred from persistent cycles.

1. Persistence Diagrams

The point clouds and Delaunay triangulation generated for single configurations of triangle rafts with four different levels of disorder are shown in figure S1. The corresponding persistence diagrams, combined across multiple configurations, are shown in figures 6(a)-(d). Discussing first the overall structure, one can see there is a systematic evolution in behaviour as the temperature

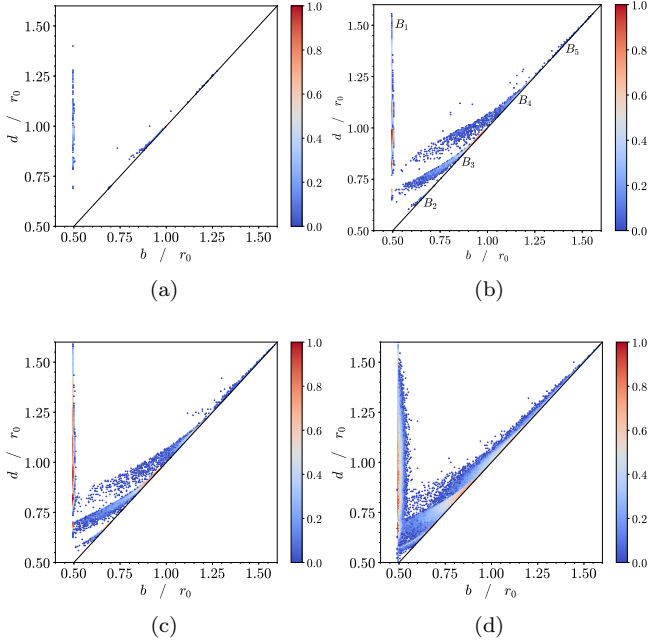


FIG. 6. Persistence diagrams for triangle rafts generated at increasing temperature and disorder: (a) $T = 10^{-4.5}$, $p_6 = 0.997$; (b) $T = 10^{-3.6}$, $p_6 = 0.668$; (c) $T = 10^{-2.7}$, $p_6 = 0.354$; (d) $T = 10^{-1.8}$, $p_6 = 0.211$. Points are coloured by multiplicity to highlight structure. The $b = d$ diagonal is shown by the black line. In (b), the bands B_n ($n = 1, 2, 3, 4$ or 5) contributing towards the persistence diagrams are identified.

(and therefore disorder) increases. To aid in the discussion, points in a persistence diagram will be referred to in terms of birth-death pairs, $(b/r_0, d/r_0)$, with filtration values given in reference to the equilibrium bond length, r_0 .

At the lowest temperature, configurations are dominated largely by regular hexagons with few defects. This scenario is reflected in the persistence diagram by bright spots at $(0.5, 1.0)$ and $(1.0, 1.0)$, which can be observed by zooming into figure 6(a). As the temperature increases, and rings of different sizes are introduced, characteristic bands form in the persistence diagram, which are most intense at short lifetimes (*i.e.*, close to the $b = d$ diagonal) and then “wash out” at smaller birth values and longer lifetimes (figures 6(b) and 6(c)). At the highest temperatures, these bands broaden and finally coalesce 6(d).

The band structure and its associated behaviour has been observed before in studies of three-dimensional amorphous silica [7, 8]. In that work, Hiraoka *et al.* [8] find that the bands correspond to the secondary rings formed within primary rings as the filtration value is increased. The interpretation is complicated, however, by the inclusion of both silicon and oxygen atoms in the persistent homology analysis. By studying the problem in two dimensions, it will be shown that bands arise from

the formation of cycles between nearest-neighbour, second and nearest-neighbour and successively higher order interactions.

2. Band Structure in Persistence Diagrams

Before rationalising the band structure observed in the persistence diagrams of triangle rafts, it is useful to consider again the process of cycle formation in persistent homology. The important points are summarised below:-

1. The initial set of persistent cycles are born at a filtration value corresponding to roughly half the atomic bond distance, $\epsilon = r_0/2$.
2. A secondary cycle is born when a 1-simplex cuts a previous cycle in two. The birth value of the new cycle is then equal to the filtration value of the 1-simplex.
3. A cycle dies when *all* of the 2-simplices within it are present. The death value is then equal to the highest filtration value of the constituent 2-simplices.
4. A cycle is counted only if the death value is strictly greater than the birth value.
5. As a result of the above, a cycle will only persist if the 1-simplex that creates it has a lower filtration value than the 2-simplex that destroys it.

Understanding these properties is key to interpreting the results of persistence diagrams.

To explore the origins of the band structure, an ideal set of polygons is considered, where all sides have unit length. This approximation is reasonable for triangle rafts due to the rigidity of the triangular subunits. In figure 6(a), the first band to consider is the vertical band at $b/r_0 = 0.5$, which displays bright spots along its length. These spots correspond to the cycles initially formed when 1-simplices, having a circumradius of half the edge length, connect adjacent atoms. This band therefore originates from cycles pertaining to the nearest-neighbour interactions, and so will be referred to as band B_1 . These cycles will die only when all the 2-simplices inside them are present. This point is illustrated in figure 7(a) for a regular octagonal cycle with unit side lengths. In this case, by definition, all simplices must have the same filtration value, as they lie on a common circle (figure 7(b)). The filtration values for these simplices are given by:

$$\Phi_k = \frac{1}{2 \sin(\pi/k)}, \quad (1)$$

where Φ_k is the circumradius in a regular k -sided polygon with unit edge lengths. It follows that a regular polygon will have a death value at Φ_k and no more cycles will be born out of this point set. Although difficult to see, these correspond well with the bright spots in the B_1 band in

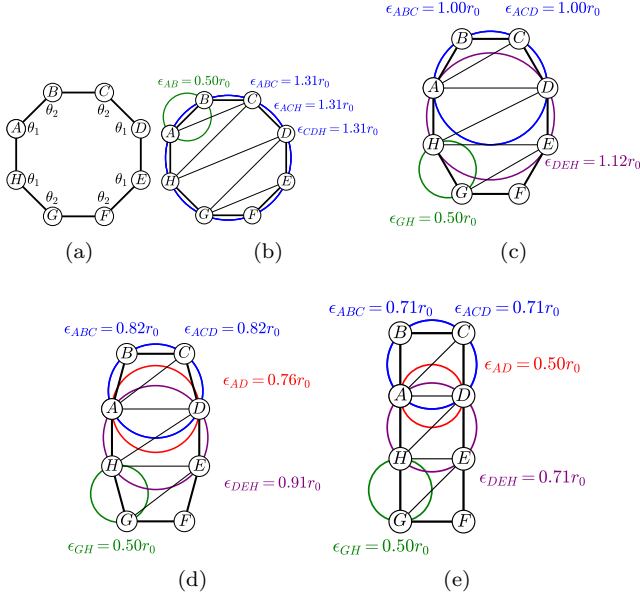


FIG. 7. Systematic distortions of an octagon with unit side lengths, defined in terms of the angles θ_1 and θ_2 shown in (a), where the polygon is compressed about an axis passing through two opposite edges and $\theta_2 = (6 \times 180 - 4\theta_1)/4$. For each geometry, the most persistent cycle is born when $b/r_0 = 0.50$, but the extent of distortion affects the lifetime and hence the coordinates (b, d) of the cycle on a persistence diagram (figure 6). Other (secondary) cycles can be born with increasing filtration. Panel (b) shows a regular octagon corresponding to $\theta_1 = 135^\circ$ with $(0.50, 1.31)$. Panels (c)-(e) show a series of distortions corresponding to (c) $\theta_1 = 150^\circ$ with $(0.50, 1.12)$; (d) $\theta_1 = 165^\circ$ with $(0.50, 0.91)$ and a second cycle with $(0.76, 0.82)$; and (e) $\theta_1 = 180^\circ$ with $(0.50, 0.71)$. The circumcircles of selected simplices are highlighted, with the corresponding filtration values colour coded. The Delaunay triangulation is given as the interior black lines. These diagrams outline the origin of the B_1 and B_3 bands.

the persistence diagrams in figure 6(a), as will become more apparent in the following discussion.

