

Quotient maps and configuration spaces of hard disks

Ozan B. Ericok^{1*} and Jeremy K. Mason¹

¹*Materials Science and Engineering, University of California, Davis, One Shields Avenue, Davis, 95616, CA, USA.

*Corresponding author(s). E-mail(s): oericok@ucdavis.edu;

Contributing authors: jkmason@ucdavis.edu;

Abstract

Hard disks systems are often considered as prototypes for simple fluids. In a statistical mechanics context, the hard disk configuration space is generally quotiented by the action of various symmetry groups. The changes in the topological and geometric properties of the configuration spaces effected by such quotient maps are studied for small numbers of disks on a square and hexagonal torus. A metric is defined on the configuration space and the various quotient spaces that respects the desired symmetries. This is used to construct explicit triangulations of the configuration spaces as α -complexes. Critical points in a configuration space are associated with changes in the topology as a function of disk radius, are conjectured to be related to the configurational entropy of glassy systems, and could reveal the origins of phase transitions in other systems. The number and topological and geometric properties of the critical points are found to depend on the symmetries by which the configuration space is quotiented.

Keywords: Configuration spaces, hard disks, phase transitions, topology, geometry

1 Introduction

The glass transition is a subject of ongoing study in condensed matter physics. Since it is related to a slowing down of the dynamics and is not accompanied by a change in any obvious structural order parameter, it is usually not considered to be a true thermodynamic phase transition. Recent computer simulations [1] suggest that the main difference between a glass and a liquid is the volume of configuration space that is available to both systems. The volume of configuration space relevant to a glassy system is often supposed to be proportional to the number of local minima of the potential energy surface. An accurate count of these minima would then allow the configurational entropy to be used as an order parameter

[1], and a popular strategy to enumerate potential energy minima was proposed by Goldstein [2] and formalized by Stillinger and Weber [3, 4]. The assumption underlying this view of the glass transition is that each local minimum of the potential energy surface corresponds to a different glassy state.

Local minima are specific examples of a larger class of points known as critical points, roughly defined as locations where the topology of a manifold changes. The number and distribution of critical points of the potential energy surface has also been implicated in the onset of phase transitions, an idea known as the Topological Hypothesis [5, 6]. Consider a system of particles with positions \bar{q}_i and potential energy $V(\bar{q}_1, \dots, \bar{q}_N)$. Previously, Franzosi et al. [7–9] claimed that a

change in the topology of the equipotential energy submanifolds $\Sigma_\nu = V^{-1}((-\infty, \nu])$ of the configuration space as a function of the energy ν is a necessary condition for a phase transition to occur in systems with smooth, stable, confining, and short-range interactions. Kastner and Mehta [10] eventually found a counterexample satisfying all the stated conditions, but for which a phase transition occurs without a change in topology. They then proposed new criteria stating that a phase transition requires either (i) the number of critical points in a narrow potential energy band to grow exponentially faster than the number of particles, or (ii) the determinant of the Hessian matrix to vanish for a significant fraction of the critical points. It is significant that the Topological Hypothesis, either the original or the revised one, has so far only been evaluated for systems simple enough to be treated at least partially analytically; there appears to not yet even be the machinery available to test the hypothesis for, e.g., a simple fluid.

In order to understand how the topology and geometry of configuration space are related to each other as a function of some thermodynamic variable, consider the one dimensional example system shown in Fig. 1. The potential energy landscape of this system consists of two local basins of attraction separated by an energy barrier. It is easy to identify that there are three distinct critical points of V , i.e., two local minima associated with the basins and one saddle point associated with the barrier. In this setting, an equipotential energy submanifold Σ_ν refers to the subset of the configuration space whose potential energy is less than a given value. Trivially, the submanifold Σ_ν is empty for energy levels $\nu < \nu_1$. When $\nu = \nu_1$, the topology of the empty set changes and the submanifold becomes the union of two points, i.e., the coordinates corresponding to the local minima. For energy levels $\nu_1 < \nu < \nu_3$, the disconnected regions corresponding to the basins grow, but stay separated until $\nu = \nu_3$. At this energy level, the saddle point appears and merges the previously disconnected components causing a topological change. Finally, for energy levels $\nu > \nu_3$, the topology of the submanifold does not change again and it grows as a single component. Suppose now that a random walker is initially positioned into the left basin with the energy level $\nu = \nu_2$ as shown by yellow in Fig. 1. Since the

submanifolds are disconnected at this energy level, the random walker can not make a transition to the right basin and can only explore the portion of the configuration space that is available to it. If the energy level of the walker is raised to $\nu = \nu_4$, the walker has now enough energy to explore both basins and. This is because previously disconnected components of the configuration space has already merged together at $\nu = \nu_3$. In other words, the topological change associated with the saddle point at $\nu = \nu_3$ causes a geometric chance in the accessible region of the configuration space for the walker.

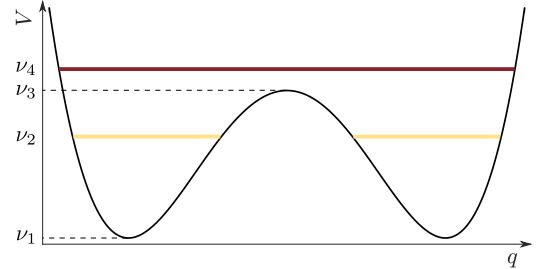


Fig. 1 Each equipotential energy submanifold Σ_ν consists of two disconnected components for energy levels $\nu_1 < \nu < \nu_3$. The topology of the submanifold changes at $\nu = \nu_3$ causing an associated geometric change in the accessible region.

In this sense, the Topological Hypothesis effectively associates the topological changes indicated by critical points with geometric changes in the accessible region of the configuration space. The specific relationship of the topology to the geometry depends on how the configuration space is constructed though. Initially consider fixing a coordinate system to identify points in a spatial region X , assigning labels to each of n particles, and representing every possible configuration of this system by a point in the product space X^n . This is not ideal from the standpoint of physically-distinguishable configurations though, e.g., a single configuration of identical particles with two distinct labelings corresponds to two distinct points in this configuration space. For specificity, consider the case of two hard disks in the hexagonal torus, with critical points as shown in Fig. 2. For a disk radius ρ , the accessible region of configuration space is the connected component of configurations where every pair of disk centers is separated by at least 2ρ . The geometry

of the accessible region can therefore be considered as a function of ρ , with the configuration space being empty in the large radius limit and the configuration space of points in the zero radius limit. As the disk radius decreases in the labeled configuration space on the left, two disconnected components initially appear. A system beginning in one of these sub-spaces cannot cross into the other unless the disk radius is further decreased, allowing the configuration to pass through one of the three saddle points. That is, the volume of the accessible region increases discontinuously at this disk radius. However, there is only ever one connected component in the unlabeled space on the right, making the volume of the accessible region a continuous function of disk radius. This has significant implications if the configurational entropy is defined as a function of the volume of the accessible region, since the configurational entropy would then be discontinuous on the left but not on the right.

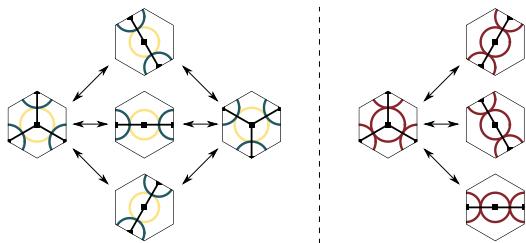


Fig. 2 The critical points of the translation invariant configuration spaces of two hard disks on a hexagonal torus (labeled on the left, unlabeled on the right). The configurations where the disks have three connections are local minima, and those with two connections are saddle points.

The canonical approach is to quotient the configuration space by all possible symmetries. For example, the homogeneity of space encourages the use of center of mass coordinates in classical mechanics [11]. When constructing regression functions for the potential energy of local atomic environments, it has been reported [12, 13] that using a configuration space that is invariant to translations, permutations and rotations decreases the number of training points required and increases the accuracy of the regression. However, the example in Fig. 2 suggests that quotienting by such symmetries could affect the geometry and topology of the configuration space in unexpected ways.

The hard disk system is often considered as a prototype for simple fluids [14]. It is governed by the hard disk potential, defined to be infinite if any pair of disk centers is separated by less than the sum of their radii and zero otherwise, and was first studied by Alder and Wainwright [15] almost sixty years ago. A number of studies suggest that the hard disk system undergoes at least one phase transition with varying packing fraction η of the disks, with the solid and liquid phases perhaps separated by an intermediate hexatic phase. A solid characterized by long-range translational and orientational order is observed when $\eta > 0.72$, whereas a liquid characterized by the absence of any long-range order is observed when $\eta < 0.70$ [16, 17]. The behavior in the $0.70 < \eta < 0.72$ interval is a subject of ongoing controversy. This was initially believed to be a two-phase region exhibiting large fluctuations in density, generally considered as a sign of a first-order phase transition. Halperin, Nelson [18] and Young [19] instead suggested that the transition could be of Kosterlitz–Thouless type, implying the existence of a hexatic phase in this interval. Conflicting results continue to be reported in the literature about the order of the transition and the phases involved. Marx et al. [16, 17] reported a single step first order phase transition, whereas Bernard and Krauth [20] and Engel et al. [21] reported a two step phase transition with a first order liquid-hexatic transition and a second order solid-hexatic transition. Given this controversy, an approach that could identify the onset of a phase transition from more fundamental considerations than a discontinuous change in the value of a thermodynamic quantity could resolve the question of what happens in the $0.70 < \eta < 0.72$ interval, and would likely be useful in a broader thermodynamic context as well. While we do not claim to complete such an undertaking here, the necessary machinery is developed and a case study suggests that such an approach is in principle possible.

