A geometric conjecture about phase transitions

O. B. Eriçok[*] and J. K. Mason[†]

Materials Science and Engineering, University of California, Davis, CA, 95616, USA.

As phenomena that necessarily emerge from the collective behavior of interacting particles, phase transitions continue to be difficult to predict using statistical thermodynamics. A recent proposal called the topological hypothesis suggests that the existence of a phase transition could perhaps be inferred from changes to the topology of the accessible part of the configuration space. This paper instead suggests that such a topological change is often associated with a dramatic change in the configuration space geometry, and that the geometric change is the actual driver of the phase transition. More precisely, a geometric change that brings about a discontinuity in the mixing time required for an initial probability distribution on the configuration space to reach steady-state is conjectured to be related to the onset of a phase transition in the thermodynamic limit. This conjecture is tested by evaluating the diffusion diameter and ϵ-mixing time of the configuration spaces of hard disk and hard sphere systems of increasing size. Explicit geometries are constructed for the configuration spaces of these systems, and numerical evidence suggests that a discontinuity in the ϵ-mixing time coincides with the solid-fluid phase transition in the thermodynamic limit.

PACS numbers:

I. INTRODUCTION

Phase transitions are essential to a variety of scientific and engineering applications, but continue to be difficult to predict from fundamental considerations. Instead, a phase transition is usually identified by an observed discontinuity in one of the derivatives of a thermodynamic potential. There have been several proposals concerning the origin of these discontinuities. For example, Landau theory [1, 2] associates first-order phase transitions with spontaneous symmetry breaking as quantified by an appropriately-constructed order parameter. Such order parameters often need to be defined ex post facto though, after the characteristics of the phases involved in the phase transition are already known. An understanding of phase transitions that derives from more fundamental considerations would therefore be valuable.

Statistical thermodynamics suggests that thermodynamic observables can be calculated as time averages of the relevant microscopic quantity along the system's trajectory through the phase space. The ergodic hypothesis [3, 4] implies that, provided such time averages are conducted over a period longer than a characteristic mixing time, they can be replaced by an average over the entire phase space using a measure that is independent of the system's initial microstate; that is, they can be replaced by an ensemble average. One question that has been raised about this procedure is whether the ensemble average could really be independent of the system's initial conditions in general. Consider that any system with a disconnected configuration space (i.e., a system that is not metrically transitive [5, 6]) is necessarily confined to the component of the configuration space where it be-

gins, meaning that the average behavior of an ensemble of such systems would not resemble the time-averaged behavior of any one system. Instead, the measure used for the average should depend on the system's initial microstate, since that defines the component of the configuration space to which the system is confined. Following this line of thought further, changes to the connectivity of the configuration space could discontinuously change the support of the measure used for thermodynamic averages, and thereby lead to the discontinuities in thermodynamic observables required for a phase transition. This is one way to motivate the topological hypothesis [7, 8] which roughly proposes that changes in the topology of the accessible part of the configuration space are necessary for a phase transition to occur.

The application of these ideas to a toy model is shown in Fig. 1. Suppose that there is an isolated system with the configuration space in the left panel, and that the potential energy is strictly monotone increasing with the y-coordinate in the figure. The system's internal energy then defines a value of y called the filtration value above which the system's trajectory cannot pass. Further suppose that the system's initial microstate corresponds to a point in the bottom left corner of the configuration space, and that the observable of interest is the x-coordinate in the figure. The middle panel shows the ensemble average of this observable as a function of filtration value for both the ergodic hypothesis and the topological hypothesis. Since the configuration space is symmetric about $x = 0$ and the ergodic hypothesis stipulates that the measure should be independent of the initial microstate, the ensemble average of the observable is zero for all filtration values. By comparison, the topological hypothesis recognizes that for sufficiently small filtration values the configuration space is disconnected, the system is confined to the left side of the configuration space, and the ensemble average of the observable should be zero only when the filtration value increases enough for the space to