The vertical B_1 band shows, however, a continuum of values, not just discrete points. Therefore, a modification to the model can be considered, whereby all edge lengths are maintained at unity, but the polygon undergoes a systematic distortion, being compressed about an axis passing through two opposite edges. This process is highlighted in figures 7(b)-(e) for an octagon where the distortion is defined by the internal angles θ_1 and θ_2 . An octagon is selected merely for illustrative purposes because it can be used to generate all of the bands shown in figure 6(b). In considering the band B_1 , only the persistent cycles born at $b/r_0 = 0.5$ are of interest. These cycles will die when the Delaunay triangulation is realised, which is achieved when the simplex with the highest filtration value, here DEH (or equivalently ADH), is present. As the polygon is compressed, it can be seen that the circumcircle for this simplex (purple cir-

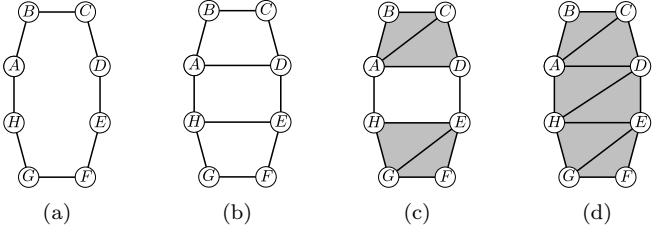


FIG. 8. Illustration of the explicit cycles present at increasing filtration value (left to right) for the octagon depicted in figure 7(d). Black lines indicate 1-simplices and grey triangles 2-simplices. Panel (a) has $\epsilon = 0.50r_0$ which marks the birth of the most persistent 1-cycle that contributes to the B_1 band. Panel (b) has $\epsilon = 0.76r_0$ and shows that the persistent cycle in (a) is cut, leading to the birth of two secondary cycles that contribute to the B_3 band. Panel (c) has $\epsilon = 0.82r_0$ which marks the death of the B_3 cycles in (b). Panel (d) has $\epsilon = 0.91r_0$ which marks the death of the persistent cycle.

cle) is reduced. In other words, distortion of the polygon acts to reduce the lifetime of the original cycle born at $b/r_0 = 0.5$. In consequence, the band B_1 manifests, instead of a simple series of discrete points.

The distortions in figure 7 demonstrate the phenomenon central to the discussion of the remaining bands in figure 6: that between some critical values of θ_1 , a second cycle is born within the first. This birth is only possible when the 1-simplex AD has a filtration value less than the 2-simplex ACD . In figure 7(c) where $\theta_1 = 150^\circ$, this criterion is *unfulfilled*, as AD and ACD share the same filtration value (*i.e.*, $\epsilon_{AD} = \epsilon_{ACD} = 1.00r_0$). This angle corresponds, in fact, to a limiting case, *i.e.*, secondary cycles will form when θ_1 is slightly greater than 150° with coordinates close to $(1.0, 1.0)$. In figure 7(d), where $\theta_1 = 165^\circ$, the circumcircle for AD (red circle) is not contained within that for ACD (blue circle), and has a smaller radius, and therefore filtration value. A second cycle is generated with $(0.76, 0.82)$. Figure 7(e) where $\theta_1 = 180^\circ$ represents the other limiting case, *i.e.*, secondary cycles will form when θ_1 is slightly less than 180° with coordinates close to $(0.5, 0.71)$. The coordinates of the secondary cycles in figure 7 will therefore range between $(0.5, 0.71)$, $(0.76, 0.82)$ and $(1.0, 1.0)$. The intermediate values can be found by numerically scanning the angle range $150^\circ < \theta_1 < 180^\circ$ but, even from these points, it is clear that the (b, d) coordinates correspond to the most prominent “central” band in the persistence diagram. Furthermore, this band arises from the cycles formed by connecting neighbours three bonds apart. To demonstrate this important point, the birth and death of the cycles in the octagon in figure 7(d) are shown in figure 8. Here it can be seen that the secondary cycles that are born in figure 8(b) and die in Figure 8(c) are the result of a cross-ring interaction between atoms that are three bonds apart. Although shown for an octagon, this finding holds for other ring sizes with sufficient nodes (*i.e.*, $k > 5$), so this band is termed B_3 .

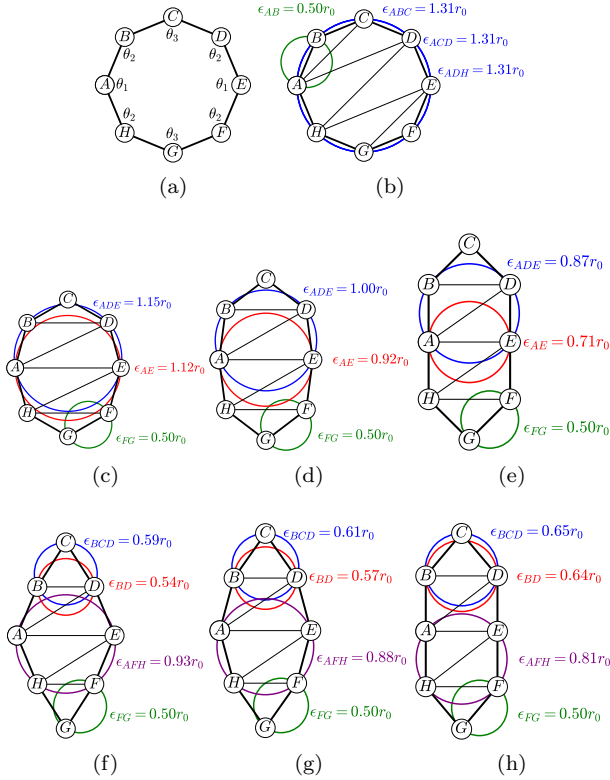


FIG. 9. Systematic distortions of an octagon with unit side lengths, defined in terms of the angles θ_1 , θ_2 and θ_3 shown in (a), where the polygon is compressed about an axis passing through two opposite vertices and $\theta_3 = (6 \times 180 - 2\theta_1 - 4\theta_2)/2$. For each geometry, the most persistent cycle is born when $b/r_0 = 0.50$, but the extent of distortion affects the lifetime and hence the coordinates (b, d) of the cycle on a persistence diagram (figure 6). Other (secondary) cycles can be born with increasing filtration. Panel (b) shows a regular octagon corresponding to $\theta_1 = \theta_2 = 135^\circ$ with (0.50, 1.31). Panels (c)-(e) show a series of distortions corresponding to (c) $\theta_1 = 150^\circ$, $\theta_2 = 135^\circ$ with (0.50, 1.15) and (1.12, 1.15); (d) $\theta_1 = 165^\circ$, $\theta_2 = 135^\circ$ with (0.50, 1.00) and (0.92, 1.00); (e) $\theta_1 = 180^\circ$, $\theta_2 = 135^\circ$ with (0.50, 0.87) and (0.71, 0.87); (f) $\theta_1 = 135^\circ$, $\theta_2 = 170^\circ$ with (0.50, 0.93) and (0.54, 0.59); (g) $\theta_1 = 150^\circ$, $\theta_2 = 160^\circ$ with (0.50, 0.88) and (0.57, 0.61); and (h) $\theta_1 = 180^\circ$, $\theta_2 = 140^\circ$ with (0.50, 0.81) and (0.64, 0.65). For panels (f)-(h) an additional cycle is present, contributing to B_4 , but is omitted for simplicity of presentation. The circumcircles of selected simplices are highlighted, with the corresponding filtration values colour coded. The Delaunay triangulation is given as the interior black lines. These diagrams outline the origin of the B_2 and B_4 bands.