Configuration spaces of hard disks have been studied previously [22, 23]. Ritchey [24] specifically studied the configuration spaces of hard disks on the hexagonal torus. They provided appropriate definitions of critical points in the configuration space of hard disks and the associated critical index in the context of Morse theory (a detailed explanation is provided in Secs. 2 and 3),

and examined the equivalence classes that critical points form under the action of translation, permutation and discrete lattice symmetries. A high density of critical points around the packing fraction of the solid-liquid transition indicated rapidly-changing configuration space topology there. This is suggestive of idea underlying the Topological Hypothesis, i.e., that a signature of two-dimensional hard disk melting should be visible in the distribution of critical points of the potential energy surface in the corresponding configuration space. One area not comprehensively addressed by this earlier work is the effect that quotienting out by the action of symmetry groups has on the number and distribution of the critical points.

Roughly speaking, a quotient map sends a set of points in the base configuration space differing only by the action of a symmetry group to a single point in the quotient space. That is, the quotient map collects configurations with identical physical properties into equivalence classes, and the quotient space describes how the equivalence classes should be related to one another to preserve sensible notions of similarity. The equivalence classes studied by Ritchey [24] effectively define a set of quotient maps and quotient spaces that are studied in more detail here. In this work, a quotient space always refers to a new space generated by quotienting the base configuration space under the action of a symmetry groups. The detailed definitions of these spaces are provided in the following sections.

As far as the authors know, explicit triangulation of the configuration spaces of hard disks, quotiented by symmetry groups or otherwise, has never been done before. Our purpose here is to establish that this can be accomplished using topological data analysis techniques, and to show that the resulting triangulation allows study of the topological and geometric properties of the configuration spaces. The approach is demonstrated for the comparatively simple but nontrivial cases of two hard disks on the square and hexagonal toruses. While these should not be expected to resolve what happens in the $0.70 < \eta < 0.72$ interval of the hard disk system in the thermodynamic limit, the insights gained from these simpler systems are envisioned as part of a larger effort to

develop a more precise formulation of the Topological Hypothesis, and eventually to evaluate whether such a hypothesis holds in practice.

More specifically, this article is concerned with using explicit triangulations of the configuration space to study the action of quotient maps induced by symmetry groups on the number and distribution of critical points. Constructing explicit triangulations of the configuration space and the various quotient spaces is not trivial even for two disks, and is sufficient to demonstrate many of the same concerns that will likely arise for more complicated systems. Three quotient spaces of the base configuration space are considered. The first quotients out only the translational symmetry. The second adds the permutation symmetry of the disk labels and the inversion. The third adds the discrete symmetries of the lattice implied by the boundary conditions. Distance functions that respect the topology of the spaces and appropriately identify symmetry-related points are proposed, and are essential to the study of these spaces. Explicit triangulations are constructed using the α -complex [25], and the isometric feature mapping (ISOMAP) algorithm [26] is used for dimensionality reduction.

Section 2 defines the configuration spaces of n disks of radius ρ using the tautological function. Section 3 briefly introduces concepts from classical Morse theory that are relevant to the discussion of critical points. Section 4 provides unambiguous definition of the symmetry groups considered here, and proposes closely-related distance functions on the base configuration space and all of the quotient spaces. Section 5 defines a procedure to map a hard disk configuration into a space with coordinates that are invariant to the desired symmetry groups. Finally, Sec. 6 presents and discusses the explicit triangulations of the quotient spaces as a function of disk radius.

2 Tautological function

The configuration space of n points on a torus T^2 is the product space of n toruses, or

$$\Lambda(n) = \{\mathbf{x} = (\bar{x}_1, \dots, \bar{x}_n) \mid \bar{x}_i \in T^2\}.$$

$\Lambda(n)$ is also considered as the base configuration space since all other spaces considered in this work can be derived from it. Figure 3 shows the square

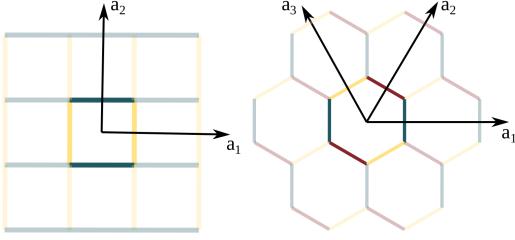


Fig. 3 A torus is obtained by identifying the opposite edges of a square (left) or a hexagon (right). These can be lifted to tilings of the plane, with the fundamental cells containing the origin and the periodic images shown in faint outline. The center to center distance of neighboring cells is always one.

and hexagonal toruses used in this study; periodic boundary conditions are imposed by identifying opposite edges of both domains. Two domains are studied to help separate the specific and general phenomena that can occur when quotienting a configuration space by the action of a symmetry group. More generally, any numerical study of the Topological Hypothesis for the hard disk system will require a choice of domain, and it will be necessary to distinguish what are consequences of that choice and what are nascent features of the thermodynamic system.

The tautological function $\tau : \Lambda(n) \rightarrow \mathbb{R}$ is defined as

$$\tau = \min_{1 \leq i < j \leq n} r_{ij}$$

where r_{ij} is half the geodesic distance between the centers of disks i and j . Intuitively, τ is the maximum radius that the disks could have without any pair overlapping given the positions of the disk centers. Observe that the configuration space

$$\Gamma(n, \rho) = \tau^{-1}([\rho, \infty)) \quad (1)$$

of n hard disks of radius ρ is the superlevel set of τ , or the set of all configurations that could accommodate disks of radius at least ρ .

3 Morse theory

Equation 1 represents the configuration space of hard disks by means of the superlevel sets of τ . This should allow a Morse-type theory to be used with the critical points of τ to identify changes in the configuration space topology. Classical Morse theory [27, 28] relates the topology of a manifold M to the critical points of a generic

smooth function f defined on that manifold. A critical point is defined as a point where the gradient ∇f of the function vanishes, and has a critical index equal to the number of negative eigenvalues of the Hessian matrix there. Intuitively, the critical index is the number of independent ways that one could move to decrease the value of f to second order.

At this point, it is worth mentioning that the critical points considered in this work are not the critical points of the potential energy surface on the configuration space $\Gamma(n, \rho)$, but the critical points of the tautological function τ on $\Lambda(n)$. Observe that $\Gamma(n, \rho)$ is some particular subset of $\Lambda(n)$ containing the points in $\Lambda(n)$ such that no overlapping will occur if disks of radius ρ are placed to these points. Keeping these in mind, the potential energy of each point in $\Gamma(n, \rho)$ is necessarily zero since they are guaranteed to be non-overlapping. In the context of Morse theory, the critical points of τ are then the locations where the topology of $\Lambda(n)$, hence that of $\Gamma(n, \rho)$ changes.

Let $M_a = \{x \in M \mid f(x) < a\}$ denote a sublevel set of M . The fundamental theorem of Morse theory states that the topology of M_a and M_b are the same if the interval $[a, b]$ doesn't contain a critical point. If it instead contains an index- p critical point, then the topology of M_a and M_b differ in a way that is equivalent to attaching a p -handle to M_a ; an n -dimensional p -handle is defined as a contractible smooth manifold $D^p \times D^{n-p}$ where D^p is the p -dimensional disk. For example, a 0-handle and a 2-handle in two dimensions are both two dimensional disks $D^0 \times D^2$ and $D^2 \times D^0$ (though they are attached in different ways), whereas a 1-handle is a rectangle $D^1 \times D^1$. The difficulty with this approach is that τ is not a smooth function, and in fact is not differentiable wherever the minimum disk separation is realized by more than one pair of disks. Our approach to handling this is explained elsewhere [24], but briefly, τ is replaced by a smooth function $E = \sum_{i < j} \exp[-w(r_{ij} - \rho)]$ that converges to the hard disk potential in the $w \rightarrow \infty$ limit. Moreover, there is a strictly monotone transformation of E that converges to τ in the same limit, suggesting that the critical points of τ be identified with the limiting critical points of E .

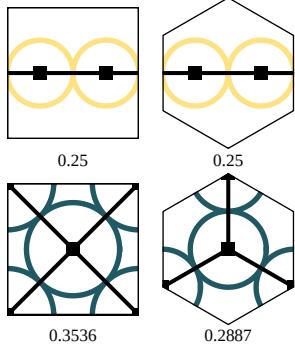


Fig. 4 Representatives of the equivalence classes of critical points for two disks on the square and hexagonal toruses. The bottom (top) row corresponds to index-0 (index-1) critical points. The disk radius is reported below each configuration.

Practically, the critical points of E for any finite w can be found by searching for the minima of the scalar function $|\nabla E|^2$ using, e.g., the conjugate gradient algorithm. Initializing the algorithm with random configurations samples critical points with a weight that depends on the construction of E . The sampled critical points are grouped into equivalence classes containing configurations related by symmetry operations. Representatives of the equivalence classes found after millions of initializations for $n = 2$ are shown in Fig. 4. Ritchey [24] suggests that every critical configuration is reproduced infinitely many times by rigid translations (usually handled by fixing one of the disks at the origin), $n!$ times by permuting the disk labels, and some number of times related to the order of the plane tiling's symmetry group.

4 Distance

The study of the configuration space geometry requires the definition of a suitable distance function. Depending on whether the space considered is the base configuration space or a quotient space, the distance could be defined between hard disk configurations or equivalence classes of configurations for given symmetry groups. For instance, the distance between two configurations that differ only by a translation should be nonzero in the base configuration space, but zero in the base configuration space modulo translations where they belong to the same equivalence class.

One natural notion of distance assigns to two configurations **on the base configuration space**

$\mathbf{p}, \mathbf{q} \in \Lambda(n)$ a distance equal to the sum of the disk displacements required to transform one into the other, or

$$d_\Lambda(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^n \|\bar{p}_i - \bar{q}_i\| \quad (2)$$

where $\|\bar{p}_i - \bar{q}_i\|$ is the geodesic distance between the two positions of the i th disk. Figure 5 shows these displacements for two configurations sampled uniformly at random on the base configuration spaces for the square and hexagonal toruses. Here, d_Λ is the sum of the lengths of the vectors pointing from one disk to the other. Observe that d_Λ is sensitive to symmetry operations in the sense that applying translations, permutations or lattice symmetries to one of the configurations changes d_Λ . That said, d_Λ satisfies the requirements of a metric on the base configuration space (identity of indiscernibles, symmetry, and the triangle inequality) with proofs provided in App. A.