—————

[*]Electronic address: oericok@ucdavis.edu
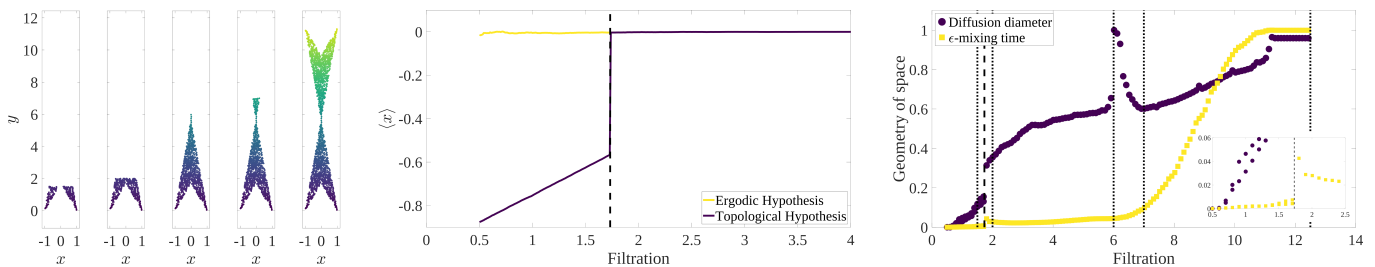[†]Electronic address: jkmason@ucdavis.edu

FIG. 1: The left panel shows the evolution of a model configuration space with points colored (shaded) by their y-coordinate. The filtration value is an upper bound for the y-coordinate of any point on the system's trajectory. The middle panel shows the average x-coordinate $\langle x \rangle$ calculated using the ergodic hypothesis and the topological hypothesis as a function of filtration value, assuming that the system starts at the lower left corner of the configuration space. The right panel shows the diffusion diameter and ϵ-mixing time of the configuration space as functions of filtration value and that they are sensitive to changes in the configuration space geometry. The diffusion diameter, but not the ϵ-mixing time, is also sensitive to the bottleneck. The five dotted lines show the filtration values used in the left panel.

become connected. Moreover, the discontinuous change in the integration measure at this filtration value leads to a discontinuous observable, precisely of the type that would be expected for a system undergoing a phase transition. It is significant that it was not necessary at any point to pass to the thermodynamic limit for there to be a discontinuity in the observable, though this line of reasoning does require that the limiting distribution of microstates be described by the microcanonical ensemble. Connecting the system to a heat bath formally introduces a nonzero probability of arbitrarily large thermal fluctuations, makes the entire configuration space accessible, and only allows discontinuous changes in observables to occur in the thermodynamic limit. That said, since experimental systems are necessarily finite, there is utility in further exploring the indications and predictors of nascent phase transitions in finite systems.

Let E be the potential energy of a system, $\boldsymbol{q}_i$ be the coordinates of the system's particles, and $V(\boldsymbol{q}_1 \ldots \boldsymbol{q}_N)$ be a smooth, stable, confining and short-range interaction potential. The topological hypothesis [9, 10] initially claimed that a topological change in the equipotential energy submanifolds $\Sigma_E = V^{-1}\left((-\infty, E]\right)$ of the configuration space was a necessary condition for a phase transition. Kastner and Mehta [11] observed a second-order phase transition in a two-dimensional Φ^4 system at an energy where no topological change occurred, disproving this claim. Gori et al. [12] subsequently refined the hypothesis and observed that the phase transition in the two-dimensional Φ^4 system was caused by a diverging transition time between two parts of the configuration space, and that this was associated with an asymptotic topological change. That is, a continuous change in the configuration space geometry brought about a divergence in the mixing time, and while such events are often accompanied by topological changes they are not always. The significance of the configuration space geometry is supported by a recent study of the topological and geometric properties of the two-dimensional XY model by Bel-Hadj-Aissa et al. [13]; they computed the mean ge-

ometric curvature of the equipotential energy level sets and observed that the location of the phase transition could be inferred from the level set curvatures.

With this as background, the fundamental conjecture of this work is:

Conjecture 1. *A necessary condition for a first-order phase transition is a discontinuity in the mixing time on the configuration space.*

This is not intended to suggest that the mixing time directly regulates the appearance of phase transitions, but merely that the mixing time is sensitive to any geometric changes that could discontinuously change the integration measure for thermodynamic averages.