It might now be clear that the other bands in figure 6 can also be related to specific-neighbour interactions. In order to demonstrate this point the octagonal case is examined again, but this time the ring distortion is caused by compression about an axis through two opposite vertices, as shown in figure 9. Figures 9(c)-(e) illustrate how secondary cycles can be born at relatively high filtration values. Compression causes the 1-simplex AE to have a

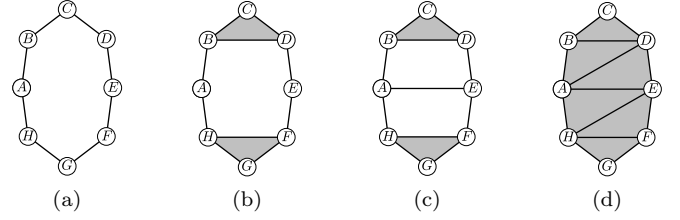


FIG. 10. Illustration of the explicit cycles present at increasing filtration value (left to right) for the octagon depicted in figure 9(d). Black lines indicate 1-simplices and grey triangles 2-simplices. Panel (a) has $\epsilon = 0.50r_0$ which marks the birth of the most persistent 1-cycle that contributes to the B_1 band. Panel (b) has $\epsilon = 0.82r_0$ and shows the addition of two 2-simplices that do not lead to new cycles. Panel (c) has $\epsilon = 0.92r_0$ which marks the birth of a secondary cycle that contributes to the B_4 band. Panel (d) has $\epsilon = 1.00r_0$ which marks the death of both the persistent and secondary cycles.

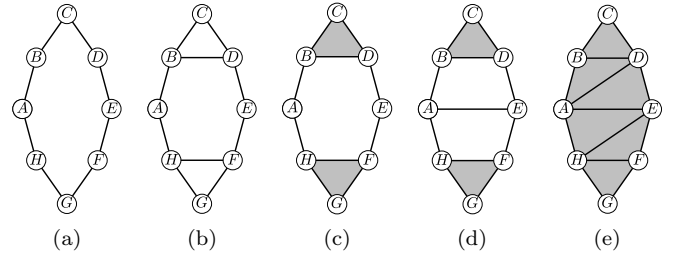


FIG. 11. Illustration of the explicit cycles present at increasing filtration value (left to right) for the octagon depicted in figure 9(g). Black lines indicate 1-simplices and grey triangles 2-simplices. Panel (a) has $\epsilon = 0.50r_0$ which marks the birth of the most persistent 1-cycle that contributes to the B_1 band. Panel (b) has $\epsilon = 0.57r_0$ and shows that the persistent cycle in (a) is cut, leading to the birth of two secondary cycles that contribute to the B_2 band. Panel (c) has $\epsilon = 0.61r_0$ which marks the death of the B_2 cycles in (b). Panel (d) has $\epsilon = 0.83r_0$ which marks the birth of a cycle contributing to the B_4 band. Panel (e) has $\epsilon = 0.88r_0$ which marks the death of both the persistent and secondary cycles.

filtration value smaller than the 2-simplex ADE, enabling the formation of a second cycle. The atoms A and E are separated by four bonds, so the corresponding band for such species is labelled B_4 . The as-yet unaccounted for B_2 band arises from atoms separated by two bonds. For an octagon, examples of the cycles contributing to the B_2 band are given in figures 9(f)-9(h) and correspond to “pinched” configurations. A demonstration of the specific cycles contributing to the B_2 and B_4 bands is given in figures 10 and 11 for the octagons in figures 9(d) and 9(g), respectively.

For the octagon it is relatively difficult to add to the B_2 band, a point that is worth some consideration. Here, it is important to note that the analysis is not unique to the octagon, which has been selected merely for illustrative purposes as it *can* display all the required behaviour.

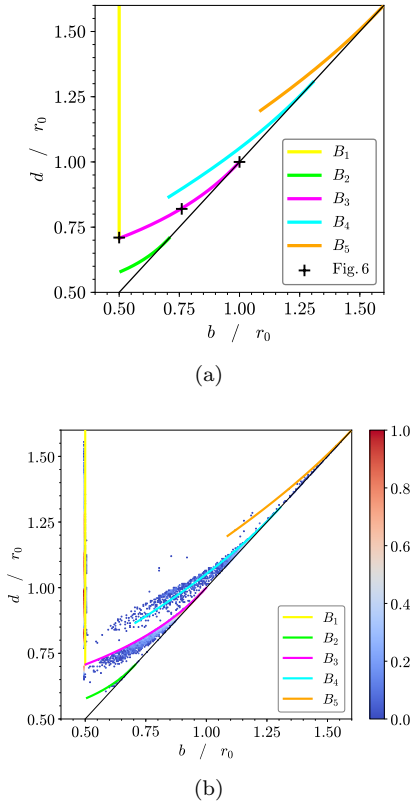


FIG. 12. Bands corresponding to different specific-neighbour interactions, as calculated numerically using idealised models. Panel (a) shows the curves calculated by distorting a range of k -sided polygons, whilst panel (b) overlays these curves on the persistence diagram for triangle rafts from figure 6(c), to which there is good agreement. In (a), the crosses on band B_3 at $(0.5, 0.71)$, $(0.76, 0.82)$ and $(1.0, 1.0)$ correspond to the distorted octagons of figure 7(c)-(e) where the θ_1 values are 150° , 165° and 180° , respectively. The secondary cycles contributing to this band span the range $150^\circ < \theta_1 < 180^\circ$.

In general, large polygons ($k > 6$) require little distortion to yield B_3 and B_4 bands, whereas significant rearrangement is needed to contribute to B_2 . On the other hand, the B_2 band readily forms for small polygons $k < 6$ and, indeed, the higher bands are inaccessible. The most common polygon, the hexagon, most easily forms secondary cycles in B_3 , and to a lesser extent B_2 (see movie *ph.hex.mp4* in the supplementary material). This explains the relative prominence of the central B_3 band in figures 6(a)-(d). Furthermore, it is equally possible with sufficiently large polygons to obtain higher bands still. In the case of the triangle rafts, the decagon leads to the presence of a weak B_5 band.

To emphasize further the above analysis, bands can be generated numerically by systematically distorting a range of realisable polygons typically observed in systems of this type, *i.e.*, for k in the range $3 \rightarrow 10$ (see movie *ph.all.mp4* in the supplementary material). Figure 12 shows the bands B_1 to B_5 (inclusive) overlaid with the

data from figure 6(c). The shapes and positions of the various bands match well with those found in the persistence diagrams from triangle rafts. Furthermore, the intensities of the bands in the persistence diagrams of triangle rafts agree with the previous discussion, with B_3 being by far the most prominent. It should be noted that the B_3 band calculated from the idealised model corresponds to the upper limit of that found from the triangle raft simulations. This finding is expected, as distortions in bond lengths will serve to reduce the death value. As shown in table I, the nature of the triangle raft model (of hinged near-rigid triangles) means that bond lengths can only be less than the equilibrium value, which leads to a concomitant decrease in the circumradii. Overall, the relative simplicity and rigidity of the triangle raft model has enabled the structure of the corresponding persistence diagrams to be well understood.

3. Cycles, Betti Numbers and Ring Statistics

In addition to the persistence diagrams, information is also available through the first Betti number, β_1 , which quantifies the number of cycles present at a specific filtration value. Plotting the β_1 against ϵ therefore reveals how the number of cycles grows and decays across the range of filtration values [7]. As will be shown, this allows characterisation of the proportion of cycles with k vertices, here termed k -cycles.

Figure 13 shows the evolution of the first Betti number with filtration for the triangle rafts constructed at different temperatures. Each configuration contains 1000 rings, but the data are averaged over all 100 configurations produced at the same temperature. Figure 13(a) shows β_1 for all cycles whilst figure 13(b) shows β_1 only for the 1000 most persistent cycles, *i.e.*, those with the longest lifetimes. Here, the rationale is that the second case will only include cycles in the B_1 band, excluding those in the higher order bands, and should therefore correspond most closely to the known ring structures. In fact, both sets of results are very similar (as will be discussed in due course), and so can be considered together.