The configuration space Λ equipped with a metric d_Λ constitutes a metric space (Λ, d_Λ) . Given a metric space and an equivalence relation \sim , there is a natural induced metric $d_{\Lambda/\sim}$ on the quotient space Λ/\sim [29]. When the equivalence relation additionally derives from a group of isometries \mathcal{S} , then the metric $d_{\Lambda/\mathcal{S}}$ on the quotient space Λ/\mathcal{S} can be written as

$$d_{\Lambda/\mathcal{S}}(\mathbf{p}, \mathbf{q}) = \inf_{S \in \mathcal{S}} \{d_\Lambda[\mathbf{p}, S(\mathbf{q})]\}. \quad (3)$$

Along with Eq. 2, this provides metrics on all the quotient spaces considered below.

Let \mathcal{T} , \mathcal{P} , \mathcal{I} and \mathcal{L} respectively be the sets of rigid translations, permutations of the disk labels, inversion about the origin, and symmetries of the tiling of the plane. Formally, a configuration \mathbf{q} is a translation of \mathbf{p} by \bar{t} if $\bar{q}_i = \bar{p}_i + \bar{t}$ for all disk indices i . Given a permutation $\pi \in \mathcal{P}$, \mathbf{q} is a permutation of \mathbf{p} if $\bar{q}_i = \pi(\bar{p}_i)$ for all i . A configuration \mathbf{q} is the inversion of \mathbf{p} if $\bar{q}_i = -\bar{p}_i$ for all i . Finally, for any symmetry element $L \in \mathcal{L}$ with representation \bar{L} , a configuration \mathbf{q} is a symmetric copy of \mathbf{p} if $\bar{q}_i = \bar{L}\bar{p}_i$ for all i . Observe that the operations belonging to all of these groups are isometric as required to use Eq. 3.

The three quotient spaces considered in this work are the base configuration space modulo translation Λ/\mathcal{T} , the base configuration space

modulo translations, permutations of disk labels, and inversion $\Lambda/\{\mathcal{T} \cup \mathcal{P} \cup \mathcal{I}\}$ and the base configuration space modulo translations, permutations of disk labels, inversion and lattice symmetries $\Lambda/\{\mathcal{T} \cup \mathcal{P} \cup \mathcal{I} \cup \mathcal{L}\}$. For the purposes of simplicity in the notation, Tab. 1 shows the symmetry groups by which the base configuration space is quotiented with a symbol \mathcal{S}_i . Table 1 also shows the distances between the configurations shown in Fig. 5 in the base configuration space and in the three quotient space considered. The $d_{\Lambda/S}$ are computed by fixing the first configuration and generating all copies of the second configuration that only differ by the action of \mathcal{S}/\mathcal{T} , i.e., the discrete symmetry elements. Finding the rigid translation $T \in \mathcal{T}$ that minimizes $d_{\Lambda}\{\mathbf{p}, T[S(\mathbf{q})]\}$ for $S \in \mathcal{S}/\mathcal{T}$ is a global optimization problem that is handled by the Tabu search algorithm [30, 31]. Finally, $d_{\Lambda/S}$ is reported as the minimum of these distances for all $S \in \mathcal{S}/\mathcal{T}$.

Table 1 Isometric symmetry groups applied to the configuration space, and the corresponding distances between the configurations in Fig. 5. \mathcal{T} , \mathcal{P} , \mathcal{I} and \mathcal{L} are the groups of translations, permutations, inversions, and symmetries of the tiling.

Space	Symmetries	$d_{\Lambda/S}^{Square}$	$d_{\Lambda/S}^{Hexagon}$
Λ	-	0.9048	0.5376
Λ/\mathcal{S}_1	\mathcal{T}	0.4396	0.4739
Λ/\mathcal{S}_2	$\mathcal{T} \cup \mathcal{P} \cup \mathcal{I}$	0.2780	0.2048
Λ/\mathcal{S}_3	$\mathcal{T} \cup \mathcal{P} \cup \mathcal{I} \cup \mathcal{L}$	0.1687	0.0704

The left column of Fig. 5 and the first row of Tab. 1 show the distance between configurations in the base configuration space Λ . The distance in Λ/\mathcal{S}_1 where $\mathcal{S}_1 = \mathcal{T}$ is the infimum of d_{Λ} over all rigid translations of one configuration with respect to the other, including those that translate the disks across the edge of the fundamental cell. The distance in Λ/\mathcal{S}_2 where $\mathcal{S}_2 = \mathcal{T} \cup \mathcal{P} \cup \mathcal{I}$ is additionally minimized over permutations of the disk labels (indicated by the uniform disk color) and inversion about the origin. The distance in Λ/\mathcal{S}_3 where $\mathcal{S}_3 = \mathcal{T} \cup \mathcal{P} \cup \mathcal{I} \cup \mathcal{L}$ is additionally minimized over the symmetries of the tiling, i.e., the symmetries of the square and hexagon. Observe that the distance between two configuration cannot increase (and generally decreases) as more symmetries are included.

5 Descriptors

As stated previously, the configuration space in Eq. 1 contains redundant information. Specifically, every configuration is equivalent to multiple other configurations related by the symmetry operations discussed by Ritchey [24]. As mentioned in the Introduction, quotienting the base configuration space by the translation group is so widespread that this operation is often not explicitly mentioned. The motivation to do so is that the resulting quotient space is much smaller than the base configuration space. That said, the quotient maps are such that it is often not clear how to explicitly parameterize the quotient spaces, though this would certainly facilitate the construction of an explicit triangulation. This section describes our procedure to do so.

Recall that the base configuration space for two disks is the product space $T^2 \times T^2$. Fixing the first disk at the origin effectively quotients the space by the translation group, making Λ/\mathcal{S}_1 equivalent to T^2 . This is explicitly parameterized starting with a rectangular region with edge lengths a and b centered at the origin in the plane. The torus formed by identifying opposite edges of the rectangle has major radius $R = a/2\pi$ and minor radius $r = b/2\pi$. The coordinates of this torus in R^3 are given by

$$\begin{aligned} x' &= (R + r \cos \theta) \cos \phi \\ y' &= (R + r \cos \theta) \sin \phi \\ z' &= r \sin \theta \end{aligned}$$

where $\phi = (a/2 - x)/R$ and $\theta = (b/2 - y)/r$. This is used for the visualizations of Λ/\mathcal{S}_1 below.

All other quotient spaces are initially embedded in an infinite-dimensional descriptor space, and a numerical approach is used to estimate the minimum number of descriptors necessary to maintain the embedding. **This idea is analogous to representing a function in the Fourier space, but using a finite subset of infinitely-many coefficients.** Given a configuration space of n disks, the distribution f is defined as a sum of Dirac-delta distributions $\delta(\bar{a}_j)$ located at the disk centers \bar{a}_j in the $a_1 a_2$ -coordinate system in Fig. 3, or

$$f(\bar{a}) = \sum_{j=1}^n \delta(\bar{a}_j) = \sum_{\bar{k}} c_{\bar{k}} \exp(2\pi i \bar{k} \cdot \bar{a}) \quad (4)$$

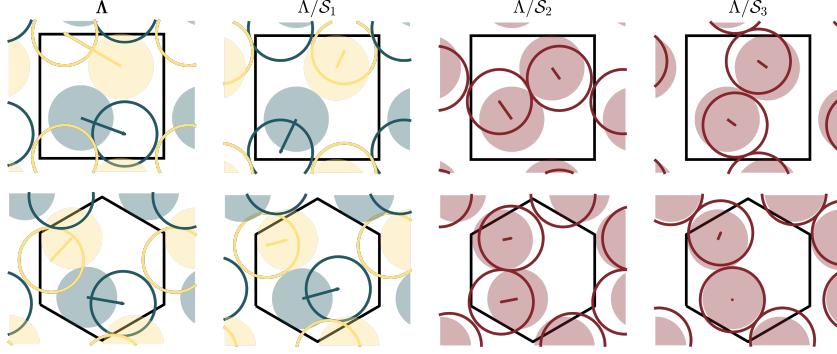


Fig. 5 Distances between two configurations in the square and hexagonal toruses. The two configurations are indicated by filled and empty circles, and colors indicate the labelling of the disks. Table 1 lists the symmetry groups used to construct the quotient spaces.

where $c_{\bar{k}}$ are the complex coefficients of the reciprocal lattice expansion and $\bar{k} = [p, q]$ for integers p and q . The infinite set of $c_{\bar{k}}$ can be calculated using the orthogonality of the complex exponentials as

$$c_{\bar{k}} = \sum_{j=1}^n \exp(-2\pi i \bar{k} \cdot \bar{a}_j). \quad (5)$$

$c_{\bar{k}}$ respects the periodicity of the lattice and is invariant to permutations of the disk labels due to the commutative property of the summation in Eq. 4. It can be shown that translating a configuration (by adding an offset to the \bar{a}_j) only changes the phase of the coefficients. This means that the moduli of the coefficients, or

$$z_{\bar{k}} = \sqrt{c_{\bar{k}}^* c_{\bar{k}}} \quad (6)$$

where $*$ denotes the complex conjugate, are a set of real-valued descriptors that are invariant to disk label permutations and rigid translations. Observe that the descriptors $z_{\bar{k}}$ also respect inversion symmetry. An illustration of the procedure above is provided in Fig. B1 in App. B. Numerical experiments indicate that the rank of the Jacobian of the map from the \bar{a}_j to the $z_{\bar{k}}$ is generically $2(n - 1)$, suggesting that some number of these descriptors could be sufficient to construct an embedding of Λ/\mathcal{S}_2 .