A natural question at this point is whether it is actually possible to measure the mixing time of a thermodynamic system by means of the convergence of an initial probability distribution on the configuration space to the limiting distribution. We hypothesize that the relevant geometric changes are so severe that any geometric quantity reasonably sensitive to the accessible volume of the configuration space and the length and number of paths connecting distant regions should exhibit measurable discontinuities for the same values of the control variable as the mixing time. The two quantities used here are the diffusion distance [14, 15] and the ϵ-mixing time. The first measures the difference between two distributions that start as Dirac delta distributions at different locations on the configuration space and evolve by diffusion; maximizing the diffusion distance over all starting locations of the two distributions gives the configuration space's diffusion diameter. The second measures the time required for a Dirac delta distribution to diffuse to the steady-state distribution within a tolerance defined by ϵ, averaged over all starting locations of the distribution in the configuration space. This is intended to resemble the thermodynamic mixing time, though the distribution evolves by the diffusion equation rather than the Liouville equation and the similarity to the steady-state distribution is quantified by the Kullback–Leibler divergence [16].

thor Of course, there have been considerable prior efforts to predict the onset of phase transitions that do not involve the topological hypothesis or considerations along the lines of Conj. 1. The Yang–Lee theory stands out as a particularly successful approach for predicting the locations and orders of equilibrium phase transitions for a variety of model systems [17–19]. Suppose that the partition function of a system can be written in the form of a polynomial with real positive coefficients; the grand canonical partition function can be written as such a polynomial in the fugacity $z = \exp[\mu/(k_B T)]$ where μ is the chemical potential. Provided that the degree M of the polynomial is finite (a generic property of finite systems), this polynomial necessarily has M roots that occur as complex-conjugat pairs on the complex-fugacity plane. A non-analyticity in the free energy as is required for there to be a phase transition can only occur at one of these roots of the partition function. The Yang–Lee theory therefore considers the distribution of these roots, and particularly their proximity to the real positive semi-axis, as a function of system size to infer the existence and properties of equilibrium phase transitions. While this is an elegant approach, not all systems of interest allow a partition function with the necessary form, and even when they do solving for the distribution of complex roots can be an extremely difficult task.

The conjecture made above is tested here by evaluating the diffusion diameter and ϵ-mixing time for the configuration spaces of hard disk and hard sphere systems, collectively called hard disk systems in the following. These systems are frequently used to model simple fluids [20], and are governed by the hard disk potential for which the energy is infinite if any disks overlap and is zero otherwise. The phase transitions for these systems have been studied extensively, starting with the seminal work of Alder and Wainwright [21] who observed a phase transition in a system of hard disks as a function of packing fraction η (the fraction of the area covered by the disks). The solid for $\eta > 0.72$ is characterized by the presence of a long-range translational and orientational order, whereas the fluid for $\eta < 0.70$ is characterized by the absence of any long-range order [22, 23]. The interval $0.70 < \eta < 0.72$ was initially believed to contain coexisting solid and fluid phases as would be expected of a first-order transition, but more recent evidence [24–26] suggests that this interval contains a hexatic phase [27, 28]. It is significant that the phases and phase transitions that have been observed are surprisingly complex and varied, even for this simple system. Similar studies of hard spheres in three dimensions include those by Isobe and Krauth [29] and Pieprzyk et al. [30] who observed solid-fluid phase coexistence in the $0.49 < \eta < 0.548$ interval.

The configuration spaces of hard disk systems have been studied before. Carlsson et al. [31] explored the topology of the configuration space of five hard disks in a unit square box by regularizing the potential energy surface (explained in Sec. IIB) and using classical