As is now expected, β_1 rises very sharply from $0 \rightarrow 1000$ at $\epsilon/r_0 = 0.5$, as the first cycles form at half the mean bond distance. The value of β_1 must then decay to zero in the limit of $\epsilon \rightarrow \infty$, but the form of the curve varies with the system temperature. At the lowest temperature, where the network is primarily a hexagonal lattice, the shape is almost a step function at $\epsilon/r_0 = 1.0$. As the temperature increases, however, more diverse ring sizes are incorporated and the curve broadens and smoothens. The decay in β_1 is not, however, smooth, even for the highest temperature. Instead, there is a series of stepwise decrements. These decrements align very closely to the values of the circumradius of k -polygons with unit edge lengths, Φ_k , which represents the upper bound on the death value for a k -polygon in B_1 for triangle rafts. Any deviations from regularity act to reduce

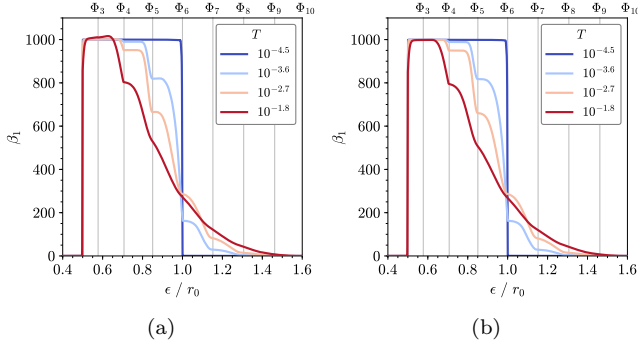


FIG. 13. Evolution of the first Betti number with filtration for triangle rafts at different levels of disorder. All of the calculated cycles are shown in panel (a), whereas only the 1000 most persistent cycles are shown in panel (b). The circumradii of regular polygons with k vertices, Φ_k , are indicated by vertical grey lines.

the lifetime.

This behaviour largely explains the properties of the curves in figure 13. The discontinuities represent the changes that occur as successive k -cycles die, curve broadening is indicative of an increased variety of k values, and curve smoothening indicates increased polygon distortion. The broadening and smoothening both increase with temperature. It is interesting to note that the secondary cycles do not seem to affect the first Betti numbers to a large degree, as evidenced by the similarity of figures 13(a) and 13(b). Only at the highest temperature is there a detectable presence in the region between $\epsilon/r_0 = 0.5 \rightarrow 0.7$. This lack of sensitivity originates from the fact that the secondary cycles are so relatively short-lived and spread throughout the filtration range that, at any filtration value, they add a virtually insignificant contribution to the total. The above analysis allows the persistent cycle statistics to be extracted as

$$p_k = \beta_1(\Phi_k) - \beta_1(\Phi_{k-1}), \quad (2)$$

normalised across all k values. The cycle statistics from this process are given in figure 14(a), which can be compared to the directly calculated ring statistics in figure 14(b).

Alternatively, the cycles statistics can also be extracted from the lifetime distribution function, $f(l)$. Equation (1) can be used to predict the lifetime of a given k -gon as:

$$l_k = \Phi_k - \frac{1}{2}, \quad (3)$$

assuming that the persistent features are born at half the inter-atomic distance. Therefore, by integrating under the probability density function between the limits l_{k-1} and l_k , the resulting value should provide an estimate of the ring statistics, p_k , *i.e.*:

$$p_k = \int_{l_{k-1}}^{l_k} f(l) dl. \quad (4)$$

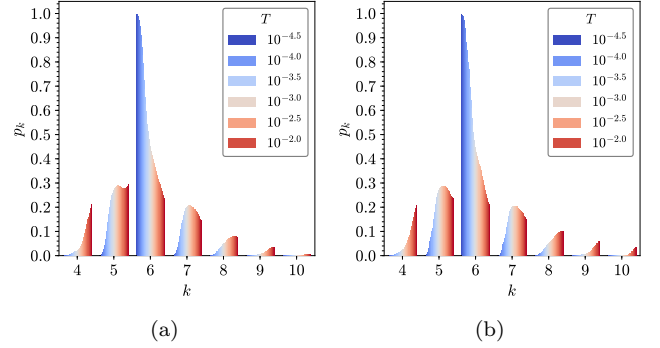


FIG. 14. Comparison of (a) the persistent k -cycle statistics obtained from equation (2) with (b) the primitive ring statistics for triangle rafts across different levels of disorder.

Figure 15(a) shows the full distribution of persistence lifetimes, $f(l)$, at fixed temperature. It also shows a truncated distribution in which the contribution from short-lived cycles with lifetimes below a certain threshold ($\sim 0.1r_0$) are eliminated, thus removing a sharp peak near zero, and the distribution is re-normalised. Here, the objective is to identify the most persistent cycles for comparison with the ring statistics. Figure 15(b) shows the truncated distributions across a range of temperatures, with blue \rightarrow red indicating cold \rightarrow hot. As the temperature increases, the distribution of lifetimes changes with the appearance of new peaks and the disappearance of others. By identifying the peaks in the truncated $f(l)$ distributions with specific k -gons, the cycle statistics can be estimated using equation (4) and compared to the ring statistics found from direct calculation. These are found to be almost identical to those calculated from the first Betti numbers, *i.e.*, those given in figure 14(a).

A direct comparison of the cycle statistics against the primitive ring statistics is provided in figure 16 for each method. By reference to figure 14, at the smallest temperature and lowest level of disorder, $p_6 \sim 1$ and $p_k \sim 0$ for $k \neq 6$, *i.e.*, the distribution is strongly centred on p_6 . This situation corresponds to the grey (p_6) line at the top right in figure 16(a) or 16(b) and to all the other lines at the extreme left. As the temperature and disorder increase, the distribution in figure 14(a) or 14(b) remains centred on p_6 but the value of p_6 decreases as the values for p_k ($k \neq 6$) increase. This scenario corresponds to moving down the grey line to the left in figure 16(a) or 16(b) and up the other lines to the right.

The comparison of figure 16 shows that, to a first approximation, the agreement between distributions is very good, particularly for ring sizes close to $k = 6$ at modest temperatures. This finding means that the persistent cycles computed via persistent homology coincide well with the primitive rings in the system, taking into account the approximative method of calculating k -cycles. Deviations are accentuated at very low temperature, when the

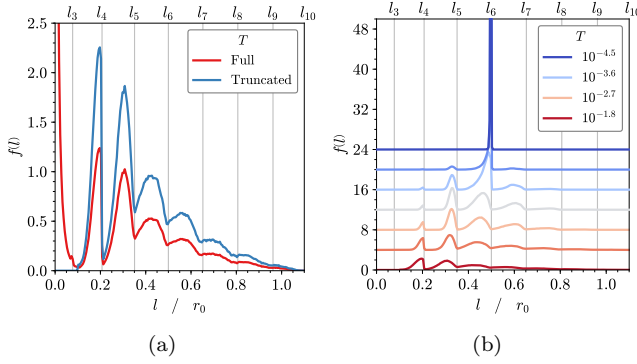


FIG. 15. Lifetime distributions for triangle raft models. (a) The distribution for a single temperature before and after a threshold is applied to remove short-lived features. (b) The distributions with the short-lived features removed show structure that evolves systematically with increasing temperature (blue \rightarrow red). The curves are offset vertically for clarity of presentation. The vertical grey lines show the lifetimes estimated from equation (3).

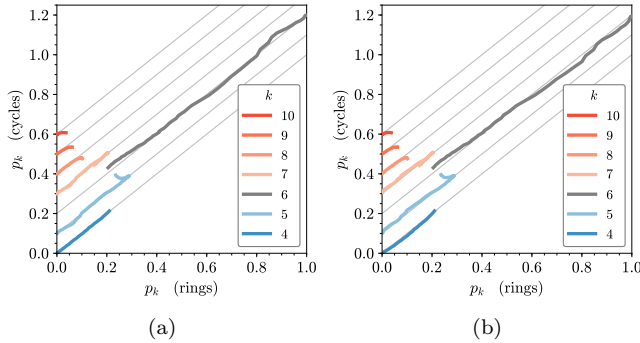


FIG. 16. Comparison of the primitive ring statistics with (a) the cycle statistics calculated from the first Betti numbers and (b) the lifetime distributions. The curves are offset vertically for clarity of presentation.

absolute ring statistics of non-six rings are small, or at high temperatures, when rings are increasingly distorted. For example, it is clear that p_{10} is systematically underestimated from cycles across the whole temperature range. In addition, with increasing temperature, the statistics of smaller rings ($k \leq 6$) are increasingly overestimated and larger rings ($k \geq 8$) increasingly underestimated. The reason for this is that when large cycles become sufficiently distorted, *e.g.*, through large elongation or even the development of a slightly non-convex shape, the cross-cycle distances become sufficiently small that they can die earlier than more regular cycles. Notwithstanding, for triangle rafts, it is reasonable to conclude that the persistent cycles found from persistent homology are in accordance with the primitive rings in the system.