Constructing an embedding of Λ/\mathcal{S}_3 further requires the descriptors to be invariant to the symmetries of the plane tiling. This is done explicitly as

$$\hat{z}_{\bar{k}} = \frac{1}{O(\mathcal{L})} \sum_{L \in \mathcal{L}} z_{\bar{k}}^L \quad (7)$$

where $z_{\bar{k}}^L$ are the descriptors $z_{\bar{k}}$ of the configuration $L\mathbf{x}$, i.e., a copy of \mathbf{x} acted upon by the symmetry operation $L \in \mathcal{L}$, and $O(\cdot)$ is the order of a group.

Appendix B provides a proof that not all of these descriptors are independent. The invariance of the descriptors $z_{\bar{k}}$ to the inversion implies that the descriptors for indices \bar{k} and $-\bar{k}$ of a given configuration are the same for both the square and the hexagonal domains. The invariance of the descriptors $\hat{z}_{\bar{k}}$ to the symmetries of the plane tiling results in more complicated relationships that are fully described in App. B. The set of independent descriptors closest to the origin in reciprocal space is always used in the analysis below.

The maps into the infinite-dimensional spaces of descriptors are conjectured to be injective, i.e., to contain all information about the original configuration up to the desired symmetries. Since the number of disks is finite, it is likely that a finite number of dimensions (descriptors) is sufficient for this purpose though. The challenge then is to find the minimum number of descriptors necessary to maintain a proper embedding. The strategy proposed here is to order the descriptors by distance from the origin in reciprocal space, sequentially remove any dependent descriptors, and numerically search for self-intersections of the image space as a function of the number of descriptors retained after truncation.

Figure 6 illustrates the idea underlying the search for self-intersections. The full circle on the left represents the base configuration space, with points related by a symmetry operation in the same color. Quotienting by the symmetry group

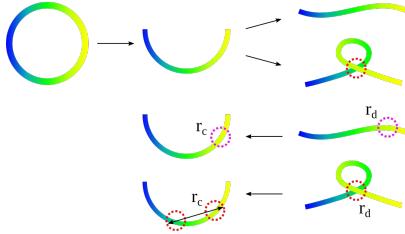


Fig. 6 An illustration of the self-intersection search. The full circle on the left represents the base configuration space, with points related by a symmetry operation in the same color. The middle half-circle represents the space quotiented by the symmetry group, and on the right are possible images of the map into a truncated descriptor space. One of these preserves the embedding, but the one that self-intersects (indicated by the red dotted circle) does not. The self-intersection is identified by considering the diameter of the preimage of a neighborhood around the intersection.

(folding the top half of the circle onto the bottom half) gives the quotient space represented by the half circle in the middle. On the right are possible images of the map of the quotient space into the truncated descriptor space. The number of descriptors could be sufficient for the image to be an embedding, as represented on the top right. The image could be self-intersecting if the number of descriptors is not sufficient though, as indicated by the region in the red dashed circle. The search for self intersections therefore involves sampling neighborhoods of radius r_d in the descriptor space and examining the preimages of these neighborhoods. If the radius r_c of the preimage scales with r_d for all such neighborhoods, then the map into the descriptor space is likely an embedding. If r_c appears to be independent of r_d for any neighborhood, then this is likely due to r_c measuring the distance between distinct neighborhoods in the preimage.

Practically, the procedure begins by sampling N configurations uniformly at random in the base configuration space. For each of these configurations, the first n_d descriptors that are invariant to the desired symmetries are computed. Small neighborhoods of radius r_d are then defined about the images of each configuration in the descriptor space; suppose that N_n images of other configurations lie within a particular neighborhood. The distances as defined in Sec. 4 are computed between these N_n configurations and the central configuration, and are used to estimate the radius r_c of the preimage in the quotient space. If r_c

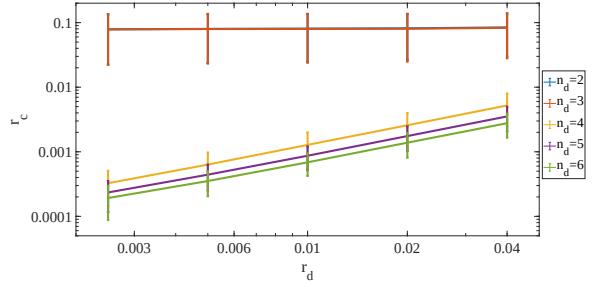


Fig. 7 The inverse analysis for $n_d = 2 \dots 6$ with different r_d values. The mean and standard deviation of r_c approach zero as r_d decreases for $n_d \geq 4$, suggesting that $n_d = 4$ is sufficient to embed the quotient space Λ/\mathcal{S}_2 .

goes to zero as r_d goes to zero for every neighborhood in the image, then the quotient space is likely embedded in the descriptor space. If not, then the image of the quotient space is likely self-intersecting as shown in Fig. 6, n_d is increased by one, and the process is repeated. Figure 7 shows the results of this analysis for the quotient space Λ/\mathcal{S}_2 and $n_d = 2 \dots 6$. It clearly shows that the mean and standard deviations of r_c go to zero as r_d goes to zero for $n_d \geq 4$, but not for $n_d \leq 3$. We conclude that four descriptors are sufficient to embed the quotient space Λ/\mathcal{S}_2 .

6 Configuration spaces

The map of the quotient space into the descriptor space can be viewed as a coordinate transformation, and the Jacobian matrix of the transformation can be found. The rank of this matrix gives the dimension of the resultant manifold at the point of evaluation [32]. Repeated sampling of the Jacobian matrix for the quotient space Λ/\mathcal{S}_2 and $n = 2$ disks suggests that the rank is generically two and that the image in the descriptor space is locally a 2-manifold. However, Fig. 7 suggests that at least four descriptors are required for the image in the descriptor space to be an embedding. Various dimensionality-reduction techniques can be used to try to reduce this further, enough to be able to visualize the space; the ISOMAP algorithm [26] is used here. Intuitively, this algorithm attempts to find a lower-dimensional embedding that preserves the geodesic distances of the points in k -nearest neighbor graphs.

Sampling hard disk configurations uniformly at random in the base configuration space and then computing the appropriate descriptors gives

a point cloud embedded in the truncated descriptor space. The study of the topological and geometric properties of the quotient space would be significantly simpler with a simplicial complex instead of a point cloud though. While there are a variety of simplicial complexes used in the literature on statistical topology (e.g., the Vietoris–Rips [33] and Cech [34] complexes), this work uses the α -complex [25] which is a subcomplex of the Delaunay triangulation [35]. Formally, let P be a set of points in R^d and Δ_k be a k -simplex where $0 \leq k \leq d$. Let r and c be the radius and the center of the circumsphere of Δ_k , respectively. Given the Delaunay triangulation $DT(P)$ of $P \subset R^d$, the α -complex $C_\alpha(P)$ of P is a simplicial subcomplex of $DT(P)$ such that a simplex $\Delta_k \in DT(P)$ is in $C_\alpha(P)$ if (i) $r < \alpha$ and the r -ball located at c is empty, or (ii) Δ_k is a face of another simplex in $C_\alpha(P)$.

A persistent question with α -complexes is the appropriate value of α . Our intention is to find a value such that the α -complex in the truncated descriptor space is a reasonable approximation of the quotient space. The heuristic used here involves a length scale analysis of the edges in the complex as a function of α . Let μ and σ respectively be the mean and standard deviation of the edge lengths. For very small α values, the α -complex contains only 0-simplices and a few 1-simplices and μ and σ are very small. For large α values, the α -complex approaches the full Delaunay triangulation, simplices that connect distant points are included, and μ and σ are large. For intermediate α values, there is presumably a plateau with intermediate values of μ and σ where the geometry of the complex is relatively stable (though this depends on the density of the sampled points). Any α within this plateau should be a reasonable value. An alternative would be to calculate the persistent homology as a function of α [36], but this would probably not provide significantly different values from the simpler length scale analysis used here. Figure 8 shows the result of this length scale analysis for the quotient space Λ/\mathcal{S}_1 , and suggests that $\alpha = 0.025$ is a reasonable value.

A lower bound on α is estimated as follows. Given n_p points in d dimensions, the Delaunay triangulation contains $O(n_p^{d/2})$ simplices [37]. This study always samples $n_p = 10^4$ points, giving

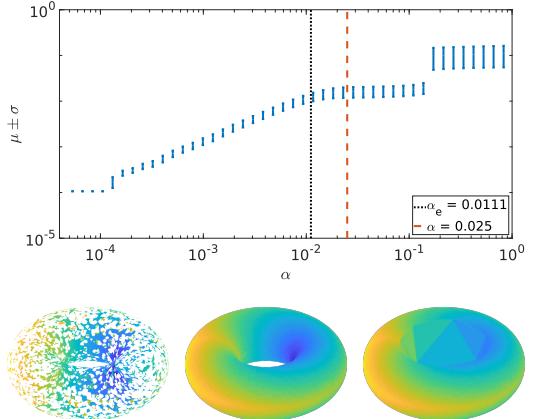


Fig. 8 The length scale analysis for Λ/\mathcal{S}_1 for the square torus. μ and σ denote the mean and standard deviation of the edge lengths of the α -complex. The black-dotted line shows the lower bound estimate α_e , whereas the red-dashed line shows the α value actually used to construct the α -complex. α -complexes for increasing α values $\{0.0001, 0.025, 0.5\}$ are shown at the bottom.

