

Research



Cite this article: Pedersen MC, Robins V, Mortensen K, Kirkensgaard JJK. 2020 Evolution of local motifs and topological proximity in self-assembled quasi-crystalline phases. *Proc. R. Soc. A* **476**: 20200170.
<http://dx.doi.org/10.1098/rspa.2020.0170>

Received: 13 March 2020

Accepted: 4 August 2020

Subject Areas:

applied mathematics, solid state physics, materials science

Keywords:

persistent homology, self-assembly, molecular dynamics, topology, quasi-crystal

Author for correspondence:

Martin Cramer Pedersen
e-mail: mcpe@nbi.ku.dk

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5099335>.

Evolution of local motifs and topological proximity in self-assembled quasi-crystalline phases

Martin Cramer Pedersen¹, Vanessa Robins²,
Kell Mortensen¹ and Jacob J. K. Kirkensgaard^{1,3}

¹Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

²Department of Applied Mathematics, Australian National University, Canberra, Australia

³Department of Food Science, University of Copenhagen, Copenhagen, Denmark

 MCP, 0000-0002-8982-7615; VR, 0000-0001-7118-8491;
KM, 0000-0002-8998-9390; JJKK, 0000-0001-6265-0314

Using methods from the field of topological data analysis, we investigate the self-assembly and emergence of three-dimensional quasi-crystalline structures in a single-component colloidal system. Combining molecular dynamics and persistent homology, we analyse the time evolution of persistence diagrams and particular local structural motifs. Our analysis reveals the formation and dissipation of specific particle constellations in these trajectories, and shows that the persistence diagrams are sensitive to nucleation and convergence to a final structure. Identification of local motifs allows quantification of the similarities between the final structures in a topological sense. This analysis reveals a continuous variation with density between crystalline clathrate, quasi-crystalline, and disordered phases quantified by ‘topological proximity’, a visualization of the Wasserstein distances between persistence diagrams. From a topological perspective, there is a subtle, but direct connection between quasi-crystalline, crystalline and disordered states. Our results demonstrate that topological data analysis provides detailed insights into molecular self-assembly.

1. Introduction

Since their initial reporting in 1984 by Shechtman *et al.* [1] and subsequent classification [2], quasi-crystals have attracted enormous interest from materials scientists, mathematicians and chemists alike. This interest culminated in the 2011 Nobel Prize in Chemistry, which was awarded to Shechtman for his discovery. Although initially regarded as an unusual and unlikely state of matter, experiments have shown quasi-crystalline phases emerging in a wide variety of systems such as: alloys [3,4], binary crystals [5], dendritic liquids [6] and star polymers [7]. The list is long and has been expanded dramatically in the last two decades. Simulations have established quasi-crystalline structures in polymeric systems [8], packings of polyhedra [9] and foams [10].

In the paper at hand, we focus our attention on the study published by Engel *et al.* [11] in which the authors describe the emergence of icosahedral quasi-crystals (IQC) from a one-component system governed by a mathematical model mimicking the behaviour of intermetallic interactions. The paper by Engel *et al.* focuses on deriving the phase diagram and characterizing the emerging phases; and poses the question: ‘How do atoms arrange themselves rapidly and with near structural perfection into a long-range ordered configuration without the guidance of a unit cell?’. In this study, we expand this understanding by characterizing the local arrangements that occur during the self-assembly of IQCs using time series of persistence diagrams—a new diagnostic tool developed within topological data analysis.

Persistent homology (PH) [12–14] is a method from the emerging field of topological data analysis, that quantifies the topological structure of a dataset over a range of length scales. To provide a stable signature, PH determines the range of length scales for which each topological feature (i.e. homology class) *persists*. For analysis of three-dimensional point clouds, PH is best computed using the sequence of alpha-shape subsets of the Delaunay triangulation [15]. By measuring and quantifying geometric properties using topology, PH signatures enable us to characterize the self-assembly process in terms of a few distinct, frequently occurring ‘polyhedral holes’, similar to Bernal’s seminal work on cavities in liquids [16].

PH has been used to analyse packings of granular material [17,18], the pore-space of reservoir rocks [19], nanoporous materials such as zeolites for gas storage [20], and atomic configurations in amorphous materials [21] to name just a few relevant applications in recent years. The method excels at extracting information about geometric configurations that create topological features such as handles and cavities in three-dimensional objects.

As PH orders the simplices in the Delaunay triangulation in an intuitive manner, the method relies on analysis of many-point constellations. Hence, it classifies structure on a larger scale than, for example, Steinhardt’s bond-orientational order parameter, but still in a local manner unlike, for example, Fourier transformation. Also, as PH does not directly rely on orientation (or a predefined neighbourhood), it is well suited for analysis of disordered and quasi-crystalline states. The gathering, clustering, and visualization of this information is automated and can be directly related to physical structures, whereas other analysis methods such as Minkowski tensors [22] or Steinhardt’s bond-orientational order parameter [23,24] tend to focus on characterizing particle-to-particle geometry (averaging the structure of nearest-neighbour particles).

Our analysis reveals the defining constellations of the simulated structures, allowing us to identify similarities and differences between phases in the simulated phase diagram and to monitor the emergence and disappearance of key geometric constellations during the self-assembly process. For the simulated quasi-crystalline states, we notably observe the emergence of pentagons and related pentagonal bipyramids (in line with the initial descriptions of the structure [11]). Using this information, we describe the convergence of our simulations from a topological perspective. We employ the Wasserstein metric for persistence diagrams to explore the relationships between the simulated structures and quantify the dimensionality of the topological phase space.

2. Methods and theory

(a) Simulations

Our molecular dynamics (MD) simulations are performed using HooMD-blue [25–27] and follow the methodology outlined by Engel *et al.* closely [11]. We reiterate the essential points in this section but redirect interested readers to the original paper for specifics. We simulate 4000 particles in a cubic box with periodic boundary conditions. The dimensions of the box are chosen so that the number density of particles is 0.03 per unit volume. The system is integrated as a Nose–Hoover NVT-ensemble using steps of $\Delta t = 0.005$. Two particles separated by a distance, r , are interacting via a so-called (simplified) oscillating pair potential (OPP) [11,28]:

$$V(r) = r^{-15} + r^{-3} \cos(k(r - 1.25) - \phi), \quad (2.1)$$

where k and ϕ are chosen for each simulation run and determine the emerging structure. Increasing k decreases the space between the minima, whereas increasing ϕ translates the minima of the potential towards higher values of r . The potential is set to zero after the first three oscillations. An example of an OPP is shown in electronic supplementary material, figure SI.7. We begin our simulations at the reported melting temperatures, T_m , [11] and cool our simulation box linearly to a temperature of 0.1 during the 10^8 simulation steps. We record the particle positions every 1000 frames during each simulation trajectory. Finally, we translate the simulated structures to put the centre-of-mass at the origin as seen in figure 2.