Morse theory [32, 33] to calculate the Betti numbers of the configuration space, roughly indicating the number of holes of various dimensions. Baryshnikov et al. [34] developed a min-type Morse theory for the configuration spaces of hard disk systems and proved that the critical points where the topology changes as a function of packing fraction are precisely the mechanically-balanced disk configurations (explained in Sec. IIB). Ritchey [35] studied configuration spaces of hard disks in the hexagonal torus (shown in the left of Fig. 2) for $n = 1 \ldots 12$ disks, created a database of the critical points, defined their critical index (specifying the nature of the associated topological change), and found their plane symmetry groups. The authors previously [36] proposed distance functions on the configuration spaces of hard disks quotiented by various symmetry groups, where quotienting means that configurations related by a symmetry operation are considered the same. They subsequently [37] used these distances to explicitly triangulate the quotients of the configuration space of two spheres in the rhombic dodecahedron (shown in the right of Fig. 2) and measure the diffusion diameters of the resulting spaces. They also observed an accumulation of critical points in the configuration space at packing fractions in the phase coexistence interval, suggesting the possibility of a topological and geometric catastrophe there in the thermodynamic limit.

The technical contribution of this work is a set of techniques described in Sec. II that collectively allow the configuration space geometries for hard disk systems with $n \leq 7$ disks and hard sphere systems with $n \leq 6$ spheres to be evaluated in practice. The diffusion diameters and ϵ-mixing times (defined in Sec. III) of these spaces are computed as functions of particle diameter and particle number, and in Sec. IV significant discontinuities are observed at the packing fractions of critical points close to the phase coexistence intervals. Along with the observation [37] that critical points accumulate in the phase coexistence intervals with increasing particle number, this suggests that there should be a discontinuity in the mixing time at the packing fraction of the phase transition in the thermodynamic limit, offering preliminary support for Conj. 1.

II. CONFIGURATION SPACES

A. Tautological function

The configuration space of n points on a d-dimensional torus T^d is

$$\Lambda(n) = \{\mathbf{x} = (\boldsymbol{x}_1 \ldots \boldsymbol{x}_n) \,|\, \boldsymbol{x}_i \in T^d\}. \qquad (1)$$

A two-dimensional torus T^2 for hard disks is obtained by identifying opposite edges of a regular hexagon, whereas a three-dimensional torus T^3 for hard spheres is obtained by identifying opposite faces of a rhombic dodecahedron. Figure 2 shows these domains with some of the periodic
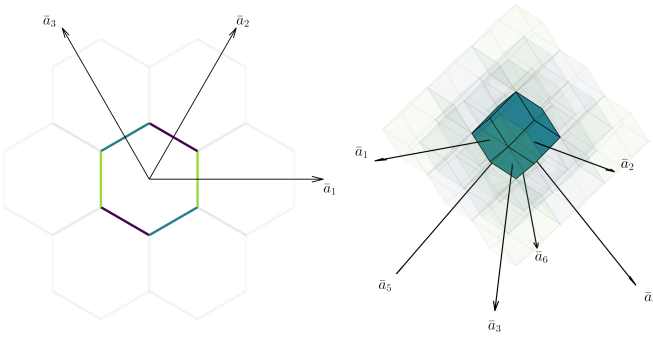
FIG. 2: A two-dimensional torus T^2 (left) and a three-dimensional torus T^3 (right) are shown with some of their periodic images. The unit vector $\boldsymbol{a}_1$ points in the x-direction in both cases. All the unit vectors pass through the centers of the corresponding faces.

images of the fundamental unit cells. The center to center distance to the neighboring cells is one in both cases.

The tautological function $\tau : \Lambda(n) \to R$ is the maximum radius that the disks could have without overlapping, or

$$\tau(\mathbf{x}) = \min_{1 \leq i < j \leq n} r_{ij} \qquad (2)$$

where r_{ij} is half the geodesic distance between the centers of disks i and j. This allows the configuration space of n disks of radius ρ to be written as

$$\Gamma(n, \rho) = \tau^{-1}\left([\rho, \infty)\right) \qquad (3)$$

or the set of all configurations of points where the minimum point separation is at least ρ.

B. Critical configurations

Morse theory [32, 33] stipulates that the topology of the sublevel sets of a generic real-valued function f defined on a smooth manifold M can change only at the critical points of f. Intuitively, a critical point is a point where the gradient of the function vanishes. The index of a critical point is defined as the number of negative eigenvalues of the Hessian matrix there, and roughly indicates the number of independent directions along which f decreases to second order. Let $M_a = \{x \in M | f(x) < a\}$ be a sublevel set of M. One of the results of Morse theory is that the topology of M_a and M_b are equivalent if the interval $[a, b]$ doesn't contain any critical points of f.