$n_t \approx 10^6$ tetrahedra in the full Delaunay triangulation of a 2-manifold embedded in R^3 . Assuming that the volume of the convex hull of Λ/\mathcal{S}_1 for two disks is covered by uniform equilateral tetrahedra would give $\alpha_e = 2^{1/6} \times (6V/n_t)^{1/3}$ for the tetrahedral edge length where V is the manifold's volume. Since the space for Λ/\mathcal{S}_1 is constructed using the rectangle $[0, 1] \times [0, 2]$, the lower bound is $\alpha_e = 0.0111$. As seen in Fig. 8, this estimate is conservative.

6.1 Adding translation invariance

The base configuration space Λ with the function τ is not amenable to Morse theory since the critical points of τ are not isolated; in fact, every critical point is related by a rigid translation to an entire critical submanifold. Partly for this reason the usual practice is to quotient out the rigid translations by, e.g., fixing the position of the first disk. This apparently innocuous operation can have the unexpected effect of identifying points related by a permutation of the disk labels though. For example, consider the index-0 critical point in the top row of Fig. 9. Translating the disks diagonally by the translation vector $\bar{t} = [0.5, 0.5]$ is equivalent to exchanging the disk labels, but is identified with the critical point on the left in the quotient space Λ/\mathcal{S}_1 . Likewise, translating the index-1 critical point in the middle row to the right by $\bar{t} = [0.5, 0]$ is equivalent to exchanging the

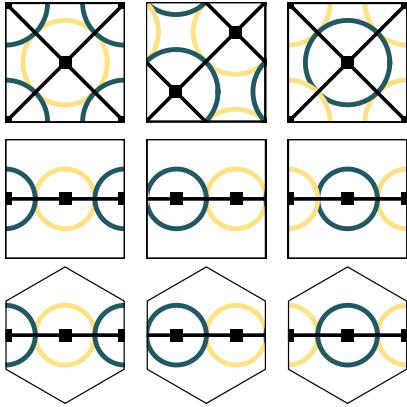


Fig. 9 Critical points can be related by both translational and permutation symmetries. This happens for both critical points of the square, but only for the index-1 critical point of the hexagon.

disk labels. That is, the submanifold that is identified when quotienting out by rigid translations can contain multiple points related by permutation symmetries. This implies that not all the equivalence classes of points related by permutation symmetries in Λ/\mathcal{S}_1 contain $n!$ elements, despite this being widely assumed (perhaps because each of these equivalence classes does contain $n!$ elements in Λ). Moreover, changing the domain of an integral from Λ/\mathcal{S}_1 to Λ/\mathcal{S}_2 is not generally as simple as dividing by a factor of $2n!$ (the factor of 2 for the inversion and $n!$ for the permutation group), despite this being standard practice in statistical mechanics [38, 39].

Figure 10 shows the translation-invariant configuration space (or quotient space $\Gamma(2, \rho)/\mathcal{S}_1$) of two disks as a function of ρ for the square torus (top) and hexagonal torus (bottom) as obtained from the α -complex of 10 000 points. Note that the square torus is constructed by extending the square to a rectangle and identifying opposite edges, but this does not affect the topological properties of the space. When $\rho > 0.25$, the space $\Gamma(2, \rho)/\mathcal{S}_1$ is comprised of a single 0-handle whereas that of the hexagonal torus is comprised of two 0-handles. This difference should be expected on the basis of Fig. 9 since the two index-0 critical points of the hexagonal torus are not related by a rigid translation. When $\rho = 0.25$, two and three 1-handles are connected for the square and the hexagonal toruses, respectively. Observe that the 1-handles provide connections between previously distant regions of the space. For $\rho <$

0.25, the space continues to grow and eventually closes in the $\rho \rightarrow 0$ limit. That is, the configuration with $\rho = 0$ acts like an index-2 critical point, even though it is not strictly within the space.

Figure 10 further confirms that some critical points are related by both translation and permutation symmetries, since the numbers of index-0 and index-1 critical points are, e.g., 1 and 2 instead of the 2 and 4 expected for the square torus on the basis of the symmetry group orders. Finally, the topology of Λ/\mathcal{S}_1 is that of a torus for both the square and the hexagon, as expected.

6.2 Adding permutation and inversion invariance

The descriptors $z_{\bar{k}}$ defined in Sec. 5 are by construction invariant to rigid translations, inversions about the origin, and permutations of disk labels. One way to construct the quotient space Λ/\mathcal{S}_2 is then to use the $z_{\bar{k}}$ as coordinates. Figure 7 suggests that four of these are sufficient for a proper embedding of Λ/\mathcal{S}_2 . The ISOMAP algorithm is used to reduce the dimension further by one, allowing visualization of the quotient space, but requires a distance function to do so. The top rows of Fig. 11 and Fig. 12 use the Euclidean distance using the corresponding descriptors, whereas the bottom rows use the distance defined in Eq. 3. This allows two versions of the quotient space $\Gamma(2, \rho)/\mathcal{S}_2$ to be constructed for both the square and hexagonal toruses; it is significant that the two versions are topologically identical, though the one using Eq. 3 better preserves the expected quotient space symmetries; analogous to the truncation of a Fourier series, the use of a distance based on a finite number of descriptors likely introduces distortions. Regardless, $\Gamma(2, \rho)/\mathcal{S}_2$ starts with the index-0 critical points and grows without topological change until $\rho = 0.25$ when the index-1 critical points appear. Unlike for $\Gamma(2, \rho)/\mathcal{S}_1$, these critical points don't appear as handles, but as singular points.

That critical points of the base configuration space do not behave in the same way in the quotient spaces should be emphasized; the index-1 critical points in Fig. 4 do appear in the $\Gamma(2, \rho)/\mathcal{S}_2$, but without any change in the topology. Instead, the critical points correspond to the appearance of sharp corners such that $\Gamma(2, 0.25)/\mathcal{S}_2$ cannot be described as a smooth manifold with boundary, but rather is a Whitney

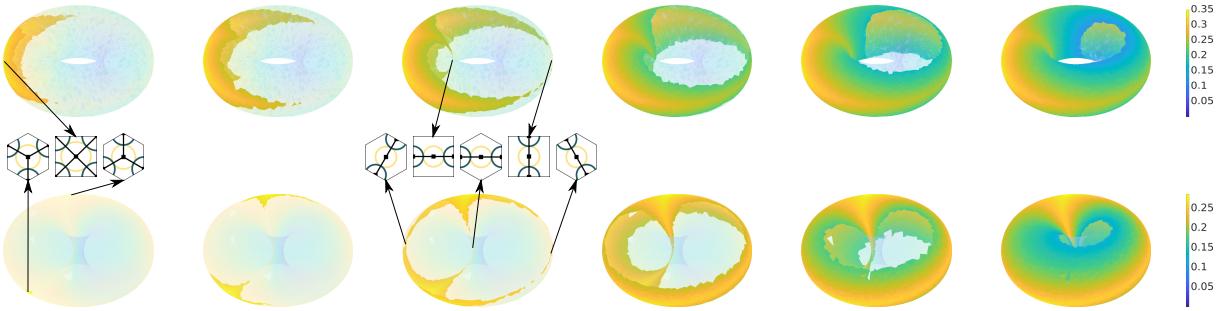


Fig. 10 The evolution of the translation invariant configuration space, or the quotient space $\Gamma(2, \rho)/\mathcal{S}_1$, for the square torus (top) and for the hexagonal torus (bottom) with $\rho = \{0.28, 0.26, 0.25, 0.21, 0.17, 0.12\}$. The locations of the critical points in this space are indicated by arrows.

stratified space. Finally, that the critical points do not connect distant regions of the space significantly affects certain geometric properties, e.g., the diameter of the space as measured by the diffusion distance [40]. As ρ is further decreased, the spaces continue to grow and eventually close up, indicating that the topology of the quotient space Λ/\mathcal{S}_2 is that of a sphere rather than a torus. That all of these changes occurred when merely quotienting out by permutations of the disk labels suggests that the ideas motivating the Topological Hypothesis need to be explored with great care.

6.3 Adding lattice invariance

The descriptors $\hat{z}_{\bar{k}}$ defined in Sec. 5 are additionally invariant to the symmetries of the plane tiling, and are used as coordinates for the embedding of the quotient space Λ/\mathcal{S}_3 . As before, dimensionality reduction is performed with the ISOMAP algorithm. The top rows of Fig. 13 and Fig. 14 use the Euclidean distance among the descriptors, whereas the bottom rows use the distance defined in Eq. 3. The two versions of $\Gamma(2, \rho)/\mathcal{S}_3$ are topologically identical as before. That said, the one using Eq. 3 better preserves the expected quotient space symmetries, with the geometric distortions introduced by using the Euclidean distance in the descriptor space much more pronounced than those in Fig. 11 and Fig. 12. Specifically, the version of $\Gamma(2, \rho)/\mathcal{S}_3$ constructed with the Euclidean distance incorrectly collapses the region for small ρ to a 1-manifold. Further examination suggests that the quotient spaces constructed with Eq. 3 are the smallest symmetric regions of their corresponding domains; the bottom row of Fig. 13 is 1/8 of the square torus, whereas that of Fig. 14

is 1/12 of the hexagonal torus. The corresponding fundamental cells can be obtained by reflecting the quotient spaces along an edge passing through the $\rho = 0$ point and applying the appropriate rotations.

Observe that the topology of the quotient space is completely changed by quotienting out the symmetries of the plane tiling. The index-0 critical point doesn't correspond to a 0-handle anymore, but to a single point, and the index-1 critical points are all identified by the symmetry operations. The $\rho = 0$ point appears as a single point as well, rather than as a 2-handle as in the other quotient spaces considered here. Finally, Λ/\mathcal{S}_3 has a boundary and is topologically equivalent to a disk, in contrast to Λ/\mathcal{S}_2 having the topology of a sphere and Λ/\mathcal{S}_1 that of a torus.