B. Persistent Homology for Networks Generated by Bond Switching

The persistent homology of triangle rafts leads to a readily identifiable band structure in the persistence diagrams, as a consequence of the tightly controlled ring geometries. Such structure is not always observed in simulations of three-dimensional systems such as Cu-Zr alloys and high density molecular liquids [8, 9]. In these cases, more diffuse diagrams are found, which accompanies an increased number of degrees of freedom. The triangle raft algorithm is unable to directly reproduce these effects in two dimensions, owing to the innate rigidity of the model. This restriction can be overcome, however, by employing a bond-switching algorithm.

Bond switching Monte Carlo algorithms have a rich history in materials science, being introduced by Wooten, Winer and Weaire to model three-dimensional amorphous silicon [28]. Since then, the algorithm has been improved and adapted to two-dimensional materials [20, 29, 30]. The bond-switching algorithm proceeds by amorphising a pristine hexagonal net through a series of bond transpositions between neighbouring atoms. After thermalising the original lattice to obtain a fully amorphous structure, the degree of disorder can be selected through a “temperature” parameter as the system is cooled. In bond switching, the choice of potential model can allow for greater freedom in the bond lengths and angles as compared to triangle rafts, and the process of randomising an existing lattice (as opposed to building a random ring structure sequentially) allows for configurations with a large range of ring sizes and a large level of ring disorder.

In this work a simple but physically motivated potential model was selected to achieve the desired flexibility. It consists of a simplified Keating potential with harmonic bond stretching and angular terms [31], augmented by a restricted bending potential to maintain ring convexity [32]. The potential has the form:

$$U = \frac{K_r}{2} \sum_{\substack{i,j \in \\ \text{stretches}}} (r_{ij} - r_0)^2 + \frac{K_\theta}{2} \sum_{\substack{i,j,k \in \\ \text{angles}}} \frac{(\cos \theta_{ijk} - \cos \theta_0)^2}{\sin^2 \theta_{ijk}}, \quad (5)$$

where r_{ij} is the distance between atoms i and j ; θ_{ijk} is the angle defined by atoms i, j and k ; r_0 and θ_0 are the bond stretching and bond angle equilibrium values; K_r and K_θ are the bond stretching and bond angle force constants.

Persistent homology analysis was performed on the configurations generated from bond switching, applying the information learned from analogous calculations on triangle rafts. The potential model used had a bond length/angle force constant ratio of $K_r/K_\theta = 16$. Configurations of $N = 1024$ rings were thermalised with 2×10^5 random moves, and annealed over a further 4×10^6 moves. For each parameter set, 100 simulations were run starting from different random seeds. Again, calculations were

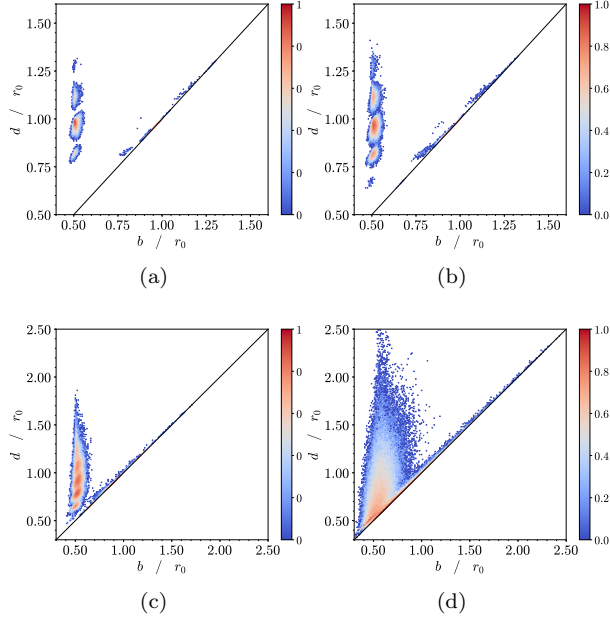


FIG. 17. Persistence diagrams for network configurations generated by bond switching with (a) $p_6 = 0.890$; (b) $p_6 = 0.666$; (c) $p_6 = 0.332$; (d) $p_6 = 0.164$. The level of disorder increases as the value of p_6 decreases.

performed with the GUDHI library [14], and birth and death filtration values are given in reference to the equilibrium bond length, r_0 . The point clouds and Delaunay triangulation generated for single configurations of bond-switch models with four different levels of disorder are shown in figure S2.

1. Persistence Diagrams

Figure 17(a)-(d) shows the persistence diagrams for four different p_6 values, representing different levels of disorder, averaged over 100 configurations. On first impression it seems that the persistence diagrams look very different to those of triangle rafts in figure 6. Closer inspection reveals, however, that this is not the case. Similar features are present and retain similar structural origins but are broadened. A key point in this broadening is that the bond lengths and angles have inherently more flexibility in the case of bond-switch networks than triangle rafts.

For example, the B_1 band is present in all the diagrams of figure 17, centred around $b/r_0 = 0.5$. The difference is that at low levels of disorder (e.g., figures 17(a)-(b)), there are distinct “islands” in B_1 , which are present in figure 6 but harder to detect due to the narrowness of the band. The upper d/r_0 limits of these spots correspond well with the expected circumradii of equation (1), i.e., they are related to the occurrence of specific k -cycles. At higher levels of disorder these spots begin to coalesce, as in figure 17(c), before combining fully to form a single

high-density feature, as in figure 17(d).

The final diagram might be considered “typical” of a highly disordered amorphous or liquid state [8, 9]. By using the bond-switching method it is clear, however, that this originates from a broadening and aggregation of features relating to individual cycle sizes. Other bands in the persistence diagram are also detectable, in particular B_3 , but only at larger p_6 values when the large cycles are less distorted. In bond switching, although there are primitive rings that are very large, these can become so distorted that they are not calculated as single persistent cycles, but rather multiple smaller cycles. This distortion precludes the formation of significant higher-order bands as seen in triangle rafts.

In the persistence diagrams, the B_1 band becomes progressively tilted from the vertical with increasing levels of disorder, driven by the appearance of a growing number of smaller cycles with $b/r_0 < 0.5$ and a growing number of larger cycles with $b/r_0 > 0.5$. This evolution is shown in the movie `ph_bondswitch.mp4` in the supplementary material. The birth filtration value of a cycle depends on the final 1-simplex to form. As the disorder is increased, it is evident that the probability of all constituent 1-simplices having filtrations less than $b/r_0 = 0.5$ becomes increasingly unlikely for larger cycles as compared to smaller cycles.

We note that these findings for the bond-switch model are not dependent on the precise choice of parameters in equation (5). Simulations using different bond stretching/force constants did not have a marked effect on the persistence diagrams.

Hiraoka *et al.* [8] state that the presence (or not) of “islands” in the persistence diagrams generated for their three-dimensional networks is direct evidence for crystalline (or amorphous) states. In the present work, however, islands may be present in the persistence diagrams for amorphous configurations as seen, for example, in figures 17(a) and 17(b), which reflects the more ordered local configurations (rings) present in these two-dimensional systems. At high enough temperatures, the islands merge as the rings become more distorted, as shown in figures 17(c) and 17(d).

2. Cycles, Betti Numbers and Ring Statistics

The first Betti number, β_1 , can again give insight into the cycle structure in configurations generated from bond switching. Figure 18(a) shows the evolution in β_1 with ϵ . The curves for the two largest p_6 values exhibit similar behaviour to those for triangle rafts. Around $\epsilon/r_0 = 0.5$, 1024 cycles form, corresponding to the primitive rings in the system, before decaying in a stepwise fashion at filtration values corresponding to the circumradius of a regular k -polygon. However, at the smaller p_6 values, the curves become smooth and the maximum decreases to around half its original value. As there remain the same number of primitive rings, this decrease implies that

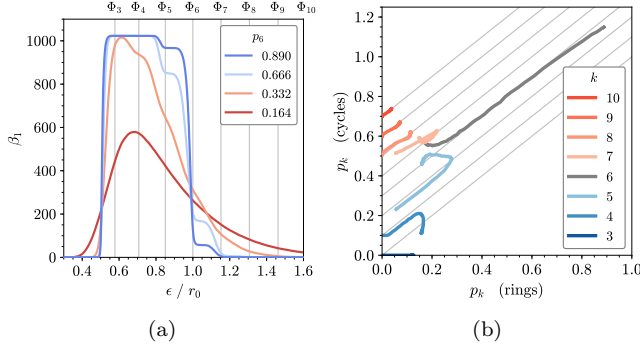


FIG. 18. (a) Evolution of the first Betti number with filtration for bond switching configurations at different levels of disorder represented by p_6 . The circumradii of regular polygons with k vertices, Φ_k , are indicated by vertical grey lines. (b) Comparison of the primitive ring statistics and cycle statistics calculated from the first Betti numbers. The curves are offset vertically for clarity of presentation.

some cycles die before others are even born so they never coexist. This in turn means that all of the rings are not detectable by persistent homology on the same length scale.