Classical Morse theory does not apply to the tautological function in Eq. 2 though since it is a min-type function that is not differentiable everywhere. For this reason, Baryshnikov et al. [34] began to develop a min-type Morse theory specifically for hard disk configuration spaces, and proved that the critical points of the tautological function are precisely the mechanically-balanced configurations. Consider the critical points of the configuration spaces of four hard disks in Fig. 3 where they are
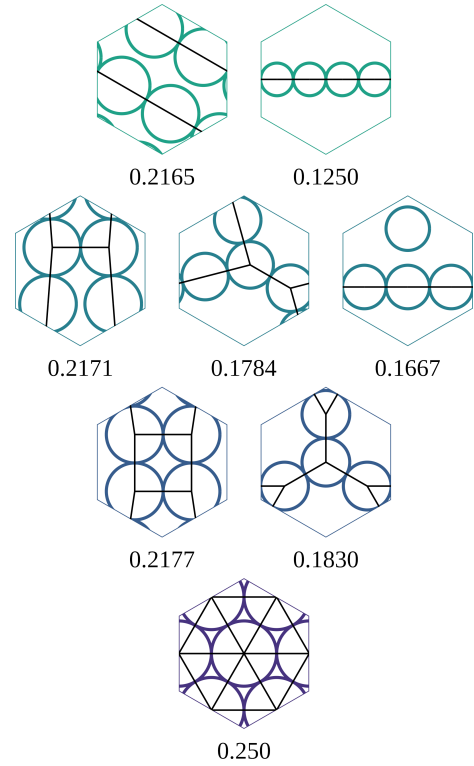

FIG. 3: The equivalence classes of the critical points of the configuration space of four hard disks. The tautological function values are reported below the figures. The critical points are ordered from index-0 (bottom) to index-3 (top).

ordered from index-0 (bottom) to index-3 (top). An edge exists between any pair of disks if the distance between the disk centers is exactly twice the value of tautological function for that configuration (reported below the figure). Suppose that the edge between disks i and j exerts a force on those disks with magnitude f_{ij} directed along the edge. A configuration is mechanically-balanced if there is a way to apply forces along the edges such that at least one force is nonzero and the net force on each disk vanishes. The result is then that such configurations are precisely those associated with changes to the topology of the superlevel sets of the tautological function.

The critical points of the configuration spaces of hard disks can also be shown to coincide with the critical points of a regularized hard disk potential energy function in the following way. Observe from the definition of the tautological function that the radius ρ of the disks in a given configuration could be increased by as much as

$$\tau - \rho = \min_{1 \leq i < j \leq n}(r_{ij} - \rho)$$
$$\approx -\frac{1}{w}\log\left\{\sum_{1 \leq i < j \leq n} \exp\left[-w(r_{ij} - \rho)\right]\right\}$$

without the disks overlapping, where the second line uses a soft-min approximation. The argument of the logarithm is interpreted as a regularized hard-disk potential

$E = \sum_{i<j} \exp\left[-w(r_{ij} - \rho)\right]$ that converges to the true hard disk potential in the $w \to \infty$ limit. Since the equation above relates τ to E by strictly monotonic transformations and these preserve the locations of any critical points, the critical points of E should converge to the critical points of τ in the limit. This allows the critical points of τ to be found by applying standard numerical minimization procedures $\|\nabla E\|^2$ since the minima of this function are by definition the critical points of E. Interactive databases of the critical points found in this way for the hard disk[1] and hard sphere[2] configuration spaces are available online. The distributions of critical points as functions of index and packing fraction in Refs. [35, 37] suggest that an accumulation of critical points in a narrow packing fraction interval with increasing number of disks could be associated with the onset of a phase transition; one of the purposes of this work is to explore this observation further.