7 Conclusion

The configuration space is essential to the statistical mechanics of glass transitions and phase transitions, and a more thorough understanding of the configuration space could shed light on these phenomena. Specifically, the distribution of critical points of the potential energy surface could constrain the differentiability of the configurational entropy, and therefore regulate the onset of a phase transition. In an effort to simplify the analysis, the base configuration space is often quotiented by various symmetries, e.g., rigid translations and permutations of particle labels. An approach to explicitly triangulate these quotient spaces is established in this work, using techniques from topological data analysis. Descriptors invariant to the desired symmetry groups are proposed,

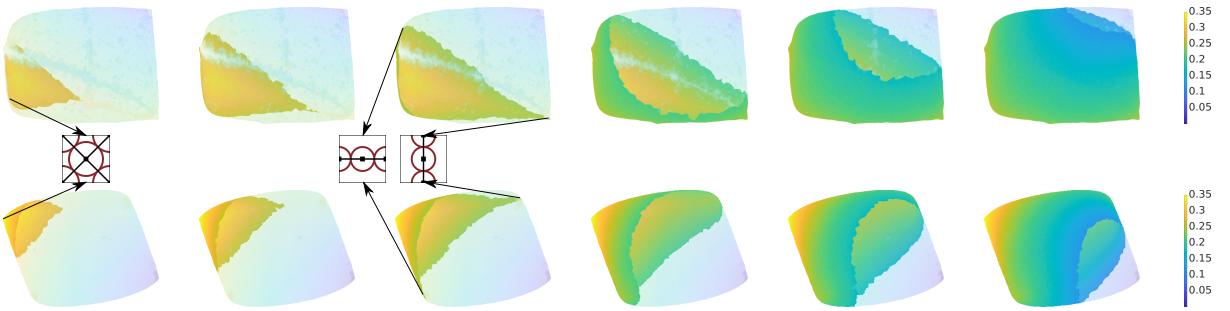


Fig. 11 The evolution of the translation, permutation and inversion invariant configuration space, or the quotient space $\Gamma(2, \rho)/\mathcal{S}_2$, for the square torus constructed with the standard Euclidean distance (top) and the distance in Eq. 3 (bottom) with $\rho = \{0.28, 0.26, 0.25, 0.21, 0.17, 0.12\}$. The locations of the critical points are indicated by arrows.

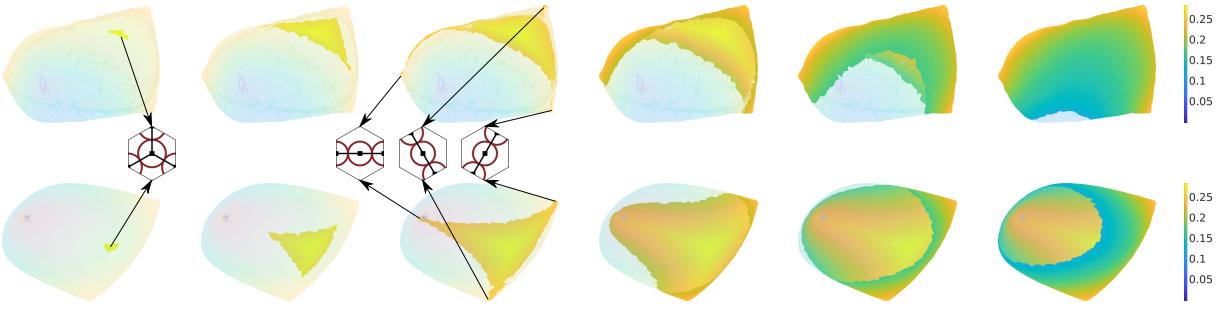


Fig. 12 The evolution of the translation, permutation and inversion invariant configuration space, or the quotient space $\Gamma(2, \rho)/\mathcal{S}_2$, for the hexagonal torus constructed with the standard Euclidean distance (top) and the distance in Eq. 3 (bottom) with $\rho = \{0.28, 0.26, 0.25, 0.21, 0.17, 0.12\}$. The locations of the critical points are indicated by arrows.

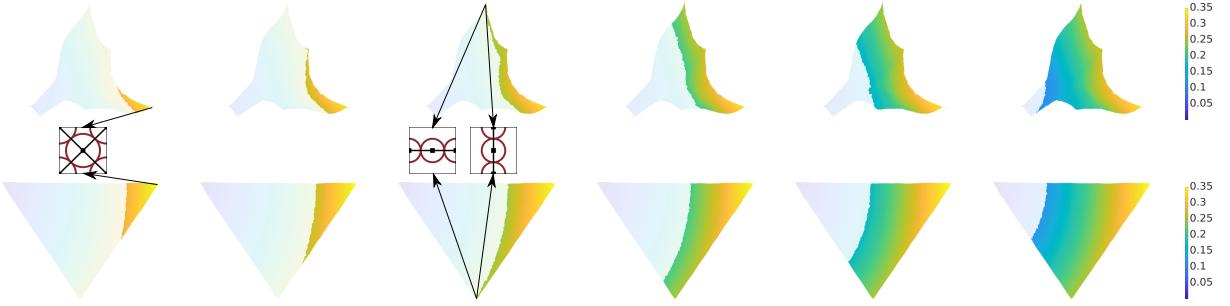


Fig. 13 The evolution of the translation, permutation, inversion and lattice symmetry invariant configuration space, or the quotient space $\Gamma(2, \rho)/\mathcal{S}_3$, for the square torus constructed with the standard Euclidean distance (top) and with the distance in Eq. 3 (bottom) with $\rho = \{0.28, 0.26, 0.25, 0.21, 0.17, 0.12\}$. The locations of the critical points are indicated by arrows.

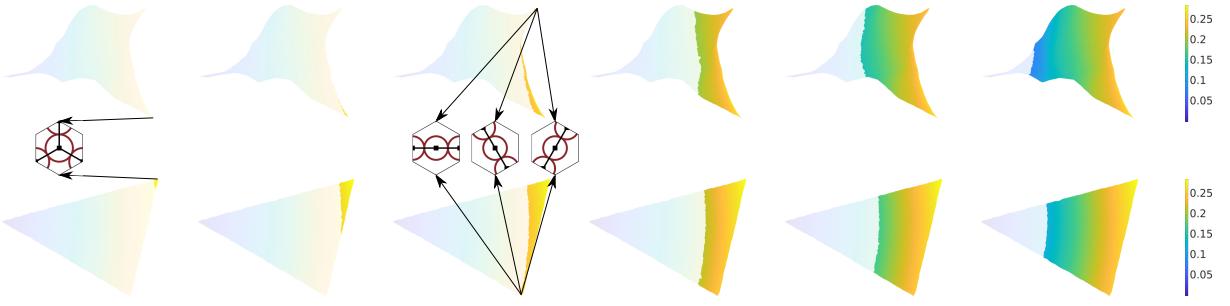


Fig. 14 The evolution of the translation, permutation, inversion and lattice symmetry invariant configuration space, or the quotient space $\Gamma(2, \rho)/\mathcal{S}_3$, for the hexagonal torus constructed with the standard Euclidean distance (top) and with the distance in Eq. 3 (bottom) with $\rho = \{0.28, 0.26, 0.25, 0.21, 0.17, 0.12\}$. The locations of the critical points are indicated by arrows.

allowing the various quotient spaces to be parameterized. Two distance functions are provided, one induced by the quotient map and the other the Euclidean distance in the descriptor space. These allow the construction of explicit triangulations of the quotient spaces as α -complexes, and thereby offer new approaches to studying the hard disk system. Specifically, the topological and geometric properties of the spaces can be directly evaluated as functions of disk radius. Some of the machinery developed is expected to be useful in other contexts as well, e.g., the proposed distance functions could be used to analyze the similarity of hard disk configurations generated by Monte Carlo simulations.

The procedure to triangulate the configuration space is developed and applied to the simple but nontrivial cases of two hard disks in the square and hexagonal toruses. The first finding is that the use of a square or hexagonal torus does not substantially affect the topology of the quotient spaces except for the number of critical points of the tautological function τ ; the overall properties of the spaces are otherwise similar. The second finding is that the number and behavior of the critical points depends on the construction of the quotient space. For example, some of the index-1 critical points are identified with one another when the base configuration space is quotiented by rigid translations. The third finding is that the topology and the geometry of the quotient spaces change dramatically as additional symmetries are quotiented out. For example, the superlevel sets of τ can no longer be described as manifolds with boundaries, and instead need to be described as stratified spaces. The $\rho = 0$ configuration, which is not identified as a critical point in the context of classical Morse theory, consistently behaves as an index-2 critical point that closes the space.

Even though this work considers only a pair of hard disks, extending and applying the techniques to the configuration spaces of more hard disks should be conceptually straightforward. The main obstacle is likely to be that the computational complexity of the distance defined in Eq. 3 grows as $n!$ (the order of the permutation group). Another future direction could be to use the stratified Morse theory of Goresky and MacPherson [41] to more thoroughly analyze the effects of the quotients maps on the topology of the spaces.

Acknowledgments. O.B.E. and J.K.M. were supported by the National Science Foundation under Grant No. 1839370.

Declarations

Data and code availability. The datasets and source codes associated with this study are available in the "quotient_hard_disks" repository, https://github.com/burakericok/quotient_hard_disks.

Compliance with ethical standards

Conflict of interest. The authors declare that they have no conflict of interest.

Appendix A Proof that d_Λ is a metric

This section proves that the distance function in Eq. 2 is a metric. Let \bar{x} be a column vector in R^2 and \bar{P} be a projection matrix whose rows contain the unit vectors of the square or hexagonal domain as in Fig. 3. By convention, the fundamental cell is defined as the set of points \bar{x} such that $w_j \in [-0.5, 0.5]$ for all j and $\bar{w} = \bar{P}\bar{x}$. Let $\mathbf{p}, \mathbf{q}, \mathbf{r} \in \Lambda(n)$ be three distinct configurations in the following.