Engel *et al.* [11] consider the 10 configurations in electronic supplementary material, table SI.1, and figure 1 a good set of representatives for the quasi-crystalline region in their phase diagram, and we shall do the same. Furthermore, we simulate the crystal structures identified in their paper as well as an additional four parameter sets resulting in disordered structures. The names describe the density of the emerging structures: low-density quasi-crystals, named L1 to L3, intermediate-density quasi-crystals, labelled I1 to I4, and high-density approximants, called H1 to H3. We label the crystalline assemblies by their Pearson symbol [31] and the disordered states by D1 to D4.

Snapshots from the trajectory of the I2 phase are shown in the top row of figure 2. Short videos of the self-assembly process along with examples of calculated diffraction patterns verifying the existence of fivefold, threefold, and twofold axes in some of the simulated assemblies can be found in the electronic supplementary material. Our motivation for grouping the D1 to D4 phases stems from the fact that they do not exhibit long-range orientational order in their diffraction patterns.

(b) Persistent homology

Homology is a mathematical theory for characterizing the connectivity and holes in a shape. It does this by representing the shape as a cell complex (a collection of simple building blocks of dimensions $0, 1, 2, \dots$) and defining abelian groups, H_k , that encode relationships between k -dimensional cycles. When the cell complex sits in three-dimensional space, H_0 represents the connected components of the cell complex, H_1 the loops and H_2 the cavities.

Persistent homology extends this formalism from a single cell complex to a growing sequence of nested cell complexes C_α called a *filtration*, indexed by a scalar parameter α that is typically a length scale. The persistent homology module PH_k is an algebraic object associated with the filtration that encodes the k -dimensional homology at all α -values as well as the way the smaller cell complexes sit inside the larger ones [32]. This information is summarized by k sets of parameter intervals, $\text{PD}k = \{(b_i, d_i)\}$ a persistence diagram. A ‘birth’ endpoint b_i is the smallest α -value for which a particular k -cycle exists, and d_i is the value at which it ‘dies’, i.e. becomes the boundary of a $(k + 1)$ -dimensional piece of the cell complex. Some filtrations may have non-trivial homology in the final cell complex; these cycles are called *essential* and are assigned a death value of infinity.

In our setting, $\text{PD}0$ death values are determined by the minimal spanning tree of the Delaunay triangulation of the point cloud; essentially making it a histogram of nearest-neighbour distances.

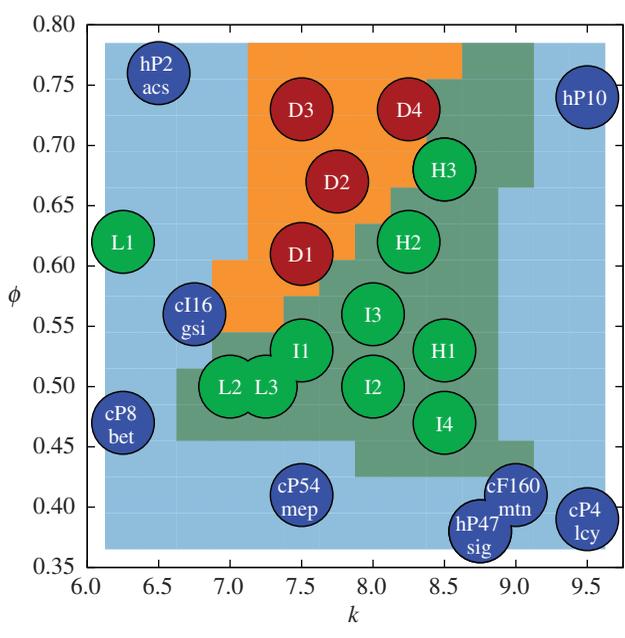


Figure 1. Visualization of our simulations in the phase diagram spanned by k and ϕ . Green points are quasi-crystalline phases, whereas blue points are crystalline and disordered points are red. The quasi-crystalline region found by Engel *et al.* is sketched in dark green, the crystalline region in teal, and the disordered region in orange. A small quasi-crystalline patch is hidden under the L1 disc. Where possible, the crystalline phases are denoted by their name in RCSR [29,30] below their Pearson symbols [31]. The figure is a reproduction of a figure by Engel *et al.* [11]. (Online version in colour.)

For the entries in PD1, the birth value is the length scale for which a loop is formed in the structure; e.g. as in the second panel of figure 3. Accordingly, the death value is the length scale for which the loop is ‘filled in’; as in the third panel of figure 3. Analogously, the births and deaths of the features in our PD2s are the distances for which all faces of a cavity are formed and the cavity ‘filled in’.

Any given frame from our MD trajectories can be considered as a point cloud in Euclidean three-space, \mathbb{E}^3 . There are a number of methods to build a filtration from this type of data. We are interested in the local geometric configurations of points, so in this study we use the union of balls of radius α growing around each data point, $X(\alpha) = \bigcup B(x, \alpha)$. The homology of $X(\alpha)$ is conveniently captured by the *alpha-shape*, a filtration of the Delaunay triangulation [33,34]. Two examples of alpha-shape filtrations are shown in figures 3 and 4.

The persistent homology of an alpha-shape filtration is particularly sensitive to short-range order. As seen in figure 2, locally ordered structures display just a few isolated features in their persistence diagrams in contrast to those of the initial purely random point cloud. The isolated PD features are associated with the repeated geometric motifs that generate cycles and cavities. A frozen crystalline point cloud would exhibit a distribution of isolated, delta-function-like features corresponding to the repeating geometric constellations. However, due to the soft nature of the OPP in equation (2.1), we observe broader distributions of the topological features in our persistence diagrams. Thus, the overall shape and distribution of points in a persistence diagram reflects the degree of order in the point cloud. The more smeared and dispersed the locations of the recorded topological features, the less geometric order the analysed structure possesses. A similar interpretation of persistence diagrams of simulated glass-forming systems can be found in the literature [21].

Analysing our alpha-shape filtration using persistent homology is effectively a physically meaningful way to identify clusters of Delaunay triangles and tetrahedra. Each point in PD2