In figure 18(a), the final broad distribution resembles closely that of previous studies of random structures [7]. This type of distribution is therefore characteristic of a highly disordered material. As mentioned for the corresponding persistence diagrams, the broadening results from the large variation in bond lengths and bond angles.

Figure 18(b) compares the cycle and primitive ring statistics. As found for the triangle rafts (figure 16(a)), there is generally a strong correlation between the calculated cycles and rings. Discrepancies emerge, however, at higher temperatures where the rings become significantly distorted. These differences are larger than for the triangle raft models because the latter are more constrained, particularly in respect to the nearest-neighbour Si-Si length scale (Table I).

C. Persistent Homology using Experimental Configurations

The atomic coordinates in two-dimensional or quasi two-dimensional structures can be identified by experimental techniques such as electron microscopy. Here, amorphous graphene is an archetypal example of such a system in which ordering can be manipulated with an electron beam [24, 33]. A single pristine graphene sheet consisting of purely six-membered rings can be systematically disordered by the introduction of (mostly) 5- and 7-membered rings, whilst the carbon atoms remain strictly threefold coordinated. Such materials are relatively simple to construct computationally (see, for example, Ref. [21]) for comparison with experiment. In addition, amorphous bilayers of silica [25, 34–36], germania

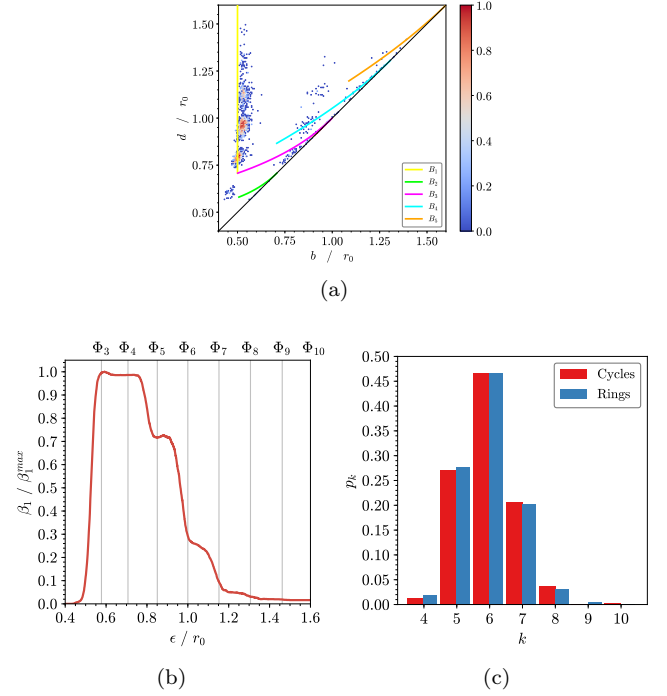


FIG. 19. The (a) persistence diagram, (b) first Betti numbers and (c) comparison of the primitive ring statistics and cycle statistics calculated from the first Betti numbers (as indicated by the legend) for a single bilayer configuration of silica extracted from an experimental image [25]. In (a), the curves correspond to the bands calculated by distorting a range of k -sided polygons (figure 12(a)). In (b), the circumradii of regular polygons with k vertices, Φ_k , are indicated by vertical grey lines.

[37, 38] and aluminosilicates [39] have been synthesised, a key motivation being control of the pore size [36]. In all cases, tetrahedral MO_4 ($\text{M} = \text{Si}, \text{Ge}$ or Al) motifs link to form a vertex-sharing bilayer network, and the projection of these tetrahedral units onto a plane leads to a network of corner-sharing triangles, the triangle nets considered above (figure 5). The configurations obtained from experiment provide detailed information on the atomic coordination environments. They are limited, however, to relatively small system sizes and the effect (if any) of the underlying surfaces on which they are deposited (in, for example, promoting defects or large-scale holes) remains unclear.

In the present work persistent homology calculations were performed for two different systems. In the first, a single configuration for an amorphous silica bilayer containing 3958 Si atoms is taken from Ref. [25]. In the second, four amorphous graphene configurations containing between 8792 and 13228 atoms are taken from Ref. [24]. These configurations show different values of p_6 and can therefore be compared to the configurations obtained by simulation at different temperatures and/or quench rates [21].

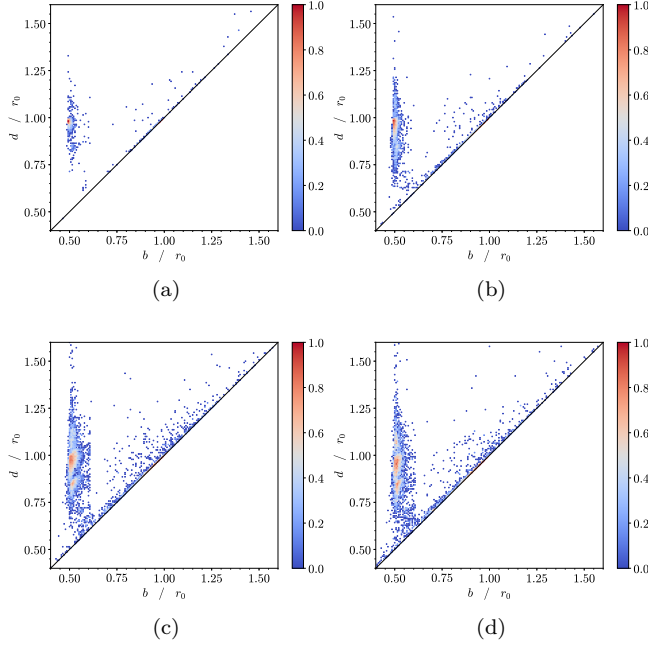


FIG. 20. Persistence diagrams generated for amorphous graphene configurations obtained from experimental images [24]. Each panel corresponds to a different value of p_6 , with (a) $p_6 = 0.97$; (b) $p_6 = 0.82$; (c) $p_6 = 0.61$; or (d) $p_6 = 0.50$.

1. Amorphous Silica Bilayers

Figure 19(a) show the persistence diagram for the amorphous silica bilayer configuration. Although the sample size was small and comprised a single configuration, there is evidence of the B_1 – B_5 bands observed for the triangle rafts (figure 12(b)). The bands in the persistence diagram from experiment are more disperse than for the triangle rafts, which is indicative of greater disorder in the bilayer material. This disorder originates from tilting of the tetrahedral SiO_4 units within the bilayer structure (figure 5). The projection of this tilted structure onto a single plane leads, therefore, to a distribution of nearest-neighbour Si-Si distances that is larger than found in the triangle raft models. Figure 19(b) shows the first Betti numbers, which display clear similarities with those generated for the triangle rafts (figure 13(a), at $T \sim 10^{-2.7}$). Large voids in the experimental configurations manifested themselves as readily identifiable long lifetime cycles in a persistence barcode or diagram.

2. Amorphous Graphene

Figures 20 and 21 show the persistence diagrams and Betti numbers for the four amorphous graphene configurations corresponding to fractions of six-membered rings in the range $p_6 = 0.50 - 0.97$. The relatively small system sizes and absence of averaging over multiple configurations

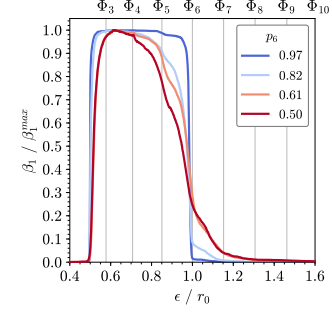


FIG. 21. First Betti numbers obtained for amorphous graphene configurations obtained from experimental images [24]. The coloured curves correspond to different values of p_6 as indicated by the legend. The circumradii of regular polygons with k vertices, Φ_k , are indicated by vertical grey lines.