C. Geometric representation

The procedure followed previously [36, 37] when studying hard disk configuration spaces involved repeatedly sampling points in the configuration space of points $\Lambda(n)$, connecting nearby points in the resulting point cloud to reconstruct $\Lambda(n)$, and restricting to the hard disk configuration space $\Gamma(n, \rho)$ by retaining only those regions with suitable values of the tautological function. The difficulty with this procedure is that the dimension and complexity of $\Gamma(n, \rho)$ increases rapidly with n, quickly requiring a point cloud with an overwhelming number of points to accurately capture the geometric details. This section describes several techniques that, while they do not solve the problem, reduce the required number of points enough to allow us to study the geometry of configuration spaces with up to 18 dimensions.

The first is to quotient the configuration space by various symmetry groups, thereby reducing the relevant volume. The three isometry groups considered here are the group of rigid translations $\mathcal{T}$, the group of disk label permutations $\mathcal{P}$, and the point group symmetries of the fundamental unit cell $\mathcal{L}$. As described previously [36, 37], the configuration space is quotiented by these groups by changing the distance function used to identify and connect nearby points in the point cloud. More precisely, there is a natural distance function $d_\Lambda(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{y}_i\|$ on Λ that is the sum of the geodesic distances each point in a configuration $\mathbf{x}$ would need to travel to be converted into configuration $\mathbf{y}$. Given an isometry group $\mathcal{S}$, the distance $d_{\Lambda/\mathcal{S}}$ in the quotient space

---

[1] https://github.com/burakericok/hard_disk_crits
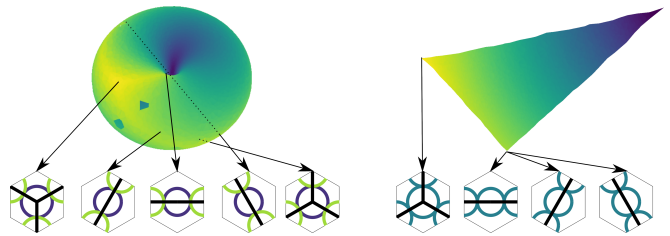[2] https://github.com/burakericok/hard_sphere_crits

FIG. 4: The configuration space quotient by rigid translations $\Lambda(2)/\mathcal{T}$ (left) and the configuration space quotient by rigid translations, disk label permutations, and point group symmetries $\Lambda(2)/\{\mathcal{T}, \mathcal{P}, \mathcal{L}\}$ (right) for two hard disks. Color (shading) represents the tautological function values and arrows indicate the locations of the critical points.

$\Lambda/\mathcal{S}$ can be written as [38]

$$d_{\Lambda/\mathcal{S}}(\mathbf{x}, \mathbf{y}) = \inf_{S \in \mathcal{S}} d_\Lambda[\mathbf{x}, S(\mathbf{y})]. \tag{4}$$

Observe that $\mathcal{T}$ is a continuous group, that Eq. 4 involves solving a global optimization problem when $\mathcal{T} \subset \mathcal{S}$, and that this problem can have multiple local minima. While algorithms like Tabu search [39, 40] can be used in such situations, the number of problems to be solved increases rapidly with the number of isometry groups considered. For example, when $\mathcal{S} = \{\mathcal{T}, \mathcal{P}, \mathcal{L}\}$, the global optimization over rigid translations needs to be solved $n! \times O(\mathcal{L})$ times to evaluate Eq. 4 where $O(\mathcal{L})$ is the order of $\mathcal{L}$. Figure 4 shows the configuration spaces of two hard disks quotiented by rigid translations $\Lambda(2)/\mathcal{T}$ (left) and by rigid translations, permutations of disk labels and point group symmetries of the hexagon $\Lambda(2)/\{\mathcal{T}, \mathcal{P}, \mathcal{L}\}$ (right) constructed using the distance in Eq. 4 with the procedure of Ref. [36]. The color represents the tautological function values at each point. These configuration spaces have only two critical point types [36], local minima (index-0) where each disk has three connections and saddle points (index-1) where each disk has two connections; these critical points are indicated with the corresponding disk configurations. Observe that the topologies and geometries of these spaces are substantially different and depend on the choice of symmetries by which to quotient. This work mainly considers the quotient spaces $\Lambda(n)/\mathcal{T}$ and $\Lambda(n)/\{\mathcal{T}, \mathcal{P}, \mathcal{L}\}$ for increasing number of disks.