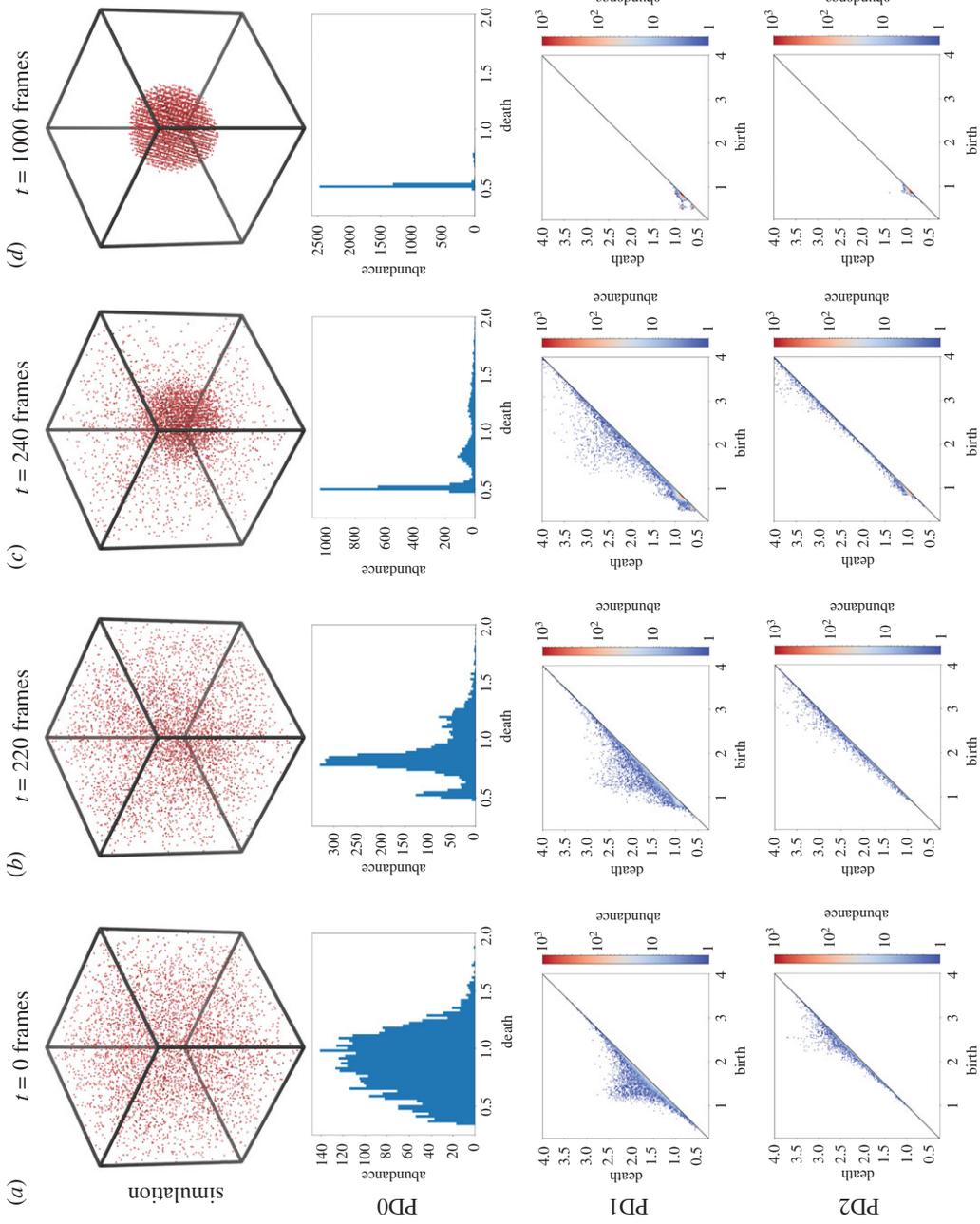


Figure 2. Examples of frames from a simulation of the I2 phase with their associated zeroth, first, and second persistence diagrams. From the initial random gas configuration in (a), we observe initial nucleation in (b) once the temperature has dropped sufficiently and the formation of a single droplet in (c). In (d), we see a single uniform droplet in the final frame after 10^8 steps in our simulation.

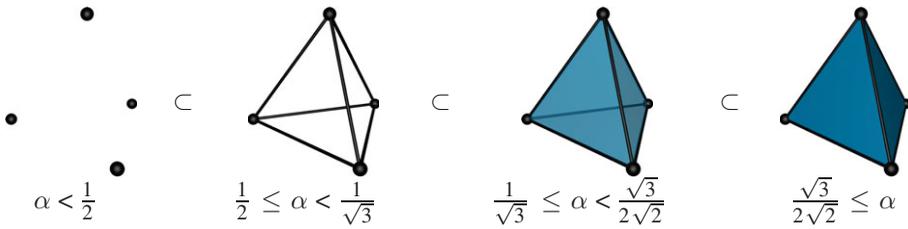


Figure 3. Alpha-shape filtration of the four corner points of a regular tetrahedron with edge length of 1. As α increases, we add simplices to the growing simplicial complex; hence each complex is a super-complex of the previous one. For $\alpha < 1/2$, our complex has the topology of four distinct points. At $\alpha = 1/2$, we connect these points by edges so that our complex now has three 1-cycles and one connected component. At $\alpha = 1/\sqrt{3}$, a 2-cycle is formed: the cavity in the centre of the tetrahedron. The 1-cycles from before have now become the boundary of this 2-cycle. Finally, at $\alpha = \sqrt{3}/2\sqrt{2}$, the 2-cycle disappears.

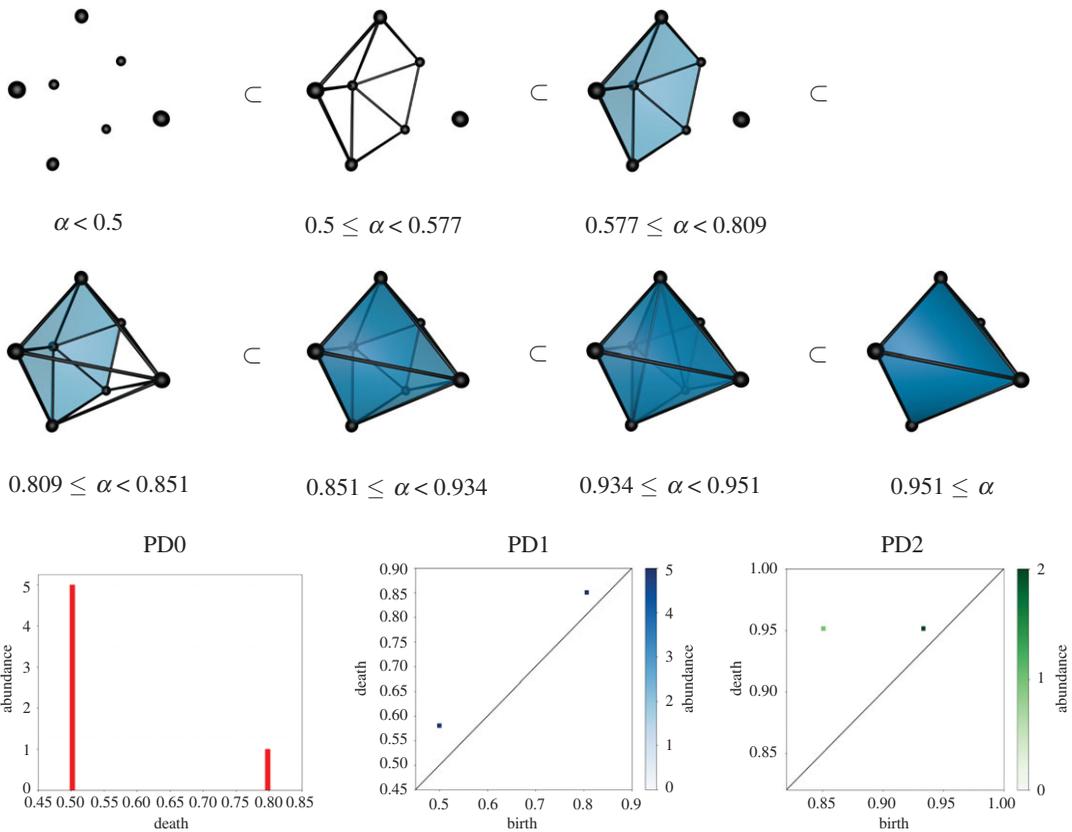


Figure 4. Top: alpha-shape filtration of the six corner points of an asymmetric pentagonal bipyramid with edge lengths of 1 or $\tau = (1 + \sqrt{5})/2 \approx 1.618$. As in figure 3, we add simplices to our complex as we increase α , which in turn changes the topology and homology of the complex. The decimal values are approximations of the exact values found in electronic supplementary material, tables SI.2 and SI.4. Consulting electronic supplementary material, tables SI.4 to SI.6, we see that the final Delaunay triangulation contains one large mixed tetrahedron, two skew tetrahedra (a), two small mixed tetrahedra, and one skew tetrahedron (b). Bottom: the persistence diagrams emerging from the filtration. (Online version in colour.)

corresponds to a ‘polyhedral hole’ as identified by Bernal in his pioneering work on atomic arrangements in ideal liquids. The birth and death values associated with each feature measure the largest enclosed sphere (d_i) and the largest sphere that can escape the cavity (b_i). PD1 encodes loop structures with birth value corresponding to the longest edge in the cycle when it is created, and death value the largest circular opening.