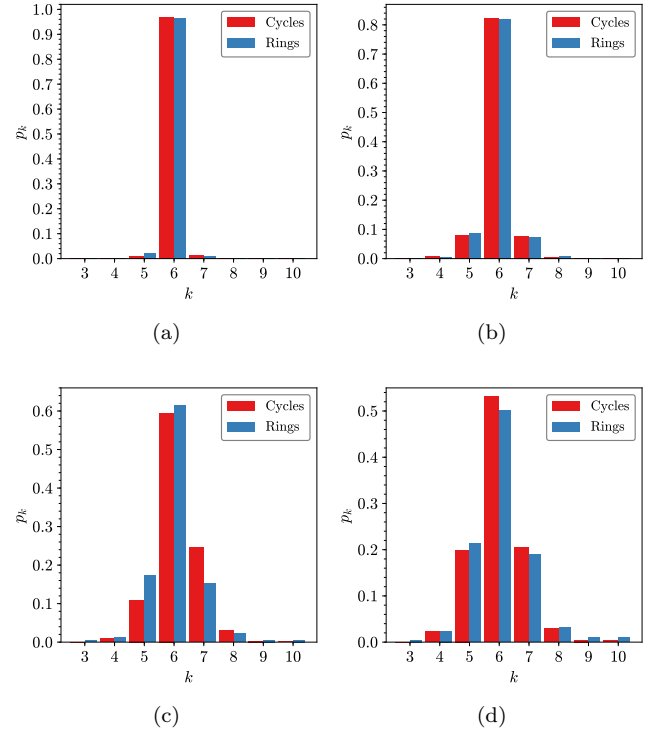


FIG. 22. Comparison of the primitive ring statistics and cycle statistics calculated from the first Betti numbers for configurations of amorphous graphene extracted from experimental images [24]. Each panel corresponds to a different value of p_6 , with (a) $p_6 = 0.97$; (b) $p_6 = 0.82$; (c) $p_6 = 0.61$; or (d) $p_6 = 0.50$.

rations makes it difficult to draw firm conclusions. However, the persistence diagrams shown in figure 20 are less structured than the persistence diagram for the amorphous silica bilayer (figure 19(a)), and are more consistent with those generated from the bond-switch models shown in figure 17. Similar remarks apply to the first Betti numbers shown in figure 21, which are consistent with those shown in figure 18(a).

3. Cycles and Rings

Figures 19(c) and 22 compare the primitive ring and cycle statistics for the experimentally-derived amorphous silica bilayers and graphene configurations, respectively. The cycle and primitive ring statistics show a strong correlation across all five configurations. Interestingly, the more disordered amorphous graphene configurations show a broad range of ring sizes (figure 22(c) and (d)) with $4 \leq k \leq 10$, where the distribution in figure 22(d) is similar to the distribution found for the amorphous silica bilayer (figure 19(c)). There are differences, however, in the underlying ring distortions in the amorphous silica bilayer versus graphene systems as found by comparing the persistence diagrams (figure 19(a) and figure 20(d)) and first Betti numbers (figure 19(b) and figure 21). The amorphous graphene configuration is more disordered, *i.e.*, the network is constructed from less regular rings. The persistent homology method is therefore sensitive to network distortion, as found for networks of densified silica glass [40].

V. DISCUSSION: EVALUATION OF PERSISTENT HOMOLOGY

Having investigated the results of persistent homology calculations for a range of two-dimensional systems, an attempt will be made to draw conclusions on the utility of persistent homology in the characterisation of these materials and more complex atomic systems.

Amorphous bilayer structures of materials such as silica, germania and aluminosilicates have relatively rigid network geometries that can be represented by a triangle-raft model. The associated persistence diagrams show band structure in which a specific band B_n originates from atoms separated by n bonds, *i.e.*, from specific-neighbour interactions. The most persistent k -cycles in the B_1 band can be identified with primitive rings, and the upper death values found for these cycles align closely to the circumradii or regular k -polygons. A persistence diagram therefore proves effective in revealing the n -body correlations associated with k -polygons, and the way in which this structure evolves with increasing disorder. It therefore contains more information than a structure factor or radial distribution function, which describe pair-correlations alone. It is also found that a good estimation of the ring statistics can be obtained by presenting the persistent homology data through the Betti numbers or lifetime distributions. For these systems, persistent homology therefore gives access to clear information that is hard to extract otherwise.

Amorphous structures of materials such as graphene have more flexible network geometries that can be represented by models generated using a bond-switching algorithm. The increased distortion of the k -polygons in the network blurs (i) the band structure in a persistence diagram and (ii) the distinction between the first Betti

numbers associated with these polygons. Nonetheless, the ring statistics can be extracted, provided the disorder is not too large. The persistent homology method therefore delivers information on the regularity of rings that is unavailable from a ring-statistics analysis.

In experiments on two-dimensional systems, the measured images of the networks are often poorly resolved and can incorporate areas that cannot be fully imaged, *e.g.*, because of the appearance of holes. Persistent homology can provide a computationally tractable way of obtaining the ring structure from these images, which must otherwise be achieved by a brute-force path search. The benefits are more nuanced than simple speed concerns. By definition, persistent homology considers cycles appearing on different length scales, so can separate the rings from large holes in a defective sample.

As the bonding develops more flexibility, the efficacy of persistent homology may become less clear cut. At lower levels of disorder, the metrics described above still apply, but in highly disordered systems, the persistent cycles no longer reflect the primitive ring structure. Rather, the persistence diagram adopts a more generic and diffuse form. There is no evidence as yet that persistent homology is able to capture information on the medium range ring order, as for instance provided by the assortativity [41] or the Aboav-Weaire parameter [42–44], which provide information on ring-ring correlations.

The general findings for two-dimensional systems should also hold in the application of persistent homology to three-dimensional systems. The more ordered the network, the easier it should be to relate traditional measures of the physical structure to persistent topological features. Also, the increased dimensionality means that topological features in the second homology group are accessible, which gives additional information, this time on the voids that are present. An advantage of persistent homology is that the features from these different groups are trivially separable. One potential downside is that, while experimental real-space data is becoming widespread for two-dimensional systems, direct imaging is less feasible for three-dimensional structures. This lack of experimental information largely limits persistent homology calculations to computational models, so it is important that they are consistent with the available experimental information if the true nature of a network's topology is to be uncovered.

VI. CONCLUSION

Persistent homology was used to characterise the topological ordering in two-dimensional atomic networks in which the level of disorder was controlled systematically. The method requires only the atomic coordinates as input and delivers, as metrics, persistence diagrams, Betti numbers and lifetime distribution functions.

For low levels of structural disorder, the persistence diagrams feature distinct bands B_n that originate from

atoms separated by n bonds, which gives information on two and higher body correlations. Such information is not accessible from structure factors and radial distribution functions, which describe pair-correlations alone. The persistent cycles give a good representation of the primitive rings, and the ring statistics can be calculated from either the first Betti numbers or lifetime distribution functions.

With increasing structural disorder, features in the persistence homology metrics become increasingly broadened to the point where the persistent cycles and primitive rings no longer coincide. Although large primitive rings may exist, they can become so distorted that they are not calculated as single persistent cycles, but rather multiple smaller cycles. The distortion precludes the formation of the higher order bands seen for more ordered networks.

The findings for two-dimensional systems provide a starting point for the interpretation of the persistent homology metrics for three-dimensional systems. Increasing levels of disorder are, therefore, unlikely to lead to persistent cycles (from the first homology group) that coincide with the primitive rings found from a shortest-path search. Conversely, it may be feasible to follow processes such as crystallisation in which the ordering increases, and link quantitatively the structural evolution to the

appearance of features in persistent homology metrics.

SUPPLEMENTARY MATERIAL

The supplementary material contains movies showing the evolution with increasing disorder of the persistence diagram for (1) a single k -sided polygon with $k = 6$, (2) a set of single k -sided polygons with $k = 3, 4, 5, 6, 7, 8, 9$ or 10 , and (3) the bond-switch model with $K_r/K_\theta = 16$. It also contains figures showing the point clouds and Delaunay triangulation for several of the (4) triangle-raft and (5) bond-switch models with different levels of disorder. A link is given to code for generating the models.