The second technique reduces the number of optimization problems that need to be solved in Eq. 4. The idea is to construct an approximation to the solution of the global optimization problem over rigid translations, allowing the discrete symmetry operations that are unlikely to realize the minimum to be quickly rejected. The procedure followed here samples a fixed number of random translations for a fixed discrete symmetry operation, and uses the minimum over the random translations as the approximation. The details of this procedure and various numerical results are provided in App. A. This reduces the computational cost enough to be able to evaluate a small number of distances $d_{\Lambda/\mathcal{S}}$, but not all the pair-

wise distances in the point cloud when identifying nearby points to reconstruct $\Lambda/\mathcal{S}$.

The third technique reduces the number of times that $d_{\Lambda/\mathcal{S}}$ needs to be evaluated. Points sampled from Λ are initially mapped to a Cartesian space with coordinates given by descriptors that are invariant to the action of $\mathcal{T}$, $\mathcal{P}$, and $\mathcal{L}$ [36, 37]. This mapping is conjectured to be an embedding and therefore to preserve the neighborhood of every point in Λ, but is not isometric in the sense that distances between configurations are distorted. That said, there are efficient algorithms to calculate k'-nearest neighbor graphs in the descriptor space with the Euclidean metric. Since neighborhoods are preserved, the k'-nearest neighbor graph of a point in the descriptor space should contain the k-nearest neighbor graph of the point in $\Lambda/\mathcal{S}$ for sufficiently large $k' > k$; our numerical experiments suggest that $k' \sim 5k$ is usually sufficient. The procedure therefore involves constructing the k'-nearest neighbor graph in the descriptor space, evaluating $d_{\Lambda/\mathcal{S}}$ for each edge of this graph, and constructing the approximate k-nearest neighbor graph in $\Lambda/\mathcal{S}$ using these distances.

It is useful at this point to discuss the relationship of the k-nearest neighbor graph in $\Lambda/\mathcal{S}$ to the geometry of the underlying space. It is clear that the apparent connectivity of the space depends on the value of k; for very small k the graph would likely contain many disconnected components, whereas for very large k every point would share an edge with every other point, regardless of the actual properties of the space. While there does not seem to be an established canonical approach to selecting the value of k, the strategy followed here involves the use of topological information. Observe that, for any filtration of the k-nearest neighbor graph by the tautological function, the number of disconnected components in the resulting graph should be at most the number of index-0 critical points with disk radius larger than or equal to the filtration value. Moreover, when the filtration value is equal to the smallest value such that all of the index-1 critical points (which correspond to saddle points) are included, only a single connected component should remain. The smallest value of k for which these conditions are satisfied seems a reasonable choice, and for the numbers of points in our point clouds (indicated in Fig. 5) is approximately $k = 100 \times n$ where n is the number of disks.

The fourth technique is to directly use the k-nearest neighbor graph rather than a simplicial complex (a triangulation) to evaluate the geometric properties of $\Lambda/\mathcal{S}$. Previously, the configuration spaces for small numbers of hard disks and spheres were triangulated as α-complexes [36, 37]. While such simplicial complexes can accurately represent all the geometric properties of the underlying space, the number of simplices required generally increases exponentially with dimension, quickly making the computational memory requirements prohibitive [41, 42]. Fortunately, the only geometric information required to calculate the diffusion diameter and the ϵ-mixing time are
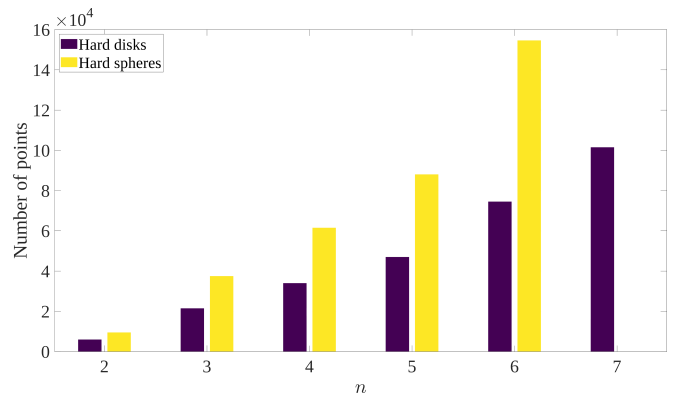

FIG. 5: Number of points used in the graph representations for hard disks and spheres.