(c) Wasserstein distance

The space of persistence diagrams, \mathcal{D} , can be endowed with various metrics. One is the p th Wasserstein distance [35,36] (also known as the Kantorovich–Rubinstein distance or the Earth Mover’s distance) between two persistence diagrams, \mathcal{A} and \mathcal{B} :

$$l_p(\mathcal{A}, \mathcal{B}) = \left(\inf_{\gamma(a,b) \in \Gamma(\mathcal{A}, \mathcal{B})} \int_{\mathcal{D} \times \mathcal{D}} d(a,b)^p d\gamma(a,b) \right)^{1/p}, \quad (2.2)$$

where $\Gamma(\mathcal{A}, \mathcal{B})$ is the set of all pairings of points between diagrams \mathcal{A} and \mathcal{B} , and $d(a,b)$ is the metric on the space; in our case $\|a-b\|_\infty$. This metric allows us to quantify the differences between the persistence diagrams. Here, we use the second Wasserstein distance, $p=2$. Algorithms for computing this distance between two PDs are available in Dionysus2 [37].

(d) Software

Our persistent homology calculations are performed using the Dionysus2 [37] framework for Python as well as the Diode [38] extension for this, which in turn uses CGAL [39] libraries to construct the underlying alpha shapes. We employ HomCloud [40] for deriving the optimal cycle representations [41] of the features in our persistence diagrams. For the calculations using multidimensional scaling, we employed the algorithm distributed with ski-kit-learn [42] for Python. The plots and diagrams in this paper were made with Gnuplot [43], PyMol [44] and Matplotlib [45].

3. Results and initial observations

The structures emerging in our simulations corroborate the phase diagram and conclusions published by Engel *et al.* We consider time series of persistence diagrams from our simulation trajectories. Videos of examples of these series can be found in the electronic supplementary material, and example frames in figure 2. We discuss each dimension of homology in turn below.

For an alpha-shape filtration, all intervals in PD0 have $b=0$. So, rather than plot the (b,d) points, we display the one-dimensional histogram of d -values for each 0-cycle. The set of d -values in PD0 is known to be the same (with a factor of $1/2$) as the set of edge-lengths in the Euclidean minimal spanning tree built from the point cloud [46], so it captures the near-neighbour distances between points. We observe how shortly after nucleation, the d -values are distributed in three peaks corresponding to the three wells in the OPP. However, as shown in figure 5, as the self-assembly progresses, more and more particles have their nearest neighbour in the first (or second) well. In fact, we note that before nucleation, our PD0s are already exhibiting these peaks, implying that the point cloud has structure before nucleation; this is particularly clear from the video sequences in the electronic supplementary material. In the time series, we see that once the temperature is lowered sufficiently, the particles leap into the first well of the potential, and the system nucleates.

With the first glance at the sequence of PD1 plots in figure 2, we observe how the births and deaths of the topological features condense from the archetypal PD1 of a random point cloud into four distinct types of feature as seen in figure 5. As discussed earlier, the concentration of (b,d) points in a few locations implies the emergence of short-range order and of specific geometric arrangements of points.

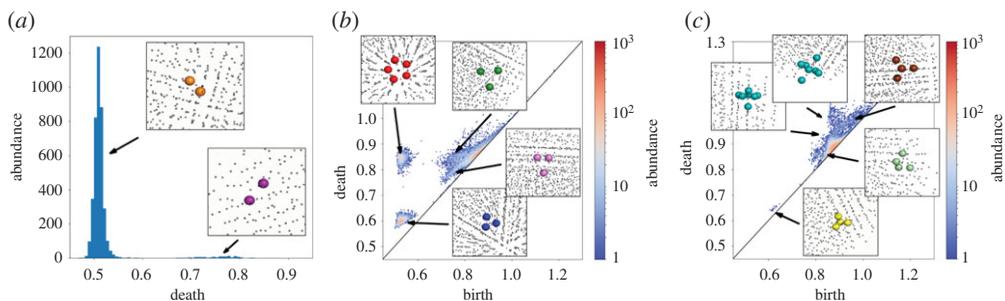


Figure 5. Identification of features in the zeroth (a), first (b), and second (c) persistence diagrams of the final frame for the 12 phase as shown in figure 2. Each characteristic feature in these diagrams can be attributed to the topology of a specific geometric shape. The configurations illustrated in PD1 and PD2 are examples of volume-optimal cycles [41] calculated using HomCloud [40]. We refer the reader to electronic supplementary material, tables SI.2, SI.4 and SI.5, for more details on the individual constellations. (Online version in colour.)

The condensation of features in PD2 highlights the same properties of the self-assembling structure. However, due to the type of cavities emerging, the classes of different features are harder to separate. In the electronic supplementary material, tables SI.2–SI.6, we calculate birth and death values of selected point constellations needed to explain the emergence of these defining features (of our quasi-crystalline phases) in our PD1s and PD2s. As seen in the rightmost entry of figure 5, our PD2s contain feature classes corresponding to regular and distorted tetrahedra and pentagonal bipyramids.

Figure 5 demonstrates the assignment of particular geometric constellations to different regions of our persistence diagrams. We draw the reader’s attention to the large pentagonal bipyramids shown in figure 5c in cyan and in electronic supplementary material, table SI.4. For the chosen values of k and ϕ in equation (2.1), the distance between the two tips of the bipyramid is roughly that of the bottom of the third well in the OPP. Engel *et al.* report that including the third well in the OPP (rather than shifting the potential to zero after two wells) in their simulations stabilizes and aids the formation of quasi-crystals. We conjecture that the inclusion of the well in the potential aids in stabilizing pentagonal bipyramids; thus supporting the formation of the quasi-crystalline structure.

The persistence diagrams from all of the simulated phases can be found in the electronic supplementary material, figures SI.2–SI.6. For the disordered phases, we observe how their PDs are clearly more ‘diffuse’. The structures appear to have more small, regular and semi-regular tetrahedral cavities than the other structures; in part due to the manner in which the first well in the potential is promoted for these values of k and ϕ .