ACKNOWLEDGEMENTS

DOM and MW are grateful for support from the EPSRC Centre for Doctoral Training in Theory and Modelling in Chemical Sciences (TMCS), under grant EP/L015722/1. PSS was supported by the Institute for Mathematical innovation at the University of Bath, Grant No. IMI/201920/011, and acknowledges helpful discussions with Dr James Hook. This paper conforms to the RCUK data management requirements. The data that support the findings of this study are available from the corresponding author upon reasonable request.

-
- [1] L. Wasserman, *Annu. Rev. Stat. Appl.* **5**, 501 (2018).
 [2] H. Edelsbrunner and J. Harer, in *Surveys on Discrete and Computational Geometry: Twenty Years Later*, Contemporary Mathematics, Vol. 453, edited by J. E. Goodman, J. Pach, and R. Richard Pollack (American Mathematical Society, Providence, Rhode Island, 2008) p. 257.
 [3] M. Saadatfar, H. Takeuchi, V. Robins, N. Francois, and Y. Hiraoka, *Nat. Commun.* **8**, 15082 (2017).
 [4] F. Jiang, T. Tsuji, and T. Shirai, *Water Resour. Res.* **54**, 4150 (2018).
 [5] L. Steinberg, J. Russo, and J. Frey, *J. Cheminform.* **11**, 48 (2019).
 [6] K. Xia, X. Feng, Y. Tong, and G. W. Wei, *J. Comput. Chem.* **36**, 408 (2015).
 [7] T. Nakamura, Y. Hiraoka, A. Hirata, E. G. Escolar, and Y. Nishiura, *Nanotechnology* **26**, 304001 (2015).
 [8] Y. Hiraoka, T. Nakamura, A. Hirata, E. G. Escolar, and K. Matsue, *Proc. Natl. Acad. Sci. USA* **113**, 7035 (2016).
 [9] Y. Onodera, S. Kohara, S. Tahara, A. Masuno, H. Inoue, M. Shiga, A. Hirata, K. Tsuchiya, Y. Hiraoka, I. Obayashi, K. Ohara, A. Mizuno, and O. Sakata, *J. Ceram. Soc. Japan* **127**, 853 (2019).
 [10] A. Gutierrez, M. Buchet, and S. Clair, *ChemPhysChem* **20**, 2286 (2019).
 [11] S. S. Sørensen, C. A. N. Biscio, M. Bauchy, L. Fajstrup, and M. M. Smedskjaer, *Sci. Adv.* **6**, eabc2320 (2020).
 [12] P. S. Salmon and A. Zeidler, *J. Stat. Mech.* **2019**, 114006 (2019).
 [13] X. Yuan and A. N. Cormack, *Comput. Mater. Sci.* **24**, 343 (2002).
 [14] C. Maria, J.-D. Boissonnat, M. Glisse, and M. Yvinec, in *Mathematical Software - ICMS 2014*, Lecture Notes in Computer Science, Vol. 8592, edited by H. Hong and C. Yap (Springer, Berlin, Heidelberg, 2014) pp. 167–174.
 [15] U. Fugacci, S. Scaramuccia, F. Iuricich, and L. De Floriani, in *Smart Tools and Apps in Computer Graphics - Eurographics Italian Chapter Conference*, edited by G. Pinatore and F. Stanco (The Eurographics Association, 2016) pp. 1–10.
 [16] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington, *EPJ Data Sci.* **6**, 17 (2017).
 [17] A. Okabe, B. Boots, K. Sugihara, and S. Nok Chiu, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, 2nd ed. (Wiley, Chichester, 2000).
 [18] A. Zomorodian and G. Carlsson, *Discrete Comput. Geom.* **33**, 249 (2005).
 [19] D. Ormrod Morley and M. Wilson, *Mol. Phys.* **117**, 3148 (2019).
 [20] D. Ormrod Morley, A. L. Thorneywork, R. P. A. Dullens, and M. Wilson, *Phys. Rev. E* **101**, 042309 (2020).
 [21] A. Kumar, M. Wilson, and M. F. Thorpe, *J. Phys.: Condens. Matter* **24**, 485003 (2012).

- [22] J. F. Shackelford and B. D. Brown, *J. Non. Cryst. Solids* **44**, 379 (1981).
- [23] A. Gervois, J. P. Troadec, and J. Lemaître, *J. Phys. A: Math. Gen.* **25**, 6169 (1992).
- [24] F. R. Eder, J. Kotakoski, U. Kaiser, and J. C. Meyer, *Sci. Rep.* **4**, 4060 (2014).
- [25] P. Y. Huang, S. Kurasch, A. Srivastava, V. Skakalova, J. Kotakoski, A. V. Krashenninnikov, R. Hovden, Q. Mao, J. C. Meyer, J. Smet, D. A. Muller, and U. Kaiser, *Nano Lett.* **12**, 1081 (2012).
- [26] D. Ormrod Morley, A. L. Goodwin, and M. Wilson, *Phys. Rev. E* **102**, 062308 (2020).
- [27] J. F. Shackelford, *J. Non. Cryst. Solids* **49**, 19 (1982).
- [28] F. Wooten, K. Winer, and D. Weaire, *Phys. Rev. Lett.* **54**, 1392 (1985).
- [29] G. T. Barkema and N. Mousseau, *Phys. Rev. B* **62**, 4985 (2000).
- [30] S. K. Jain and G. T. Barkema, *Phys. Chem. Chem. Phys.* **20**, 16966 (2018).
- [31] S. von Althaus, A. Kuronen, and K. Kaski, *Phys. Rev. B* **68**, 073203 (2003).
- [32] M. Bulacu, N. Goga, W. Zhao, G. Rossi, L. Monticelli, X. Periole, D. P. Tieleman, and S. J. Marrink, *J. Chem. Theory Comput.* **9**, 3282 (2013).
- [33] J. Kotakoski, A. V. Krashenninnikov, U. Kaiser, and J. C. Meyer, *Phys. Rev. Lett.* **106**, 105505 (2011).
- [34] L. Lichtenstein, C. Büchner, B. Yang, S. Shaikhutdinov, M. Heyde, M. Sierka, R. Włodarczyk, J. Sauer, and H. J. Freund, *Angew. Chem. Int. Ed.* **51**, 404 (2012).
- [35] L. Lichtenstein, M. Heyde, and H.-J. Freund, *J. Phys. Chem. C* **116**, 20426 (2012).
- [36] C. Büchner and M. Heyde, *Prog. Surf. Sci.* **92**, 341 (2017).
- [37] A. L. Lewandowski, P. Schlexer, C. Büchner, E. M. Davis, H. Burrall, K. M. Burson, W.-D. Schneider, M. Heyde, G. Pacchioni, and H.-J. Freund, *Phys. Rev. B* **97**, 115406 (2018).
- [38] L. Lewandowski, S. Tosoni, L. Gura, P. Schlexer, P. Marschalik, W.-D. Schneider, M. Heyde, G. Pacchioni, and H.-J. Freund, *Angew. Chem. Int. Ed.* **58**, 10903 (2019).
- [39] S. Shaikhutdinov and H.-J. Freund, *ChemPhysChem* **14**, 71 (2013).
- [40] Y. Onodera, S. Kohara, P. S. Salmon, A. Hirata, N. Nishiyama, S. Kitani, A. Zeidler, M. Shiga, A. Masuno, H. Inoue, S. Tahara, A. Polidori, H. E. Fischer, T. Mori, S. Kojima, H. Kawaji, A. I. Kolesnikov, M. B. Stone, M. G. Tucker, M. T. McDonnell, A. C. Hannon, Y. Hiraoka, I. Obayashi, T. Nakamura, J. Akola, Y. Fujii, K. Ohara, T. Taniguchi, and O. Sakata, *NPG Asia Materials* **12**, 85 (2020).
- [41] M. E. J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).
- [42] D. A. Aboav, *Metallography* **3**, 383 (1970).
- [43] D. Weaire, *Metallography* **7**, 157 (1974).
- [44] S. N. Chiu, *Mater. Charact.* **34**, 149 (1995).