First, the proposed distance function satisfies the identity of indiscernibles, or $d_\Lambda(\mathbf{p}, \mathbf{q}) = 0 \iff \mathbf{p} = \mathbf{q}$.

Proof That $\mathbf{p} = \mathbf{q} \implies d_\Lambda(\mathbf{p}, \mathbf{q}) = 0$ is true by inspection. For the other direction, observe that $d_\Lambda(\mathbf{p}, \mathbf{q}) = 0$ implies that $\|\bar{p}_i - \bar{q}_i\| = 0$ for all i . Consider the i th disk, and drop the index in the following.

For the square torus in Fig. 3, the geodesic distance reduces to

$$\begin{aligned} \|\bar{p} - \bar{q}\| &= \sqrt{a^2 + b^2} \\ a &= \min(|p_x - q_x|, 1 - |p_x - q_x|) \\ b &= \min(|p_y - q_y|, 1 - |p_y - q_y|) \end{aligned}$$

That $\|\bar{p} - \bar{q}\| = 0$ implies that $a = 0$ and $b = 0$. Since \bar{p} and \bar{q} are assumed to be in the fundamental cell, $p_x, q_x \in [-0.5, 0.5]$ and $|p_x - q_x| < 1$. Then $a = 0$ requires that $|p_x - q_x| = 0$, or that $p_x = q_x$. $b = 0$ implies that $p_y = q_y$ by a similar argument, or that $\|\bar{p}_i - \bar{q}_i\| = 0$ if and only if $\bar{p}_i = \bar{q}_i$. Then $d_\Lambda(\mathbf{p}, \mathbf{q}) = 0 \implies \mathbf{p} = \mathbf{q}$.

For the hexagonal torus in Fig. 3, $\bar{P}_h = [1, 0; 1/2, \sqrt{3}/2; -1/2, \sqrt{3}/2]$ and the geodesic distance reduces to

$$\begin{aligned}\|\bar{p} - \bar{q}\| &= \{\min[a^2 + b^2, (1-a)^2 + b^2, \\ &\quad (a-0.5)^2 + (b-\sqrt{3}/2)^2]\}^{1/2} \\ a &= |p_x - q_x| \\ b &= |p_y - q_y|.\end{aligned}$$

Let $\bar{P}_h\bar{p} = \bar{t}$ and $\bar{P}_h\bar{q} = \bar{w}$. Since \bar{p} and \bar{q} are assumed to be in the fundamental cell, $t_j, w_j \in [-0.5, 0.5]$ for all j . The seven possible ways for $\|\bar{p} - \bar{q}\| = 0$ are shown in Tab. A1. Observe that only the first satisfies the assumption that $w_j \in [-0.5, 0.5]$ for all j , and therefore $\bar{t} = \bar{w}$. This implies that $\|\bar{p}_i - \bar{q}_i\| = 0$ if and only if $\bar{p}_i = \bar{q}_i$. Then $d_\Lambda(\mathbf{p}, \mathbf{q}) = 0 \implies \mathbf{p} = \mathbf{q}$.

□

Second, the proposed distance function is symmetric, or $d_\Lambda(\mathbf{p}, \mathbf{q}) = d_\Lambda(\mathbf{q}, \mathbf{p})$, by the symmetry of $\|\bar{p}_i - \bar{q}_i\|$.

Third, the proposed distance function satisfies the triangle inequality, or $d_\Lambda(\mathbf{p}, \mathbf{r}) \leq d_\Lambda(\mathbf{p}, \mathbf{q}) + d_\Lambda(\mathbf{q}, \mathbf{r})$.

Proof The triangle inequality can be explicitly rewritten as $\sum_{i=1}^n \|\bar{p}_i - \bar{r}_i\| \leq \sum_{i=1}^n \|\bar{p}_i - \bar{q}_i\| + \sum_{i=1}^n \|\bar{q}_i - \bar{r}_i\|$. Observe that the equation is true if the inequality holds separately for the i th term in each of the sums, and that this is true since $\|\cdot\|$ is the geodesic distance.

□

Appendix B Relations between the $z_{\bar{k}}$

This section provides proofs and supporting materials to show the relations among the descriptors. The procedure to calculate the descriptors $z_{\bar{k}}$ is illustrated in Fig. B1. For a configuration \mathbf{p} whose coordinates are given in xy -coordinate system, the first step is to find the corresponding coordinates of the disks in the a_1a_2 -coordinate system, and define the function $f(\bar{a})$ using Eq. 4. Observe that the resulting function will be the same for each $n!$ permutations of the disk labels. Then, the complex coefficients of the infinite expansion in Eq. 4 are calculated using the orthogonality of complex exponentials as shown in Eq. 5. The final step is then to take the modulus of these complex coefficients and the descriptors are sorted with respect to the distance of \bar{k} to the origin. Moreover, observe that the complex coefficients $c_{\bar{k}}$

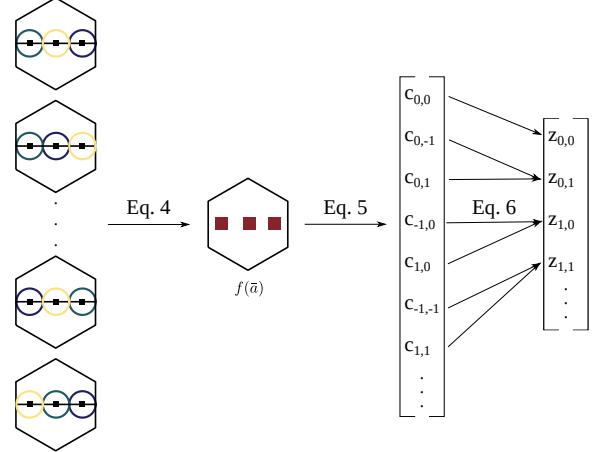


Fig. B1 $3!$ permutations of a given configuration will result in the same function $f(\bar{a})$ after using Eq. 4. The complex coefficients are calculated using Eq. 5. Finally, the descriptors $z_{\bar{k}}$ are then defined as the modulus of the complex coefficients.

and $c_{-\bar{k}}$ have the same descriptors $z_{\bar{k}}$ for a given configuration.

The complex coefficients $c_{\bar{k}}$ is equal to $c_{-\bar{k}}^*$ where $*$ denotes complex conjugate.

Proof Let \mathbf{p} be a configuration and \bar{p} be the coordinates of the disks in a_1a_2 -coordinate system.

$$\begin{aligned}c_{\bar{k}}(\mathbf{p}) &= \sum_{j=1}^n \exp(-2\pi i \bar{k} \cdot \bar{p}_j) \\ &= \left\{ \sum_{j=1}^n \exp(2\pi i \bar{k} \cdot \bar{p}_j) \right\}^* \\ &= \left\{ \sum_{j=1}^n \exp[-2\pi i(-\bar{k} \cdot \bar{p}_j)] \right\}^* \\ &= c_{-\bar{k}}^*(\mathbf{p})\end{aligned}$$

□

Let \mathbf{p} be a configuration and \bar{p} be the coordinates of the disks in a_1a_2 -coordinate system.

Proof Using the fact that $c_{\bar{k}} = c_{-\bar{k}}^*$

$$z_{\bar{k}}(\mathbf{p}) = \sqrt{c_{\bar{k}}(\mathbf{p})c_{\bar{k}}^*(\mathbf{p})} = \sqrt{c_{-\bar{k}}^*(\mathbf{p})c_{-\bar{k}}(\mathbf{p})} = z_{-\bar{k}}(\mathbf{p})$$

□

The descriptors $z_{\bar{k}}$ in Eq. 6 are invariant under inversion.

Table A1 The seven possibilities that might have zero distance for the hexagonal torus. Recall that $t_1, t_2, t_3 \in (-0.5, 0.5)$. Then, only the first one can be in the fundamental cell.

i	q_x	q_y	w_1	w_2	w_3
1	p_x	p_y	t_1	t_2	t_3
2	$p_x - 1$	p_y	$t_1 - 1$	$t_2 - 1/2$	$t_3 + 1/2$
3	$p_x + 1$	p_y	$t_1 + 1$	$t_2 + 1/2$	$t_3 - 1/2$
4	$p_x - 1/2$	$p_y - \sqrt{3}/2$	$t_1 - 1/2$	$t_2 - 1$	$t_3 - 1/2$
5	$p_x - 1/2$	$p_y + \sqrt{3}/2$	$t_1 - 1/2$	$t_2 + 1/2$	$t_3 + 1$
6	$p_x + 1/2$	$p_y - \sqrt{3}/2$	$t_1 + 1/2$	$t_2 - 1/2$	$t_3 - 1$
7	$p_x + 1/2$	$p_y + \sqrt{3}/2$	$t_1 + 1/2$	$t_2 + 1$	$t_3 + 1/2$

Proof Let \mathbf{p} and \mathbf{q} be two configurations that differ by an inversion symmetry such that $\bar{p}_i = -\bar{q}_i$ for all i . Then

$$\begin{aligned} c_{\bar{k}}(\mathbf{p}) &= \sum_{j=1}^n \exp(-2\pi i \bar{k} \cdot \bar{p}_j) \\ &= \sum_{j=1}^n \exp(2\pi i \bar{k} \cdot \bar{q}_j) \\ &= c_{\bar{k}}^*(\mathbf{q}) \\ z_{\bar{k}}(\mathbf{p}) &= \sqrt{c_{\bar{k}}(\mathbf{p})c_{\bar{k}}^*(\mathbf{p})} = \sqrt{c_{\bar{k}}^*(\mathbf{q})c_{\bar{k}}(\mathbf{q})} = z_{\bar{k}}(\mathbf{q}) \end{aligned}$$

□

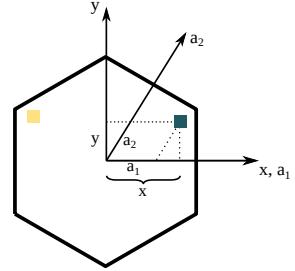
The descriptors $z_{\bar{k}}$ in Eq. 6 are invariant under translation.