4. Analysis and discussion

(a) Evolution of PD1 features

Due to the well-separated nature of the features in our first persistence diagrams, we analyse these further to characterize the self-assembly process (of the quasi-crystalline phases).

For each well in the simulated potential, we consider the width of that well to be the full width at half of the maximum (the difference between the potential value at the bottom of the well and at the peak of the subsequent hill). Based on these values, we define four windows for each of our quasi-crystals in which we expect (1) small, equilateral triangles, (2) pentagons, (3) large, equilateral triangles and (4) isosceles triangles to appear. The concept is sketched in electronic supplementary material, table SI.7, and explained there in detail. Note that these windows and their positioning differ slightly for the various simulation trajectories, as the simulated OPP is different for each one.

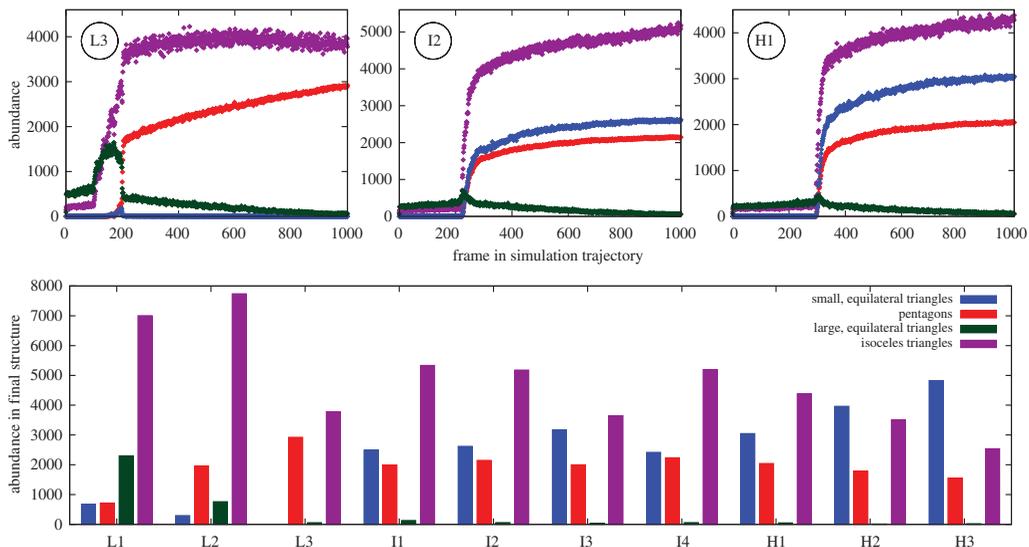


Figure 6. On top, examples of the evolution of the ‘intensity’ of selected spots (or rather the abundance of features inside our specified windows) in our PDIs for our simulation trajectories. Below, a comparison of the abundance of the same features for the final frames in all of our quasi-crystalline trajectories. (Online version in colour.)

Time evolutions of the abundance of topological features inside the specified windows are shown in figure 6. The full set of these time series is shown in electronic supplementary material, figures SI.8 and SI.9.

The analysis shows how—after the initial nucleation, which happens after around 220 frames in the three time series shown in figure 6—the abundance of features inside the specified windows differs between the simulated structures. We see how for the four chosen motifs, the large, equilateral triangles are unfavourable motifs which largely vanish during the self-assembly. Note that this is not the case for the L1 and L2 phase, where large, equilateral triangles appear thermodynamically competitive throughout the simulation.

As the structure is cooled, we note how different constellations become more and more prevalent; notably, the amount of pentagonal constellations continues to increase throughout the simulation for the samples shown in figure 6. We note that at the end of our simulation the structure has yet to fully equilibrate, as the amount of various features have yet to converge as we, for example, observe that the number of isosceles triangles is still increasing at the final frame in our trajectory.

In the lower half of figure 6, we plot the amount of the various features inside the discussed windows for the last frame in our trajectories as a means of distinguishing the different phases. We see how larger triangular constellations are preferred for low-density structures, whereas the smaller triangles are favourable for the simulated high-density structures.

Extending this methodology to the rest of the simulated structures, we correlate the abundance of the different features to each other in electronic supplementary material, figure SI.10; and to the estimated density of the structure in electronic supplementary material, figure SI.11. They show a continuity of structural motifs between the clathrates (cF160, cP54 and hP47), the quasi-crystals and the disordered phases. Furthermore, we see how the crystals cP4 and cP8 are quite distinct, whereas the PDIs of hP2 and hP10 stand out as containing local structural motifs that match higher-density, disordered structures.

(b) Convergence of Wasserstein distances

The quantity defined in equation (2.2) allows us to measure the distance between each frame in an MD trajectory and the final frame of the same trajectory, yielding a visualization of the

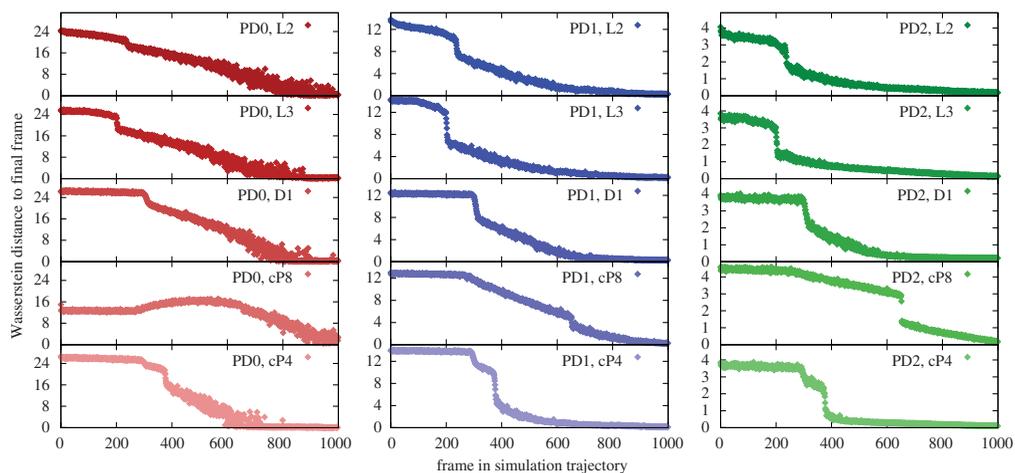


Figure 7. Examples of the evolution of the Wasserstein distance between a given frame in a trajectory and the final frame in the same trajectory. The full set of plots can be found in electronic supplementary material, figures SI.12 to SI.16. As the Wasserstein distance expresses differences between our persistence diagrams, we can think of these plots as detailing the manner in which our system converges in a topological sense. (Online version in colour.)

convergence that supports the feature counts studied in the previous section. Figure 7 shows examples of these time series. The full set of plots showing the convergence of the Wasserstein distances can be found in electronic supplementary material, figures SI.12–SI.16.

In most simulations, we observe two or three distinct stages in the Wasserstein distance time series. The first one or two stages are either roughly constant or slowly decreasing and this is followed by a noticeable drop at the same frame number in each dimension signalling nucleation. The final stage is a roughly monotonic but noisy decrease towards zero for the remainder of the trajectory. We observe a fundamentally different behaviour with respect to dimension in the convergence of PD0, PD1 and PD2 distances to final frames. After nucleation, the PD0 distances converge in a roughly linear fashion towards the final frame, while the PD1 and PD2 distances have a more exponential-like asymptotic convergence.

Examining the time series for different types of structures shows that the cP8 phase has a significantly different behaviour from the others, particularly for PD0 distances as it is the only one with a significant increase in distance for intermediate frames (300–600) before the final nucleation and convergence. The PD1 and PD2 distance time series also show roughly linear decrease over this time interval, before the nucleation jump and final asymptotic convergence. Note from electronic supplementary material, figure SI.4, that the final frame of the cP8 simulation has almost all particles with their nearest neighbour in the second well of the OPP. Closer inspection of the cP8 trajectory reveals that during the self-assembly process the number of particles whose nearest neighbour is in the first well of the OPP increases during the intermediate frames, leading to the increase in PD0 distances to the final frame.

In a few of the crystalline cases, e.g. cP4, we see two significant drops in the Wasserstein distance plots. Closer inspection of the evolution of PD0 histograms for the cP4 trajectory shows that this behaviour stems from the period in which a condensed droplet coexists with the gas. The second transition is easily identified in the video in the electronic supplementary material. We see that the nucleation jump tends to be significantly larger for the crystalline phase trajectories. Our interpretation of this is that the crystalline phase nucleation is topologically closer to its final state, whereas the disordered and quasi-crystalline phases equilibrate towards the final structure in a slower fashion.

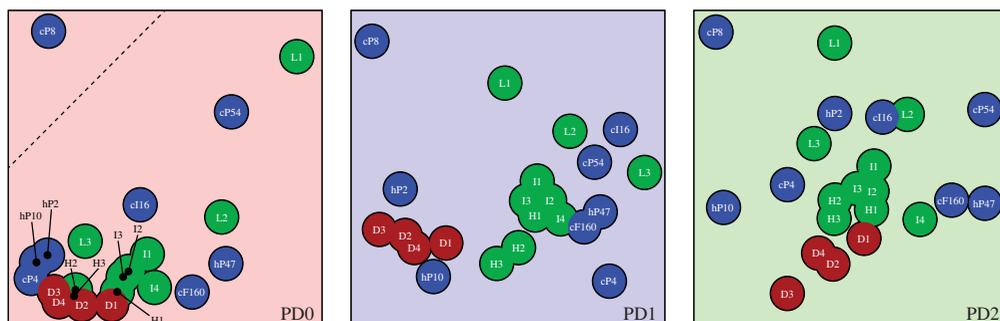


Figure 8. The simulated structures positioned by the MDS algorithm by attempting to space them according to the Wasserstein distances tabulated in electronic supplementary material, tables SI.8 to SI.10. The cP8 structure should be positioned considerably further away than shown in the PD0 diagram. As in figure 1, quasi-crystalline structures are plotted in green, crystalline in blue, and disordered in red. The axes of these diagrams are the coordinates of the embedding in arbitrary (but equal for the first and second axis) units. (Online version in colour.)

(c) Topological proximity of phases

Extending the methodology from the previous section, we calculate Wasserstein distances between different phases in an effort to quantify the topological relationship and similarity. These distances are tabulated in electronic supplementary material, tables SI.8 to SI.10, between the final frames in the trajectories.

We can visualize these relationships by employing multidimensional scale (MDS) [47] using the Wasserstein distances between the structures as input. The algorithm assigns a position in Euclidean space of prespecified dimension D to each of our simulated structures. These positions are chosen so that they minimize the squared error between the distances in the embedding and the provided distances. The two-dimensional embeddings are shown in figure 8. At first glance, we observe how the geometry and topology of the cP8 structure is rather different from any of the other structures for PD0, PD1 and PD2—as a consequence of the potential promoting the second well to a degree that renders the first well largely irrelevant.

The leftmost plot of figure 8 can be interpreted as a map over the density of the structures. Specifically, it displays the position and the occupancy of the wells in the OPP. We see how the structures plotted topmost in the diagram are the ones with the most particles having their nearest neighbour in the second potential well; cP8 being the extreme case of this. As we move downward, we encounter the denser structures. Glancing only at the leftmost plot, one would be tempted to label cP54 as an outlier as well. However, as we see in the PD1 and PD2 figure, the structure cannot be classified as an outlier in terms of the higher order homology. In general, we note that proximity in one of our diagrams does not necessarily preclude proximity in another.

The PD1 entry in figure 8 resonates well with our findings in figure 6. We see how the low-density IQCs are considered somewhat different as they promote larger triangular constellations rather than smaller. Similarly, cP54 and cI16—which are also low-density states—are found in the same region of the plot. In the bottom left of the figure, we find a clustering of states with high ϕ : hP2, hP10, and the disordered states. Finally, we note that cP4 is also found outside the main clusters as the structure contains few to no large triangular constellations. The clathrates, cF160, cP54 and hP47, can all be found somewhat centrally as the structures contain many pentagonal constellations giving the PD1s similar to those of the quasi-crystalline phases.

In the bottom left, we find the disordered states and the hP2 and hP10 states; structures where the majority of PD1 features are identified as small, triangular constellations. This is not the case for the cP8 structure, as these triangular constellations are formed by the second well in the OPP.

Lastly, the PD2 entry allows us to ascertain the similarity of the cavity structure of the simulated structures. We note how the aforementioned clathrates cF160, cP54 and hP47 can all be found at the right of the diagram. Similarly, we note how the second persistence diagram

of the hP10 structure exhibits unique clusters of features not found in other diagrams, which consequently makes it a slight outlier.

The disordered states appear to cluster in the bottom of our diagram. Glancing at their PD2s, we see that they are less condensed into sharp clusters of features than for the ordered states. In this frame of mind, we note that the D3 appears to be the least ordered state, placing it near the edge of the PD2 entry in figure 8. However, the clusterings of the simulated structures in figure 8 indicate that no single topological feature determines the crystallinity or disorder.

In electronic supplementary material, figure SI.17, we show the aforementioned error as a function of the dimension of the embedding, D , to give an indication of the effective dimensionality of the space we reduce using the MDS algorithm. We see that the two-dimensional embeddings in figure 8 can account for roughly 95%, 85% and 80% of the observed variance between the structures. To explain 95% of the variance, one would have to construct embeddings in \mathbb{E}^4 and \mathbb{E}^6 for the PD1 and PD2 entry, respectively. This result is corroborated by a similar analysis using principal component analysis (PCA) [48] of the homology rank functions [49] as shown in electronic supplementary material, figure SI.18; though we note that the PCA produces slightly lower dimensionalities. As shown in electronic supplementary material, figure SI.19, our PCA shows that the first principal score correlates with densities of the simulated structures, implying that density is a strong classifier for clustering.

5. Conclusions and outlook

We have demonstrated the application of persistent homology as an analysis tool for MD trajectories and self-assembly in general and on quasi-crystalline, disordered, and crystalline self-assemblies specifically and emphasize the generality of our approach. Persistent homology proved to be a powerful tool for quantifying and representing the local motifs; the method synergizes well with MD and is able to condense information and data from the technique in a way that is not easily obtained via other methods. In an automated manner, persistent homology identifies and clusters defining physical cavity and ring structures not captured in the Delaunay triangulation; in turn facilitating comparison between different structures and their homology.