Proof Let \mathbf{p} and \mathbf{q} be two configurations that differ by a translation $\bar{\Delta}$ such that $\bar{p}_i = \bar{q}_i + \bar{\Delta}$ for all i . Then

$$\begin{aligned} z_{\bar{k}}(\mathbf{p}) &= \left\| \sum_{j=1}^n \exp(-2\pi i \bar{k} \cdot \bar{p}_j) \right\| \\ &= \left\| \sum_{j=1}^n \exp[-2\pi i \bar{k} \cdot (\bar{q}_j + \bar{\Delta})] \right\| \\ &= \left\| \exp(-2\pi i \bar{k} \cdot \bar{\Delta}) \sum_{j=1}^n \exp(-2\pi i \bar{k} \cdot \bar{q}_j) \right\| \\ &= \underbrace{\left\| \exp(-2\pi i \bar{k} \cdot \bar{\Delta}) \right\|}_{1} \left\| \sum_{j=1}^n \exp(-2\pi i \bar{k} \cdot \bar{q}_j) \right\| \\ &= z_{\bar{k}}(\mathbf{q}) \end{aligned}$$

□

Finally, the descriptors $\hat{z}_{\bar{k}}$ in Eq. 7 are not all independent. Figure B2 shows two different coordinate systems (xy - and a_1a_2 -) associated with the hexagonal domain. The reason why there are two



- $\mathbf{p} = [x_1 \ y_1 \ \dots \ x_n \ y_n]$
- $\mathbf{p} = [a_{1,1} \ a_{1,2} \ \dots \ a_{n,1} \ a_{n,2}]$
- $\mathbf{p}' = [x'_1 \ y'_1 \ \dots \ x'_n \ y'_n]$
- $\mathbf{p}' = [a'_{1,1} \ a'_{1,2} \ \dots \ a'_{n,1} \ a'_{n,2}]$

Fig. B2 caption here.

different coordinate systems is that representing configurations in the xy -coordinate system is more intuitive but descriptor calculations are easier in a_1a_2 -coordinate systems. Let \bar{x} and \bar{a} respectively be the coordinates of a point in the xy -coordinate system and a_1a_2 -coordinate system. A given configuration \mathbf{p} can then be represented by different coordinates. Let \bar{T} be the forward coordinate transformation matrix between xy - and a_1a_2 -coordinate systems such that $\bar{a} = \bar{T}\bar{x}$ and $\bar{x} = \bar{T}^{-1}\bar{a}$.

Let \mathcal{L} be the set of symmetries of the tiling of the plane, and \bar{L} be one of the corresponding matrices written using the xy -coordinate system. Let \bar{x} be the position of a disk and $\bar{x}' = \bar{L}\bar{x}$ the position of the disk under the action of $L \in \mathcal{L}$. As an example, consider the yellow square in Fig. B2 that is obtained by rotating the green square by $2\pi/3$ counter clockwise. The corresponding rotation matrix \bar{L} represents this symmetry operation

using the xy representations of \mathbf{p} and \mathbf{p}' . The same operation can also be represented using the a_1a_2 representations of \mathbf{p} and \mathbf{p}' as

$$\bar{a}' = \bar{T}\bar{x}' = \bar{T}\bar{L}\bar{x} = \bar{T}\bar{L}\bar{T}^{-1}\bar{a} = \bar{U}\bar{a}$$

where $\bar{U} = \bar{T}\bar{L}\bar{T}^{-1}$. \bar{U} is equivalent to \bar{L} , but is written using the a_1a_2 -coordinate system. This makes the descriptors $z_{\bar{k}}$ of two symmetric configurations be related to each other with the relation is precisely defined by \bar{U} .

For the square torus, $\bar{T} = [1, 0; 0, 1]$ and \mathcal{L} is the dihedral symmetry group of order eight (D_4). Equation 7 is

$$\hat{z}_{\bar{k}} = \frac{1}{8} \sum_{j=1}^8 z_k^{L_j}$$

where $z_k^{L_j}$ are the descriptors $z_{\bar{k}}$ of the configuration $L_j\mathbf{x}$ with $L_j \in \mathcal{L}$. The $z_k^{L_j}$ can instead be written as

$$\begin{aligned} z_k^{L_j} &= \left\| \sum_{l=1}^n \exp(-2\pi i \bar{k} \cdot \bar{U}_j \bar{a}_l) \right\| \\ &= \left\| \sum_{l=1}^n \exp(-2\pi i \bar{k}' \cdot \bar{a}_l) \right\| \end{aligned}$$

where $\bar{k}' = \bar{U}_j^T \bar{k}$. Computing \bar{k}' for the elements in \mathcal{L} for the square torus yields

$$\begin{aligned} [p, q] &\sim [-p, q] \sim [q, -p] \sim [-q, -p] \sim \\ &[-p, -q] \sim [p, -q] \sim [-q, p] \sim [q, p]. \end{aligned}$$

For the hexagonal torus, $\bar{T} = [1, -1/\sqrt{3}; 0, 2/\sqrt{3}]$ and \mathcal{L} is the dihedral symmetry group of order twelve (D_6). Repeating the procedure above yields

$$\begin{aligned} [p, q] &\sim [q, q-p] \sim [q-p, -p] \sim [-p, -q] \sim \\ &[-q, p-q] \sim [p-q, p] \sim [-p, q-p] \sim [-q, -p] \sim \\ &[p-q, -q] \sim [p, p-q] \sim [q, p] \sim [q-p, q]. \end{aligned}$$

References

- [1] Berthier L, Ozawa M and Scalliet C 2019 *arXiv preprint arXiv:1902.07679*
- [2] Goldstein M 1969 *The Journal of Chemical Physics* **51** 37283739
- [3] Stillinger F H and Weber T A 1982 *Physical Review A* **25** 978
- [4] Stillinger F H 1995 *Science* **267** 19351939
- [5] Caiani L, Casetti L, Clementi C and Pettini M 1997 *Physical review letters* **79** 4361
- [6] Franzosi R and Pettini M 2004 *Physical review letters* **92** 060601
- [7] Franzosi R, Pettini M and Spinelli L 2007 *Nuclear Physics B* **782** 189218
- [8] Franzosi R and Pettini M 2007 *Nuclear Physics B* **782** 219240
- [9] Gori M, Franzosi R and Pettini M 2018 *Journal of Statistical Mechanics: Theory and Experiment* **2018** 093204
- [10] Kastner M and Mehta D 2011 *Physical review letters* **107** 160602
- [11] Denman H H 1965 *Journal of Mathematical Physics* **6** 16111616
- [12] Behler J and Parrinello M 2007 *Phys. Rev. Lett.* **98**(14) 146401
- [13] Kocer E, Mason J K and Erturk H 2019 *The Journal of Chemical Physics* **150** 154102
- [14] Dyre J C 2016 *Journal of Physics: Condensed Matter* **28** 323001
- [15] Alder B and Wainwright T 1962 *Physical Review* **127** 359
- [16] Mitus A, Weber H and Marx D 1997 *Physical Review E* **55** 6855
- [17] Weber H, Marx D and Binder K 1995 *Physical Review B* **51** 14636
- [18] Halperin B and Nelson D R 1978 *Physical Review Letters* **41** 121
- [19] Young A 1979 *Physical Review B* **19** 1855

- [20] Bernard E P and Krauth W 2011 *Physical review letters* **107** 155704
- [21] Engel M, Anderson J A, Glotzer S C, Isobe M, Bernard E P and Krauth W 2013 *Physical Review E* **87** 042134
- [22] Carlsson G, Gorham J, Kahle M and Mason J 2012 *Phys. Rev. E* **85**(1) 011303
- [23] Baryshnikov Y, Bubenik P and Kahle M 2014 *International Mathematics Research Notices* **2014** 25772592
- [24] Ritchey K 2019 *Computational Topology for Configuration Spaces of Disks in a Torus* Ph.D. thesis Ohio State University Electronic Thesis or Dissertation
- [25] Edelsbrunner H, Kirkpatrick D and Seidel R 1983 *IEEE Transactions on information theory* **29** 551559
- [26] Tenenbaum J B, Silva V d and Langford J C 2000 *Science* **290** 23192323
- [27] Morse M 1934 *The Calculus of Variations in the Large* vol 18 (American Mathematical Society Colloquium Publication)
- [28] Milnor J 2016 *Morse theory.* (AM-51) vol 51 (Princeton university press)
- [29] Burago D, Burago I D, Burago Y, Ivanov S and Ivanov S A 2001 *A course in metric geometry* vol 33 (American Mathematical Soc.)
- [30] Chelouah R and Siarry P 2000 *European journal of operational research* **123** 256270
- [31] Glover F 1989 *ORSA Journal on computing* **1** 190206
- [32] Krantz S and Parks H 2012 *The Implicit Function Theorem: History, Theory, and Applications* Modern Birkhäuser Classics (Springer New York)
- [33] Vietoris L 1927 *Mathematische Annalen* **97** 454472
- [34] Hatcher A 2002 *Algebraic topology* (Cambridge University Press)
- [35] Delaunay B et al. 1934 *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk* **7** 12
- [36] Edelsbrunner H and Harer J 2008 *Contemporary mathematics* **453** 257–282
- [37] Seidel R 1995 *Computational Geometry* **5** 115116
- [38] Gould H and Tobochnik J 2010 *Statistical and Thermal Physics* (Princeton University Press)
- [39] Peliti L 2011 *Statistical mechanics in a nutshell* vol 10 (Princeton University Press)
- [40] Coifman R R and Lafon S 2006 *Applied and computational harmonic analysis* **21** 530
- [41] Goresky M and MacPherson R 1988 Stratified morse theory *Stratified Morse Theory* (Springer) p 322