From our persistence diagrams, we identified defining constellations of points in the various assemblies, which explains the observation by Engel *et al.* that the inclusion of the third well in the OPP aids in stabilizing quasi-crystals, as the third well favours pentagonal bipyramids in the structure.

Our approach highlights different aspects of the self-assembly behaviour for different phases. We observed how some constellations are used as intermediate configurations before the systems transition towards their final state, whereas others increase monotonically throughout the process.

We employed the Wasserstein distance as a means of quantifying the convergence of our simulations in a topological sense, and ultimately to relate the different simulations to each other. Using this metric, we grouped structures into different classes based on their topological proximity; demonstrating a concise and appealing way of relating larger numbers of simulated structures to each other.

Future work could elucidate the effect of the inclusion of the higher-order wells in the OPP, which is considered influential in the self-assembly process [11]. The relationship is not detectable using the approach here; one would have to employ higher order k th nearest neighbour diagrams for the analysis [50]. Based on our results here, we see this as a potential new way to understand phason behaviour and ultimately describe why quasi-crystals form long-range ordered configurations as posed in the introduction.

Our descriptions of the self-assembly of our phases relied on MD simulations of an interaction potential. However, other approaches are available for this task; as an example various implementation of field theories [51,52]. Persistent homology can also be applied to these descriptions of matter. If the field is an electron density, say, then the sequences of geometric shapes that form the filtration are built as upper level sets of electron density. This ensures that the

highest-density regions, which correspond to centres of particles in an MD simulation, generate PD0 births, while the lowest density regions, being cavities, appear as features in PD2. Algorithms for computing persistent homology from level sets of a real-valued function evaluated on a discretized grid are developed [53,54]. The resulting persistence diagrams can then be analysed using the same techniques as described here.

Data accessibility. This article has no additional data.

Authors' contributions. V.R., K.M. and J.J.K.K. conceived the idea of the study. M.C.P. planned and conducted the research with aid from V.R., K.M. and J.J.K.K. All authors discussed the analysis and presentation of the results. M.C.P. wrote the paper with aid from V.R., K.M. and J.J.K.K. All authors approved the final version.

Competing interests. The authors declare no competing interests.

Funding. The Villum Foundation is gratefully acknowledged by M.C.P. for financial support via grant nos. 17363 and 22833. V.R. is supported by the Australian Research Council as a Future Fellow on grant no. FT140100604.

Acknowledgements. The authors thank the staff at the Raijin supercomputer at NCI Australia for providing the computational resources needed for this project via the ANUMAS scheme. We thank the developers of the used software packages and in particular Dmitriy Morozov for his help with and insights into Dionysus' inner workings, and Kate Turner, Tomonari Dotera and Gerd Schröder-Türk for discussions on the manuscript and data analysis. We thank Engel *et al.* for an excellent paper and for making the information and software necessary to continue their work available.

References

1. Shechtman D, Blech I, Gratias D, Cahn JW. 1984 Metallic phase with long-range orientational order and no translational symmetry. *Phys. Rev. Lett.* **53**, 1951–1953. (doi:10.1103/PhysRevLett.53.1951)
2. Levine D, Steinhardt PJ. 1984 Quasicrystals: a new class of ordered structures. *Phys. Rev. Lett.* **53**, 2477–2480. (doi:10.1103/PhysRevLett.53.2477)
3. Guryan CA, Goldman AI, Stephens PW, Hiraga K, Tsai AP, Inoue A, Masumoto T. 1989 Al-Cu-Ru: an icosahedral alloy without phason disorder. *Phys. Rev. Lett.* **62**, 2409–2412. (doi:10.1103/PhysRevLett.62.2409)
4. Bindi L, Steinhardt PJ, Yao N, Lu PJ. 2011 Icosahedrite, Al₆₃Cu₂₄Fe₁₃, the first natural quasicrystal. *Am. Mineral.* **96**, 928–931. (doi:10.2138/am.2011.3758)
5. Talapin DV, Shevchenko EV, Bodnarchuk MI, Ye X, Chen J, Murray CB. 2009 Quasicrystalline order in self-assembled binary nanoparticle superlattices. *Nature* **461**, 964–967. (doi:10.1038/nature08439)
6. Zeng X, Ungar G, Liu Y, Percec V, Dulcey AE, Hobbs JK. 2004 Supramolecular dendritic liquid quasicrystals. *Nature* **428**, 157–160. (doi:10.1038/nature02368)
7. Hayashida K, Dotera T, Takano A, Matsushita Y. 2007 Polymeric quasicrystal: mesoscopic quasicrystalline tiling in ABC star polymers. *Phys. Rev. Lett.* **98**, 195502. (doi:10.1103/PhysRevLett.98.195502)
8. Dotera T, Gemma T. 2008 Dodecagonal quasicrystal in a polymeric alloy. *Phil. Mag.* **88**, 2245–2251. (doi:10.1080/14786430701861502)
9. Haji-Akbari A, Engel M, Keys AS, Zheng X, Petschek RG, Palffy-Muhoray P, Glotzer SC. 2009 Disordered, quasicrystalline and crystalline phases of densely packed tetrahedra. *Nature* **462**, 773–777. (doi:10.1038/nature08641)
10. Cox SJ, Graner F, Mosseri R, Sadoc JF. 2017 Quasicrystalline three-dimensional foams. *J. Phys. Condens. Matter* **29**, 114001. (doi:10.1088/1361-648X/aa5712)
11. Engel M, Damasceno PF, Phillips CL, Glotzer SC. 2015 Computational self-assembly of a one-component icosahedral quasicrystal. *Nat. Mater.* **14**, 109–116. (doi:10.1038/nmat4152)
12. Verri A, Uras C, Frosini P, Ferri M. 1993 On the use of size functions for shape analysis. *Biol. Cybern.* **70**, 99–107. (doi:10.1007/BF00200823)
13. Robins V. 1999 Towards computing homology from finite approximations. *Topology Proc.* **24**, 503–532.
14. Edelsbrunner H, Letscher D, Zomorodian A. 2002 Topological persistence and simplification. *Discrete Comput. Geom.* **28**, 511–533. (doi:10.1007/s00454-002-2885-2)
15. Edelsbrunner H, Kirkpatrick DG, Seidel R. 1983 On the shape of a set of points in the plane. *IEEE Trans. Inf. Theory* **29**, 551–559. (doi:10.1109/TIT.1983.1056714)

16. Bernal JD. 1960 Geometry of the structure of monatomic liquids. *Nature* **185**, 68–70. (doi:10.1038/185068a0)
17. Kramár M, Goulet A, Kondic L, Mischaikow K. 2014 Evolution of force networks in dense particulate media. *Phys. Rev. E* **90**, 052203. (doi:10.1103/PhysRevE.90.052203)
18. Saadatfar M, Takeuchi H, Robins V, Francois N, Hiraoka Y. 2017 Pore configuration landscape of granular crystallization. *Nat. Commun.* **8**, 15082. (doi:10.1038/ncomms15082)
19. Robins V, Saadatfar M, Delgado-Friedrichs O, Sheppard AP. 2016 Percolating length scales from topological persistence analysis of micro-CT images of porous materials. *Water Resour. Res.* **52**, 315–329. (doi:10.1002/2015WR017937)
20. Lee Y, Barthel SD, Dłotko P, Moosavi SM, Hess K, Smit B. 2017 Quantifying similarity of pore-geometry in nanoporous materials. *Nat. Commun.* **8**, 15396. (doi:10.1038/ncomms15396)
21. Hiraoka Y, Nakamura T, Hirata A, Escolar EG, Matsue K, Nishiura Y. 2016 Hierarchical structures of amorphous solids characterized by persistent homology. *Proc. Natl Acad. Sci. USA* **113**, 7035–7040. (doi:10.1073/pnas.1520877113)
22. Kapfer SC, Mickel W, Mecke K, Schröder-Turk GE. 2012 Jammed spheres: Minkowski tensors reveal onset of local crystallinity. *Phys. Rev. E* **85**, 030301. (doi:10.1103/PhysRevE.85.030301)
23. Steinhardt PJ, Nelson DR, Ronchetti M. 1983 Bond-orientational order in liquids and glasses. *Phys. Rev. B* **28**, 784–805. (doi:10.1103/PhysRevB.28.784)
24. Mickel W, Kapfer SC, Schröder-Turk GE, Mecke K. 2013 Shortcomings of the bond orientational order parameters for the analysis of disordered particulate matter. *J. Chem. Phys.* **138**, 044501. (doi:10.1063/1.4774084)
25. Glaser J, Nguyen TD, Anderson JA, Lui P, Spiga F, Millan JA, Morse DC, Glotzer SC. 2015 Strong scaling of general-purpose molecular dynamics simulations on GPUs. *Comput. Phys. Commun.* **192**, 97–107. (doi:10.1016/j.cpc.2015.02.028)
26. Anderson JA, Lorenz CD, Travesset A. 2008 General purpose molecular dynamics simulations fully implemented on graphics processing units. *J. Comput. Phys.* **227**, 5342–5359. (doi:10.1016/j.jcp.2008.01.047)
27. The Regents of the University of Michigan. 2020 HOOMD-blue. <https://glotzerlab.engin.umich.edu/hoomd-blue/index.html>.
28. Mihalkovič M, Henley CL. 2012 Empirical oscillating potentials for alloys from *ab initio* fits and the prediction of quasicrystal-related structures in the Al-Cu-Sc system. *Phys. Rev. B* **85**, 092102. (doi:10.1103/PhysRevB.85.092102)
29. O’Keeffe M, Peskov MA, Ramsden SJ, Yaghi OM. 2008 The reticular chemistry structure resource (RCSR) database of, and symbols for, crystal nets. *Acc. Chem. Res.* **41**, 1782–1789. (doi:10.1021/ar800124u)
30. O’Keeffe M, Delgado-Friedrichs O, Hyde S, Yaghi O. 2020 The Reticular Chemistry Structure Resource. See <http://rcsr.anu.edu.au/>.
31. Pearson WB. 1967 *A handbook of lattice spacings and structures of metals and alloys*, vol. 2. Oxford, UK: Pergamon Press.
32. Zomorodian A, Carlsson G. 2005 Computing persistent homology. *Discrete Comput. Geom.* **33**, 249–274. (doi:10.1007/s00454-004-1146-y)
33. Edelsbrunner H, Mücke E. 1994 Three-dimensional alpha shapes. *ACM Trans. Graph.* **13**, 43–72. (doi:10.1145/174462.156635)
34. Edelsbrunner H. 1995 The union of balls and its dual shape. *Discrete Comput. Geom.* **13**, 415–440. (doi:10.1007/BF02574053)
35. Vaserštejn LN. 1969 Markov processes over denumerable products of spaces describing large systems of automata. *Probl. Inform. Transm.* **5**, 47–52.
36. Dobrushin RL. 1970 Definition of a system of random variables by conditional distributions. *Teor. Veroyatnost. i Primenen.* **15**, 469–497.
37. Morozov D. 2020 Dionysus 2: a computational topology package focused on persistent homology. <http://mrzv.org/software/dionysus2>.
38. Morozov D. 2020 Diode: alpha shape generator using CGAL. <https://github.com/mrzv/diode>.
39. The CGAL Project. 2020 CGAL: The Computational Geometry Algorithms Library. <https://www.cgal.org>.
40. The Hiraoka Laboratory. 2020 HomCloud 2.5.1. See https://www.wpi-aimr.tohoku.ac.jp/hiraoka_lab/homcloud/.
41. Obayashi I. 2018 Volume-optimal cycle: tightest representative cycle of a generator in persistent homology. *SIAM J. Appl. Algebra Geom.* **2**, 508–534. (doi:10.1137/17M1159439)

42. Pedregosa F *et al.* 2011 Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830.
43. Williams T *et al.* 2020 Gnuplot 5.2: an interactive plotting program. <http://www.gnuplot.info/>.
44. Schrödinger LLC. 2020 The PyMOL Molecular Graphics System 2.2. <https://pymol.org/>.
45. Hunter JD. 2007 Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 99–104. (doi:10.1109/MCSE.2007.55)
46. Robins V. 2002 Computational topology for point data: Betti numbers of α -shapes. In *Morphology of condensed matter: physics and geometry of spatially complex systems* (eds K Mecke, D Stoyan), pp. 261–274. Berlin, Germany: Springer.
47. Kruskal JB. 1964 Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**, 1–27. (doi:10.1007/BF02289565)
48. Pearson K. 1901 On lines and planes of closest fit to systems of points in space. *Phil. Mag.* **2**, 559–572. (doi:10.1080/14786440109462720)
49. Robins V, Turner K. 2016 Principal component analysis of persistent homology rank functions with case studies of spatial point patterns, sphere packing and colloids. *Physica D* **334**, 99–117. (doi:10.1016/j.physd.2016.03.007)
50. Edelsbrunner H, Osang G. 2018 The multi-cover persistence of Euclidean balls. *34th Int. Symp. on Computational Geometry (SoCG), Budapest, Hungary, 11–14 June 2018*, article no. 34.
51. Lloyd DJ, Sandstede B, Avitabile D, Champneys AR. 2008 Localized hexagon patterns of the planar Swift-Hohenberg equation. *SIAM J. Appl. Dyn. Syst.* **7**, 1049–1100. (doi:10.1137/070707622)
52. Subramanian P, Archer AJ, Knobloch E, Rucklidge AM. 2018 Spatially localized quasicrystalline structures. *New J. Phys.* **20**, 122002. (doi:10.1088/1367-2630/aaf3bd)
53. Robins V, Wood PJ, Sheppard AP. 2011 Theory and algorithms for constructing discrete morse complexes from grayscale digital images. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**, 1646–1658. (doi:10.1109/TPAMI.2011.95)
54. Delgado-Friedrichs O, Robins V, Sheppard A. 2015 Skeletonization and partitioning of digital images using discrete morse theory. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 654–666. (doi:10.1109/TPAMI.2014.2346172)