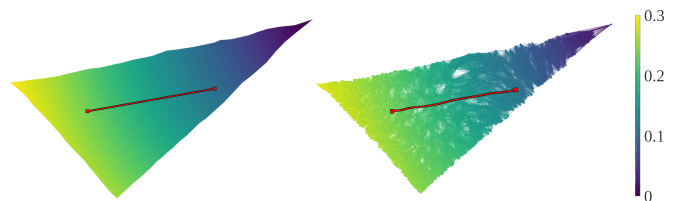
FIG. 6: The quotient space $\Lambda(2)/\{\mathcal{T}, \mathcal{P}, \mathcal{L}\}$ of two points on T^2 represented as an α-complex (left) and a graph (right), with color (shading) indicating the value of the tautological function. The red (outlined) path connecting two points is a straight line on the left, but is slightly longer on the right.

geodesic distances, and these can be reasonably approximated from the k-nearest neighbor graph alone. Figure 6 shows the quotient space $\Lambda(2)/\{\mathcal{T}, \mathcal{P}, \mathcal{L}\}$ of two points on T^2 represented as an α-complex (left) and as a graph (right). While the shortest path connecting two points is a straight line on the left, the path through the edges of the graph on the right is slightly longer. Since both the diffusion distance and the ϵ-mixing time are designed to be robust to small geometric perturbations like these, using the k-nearest neighbor graph of $\Lambda/\mathcal{S}$ directly is sufficient for our purposes.

The fifth and final technique is related to the construction of the point cloud on Λ, and is intended to reduce the number of points required to accurately represent the relevant geometric features of $\Lambda/\mathcal{S}$. Figure 7 shows the distributions of tautological function values ρ for point clouds on $\Lambda(4)$ sampled by three different procedures, with the dashed lines indicating the radii of previously-identified critical points. The left panel shows the distribution of points sampled uniformly on $\Lambda(4)$, and reveals that the overwhelming majority of the space's volume has comparatively small values of ρ where only a single pair of disks would be in contact. The critical points are concentrated at relatively high values of ρ though. This motivates the use of importance sampling [43] to sample points more uniformly over ρ, concentrating points in the regions where the topology and geometry of $\Gamma(4, \rho)$ are most likely to change and giving the distribution in the

middle panel. The sampling density around the critical points can be increased further by perturbing known critical configurations and adding the resulting configurations directly to the point cloud, giving the distribution in the right panel. The intention is to ensure that the density of sampled points is high enough to accurately reflect the parts of Λ that are most relevant to the hypothesis without unnecessarily sampling points elsewhere in the space.

III. CONFIGURATION SPACE GEOMETRY

Diffusion is a smoothing process. Flows from regions of high concentration to low concentration necessarily make a concentration field more uniform with time. This also makes diffusion processes relatively insensitive to small perturbations in the initial concentration field or to the geometry of the underlying space. This property is likely one of the reasons that diffusive processes are often used to learn about the geometry of a space represented by a sampled point cloud, since the resulting insights do not depend sensitively on the distribution of the points.

A. Diffusion Distance

Let G be the k-nearest neighbor graph on $\Lambda/\mathcal{S}$ constructed by the procedure in Sec. II C. The diffusion distance $d_{ij,t}$ on G measures the L^2 distance between two distributions that begin as Dirac delta distributions on vertices i and j and diffuse on G for a time t. Let the kernel matrix $\mathbf{K}$ have entries $k_{ij} = \exp(-l_{ij}^2/\sigma^2)$ where l_{ij} is Dijkstra's distance between vertices i and j and σ is twice the median length of edges in G, and the diagonal degree matrix $\mathbf{D}$ has entries $d_{ii} = \sum_j k_{ij}$. Then $\mathbf{P} = \mathbf{D}^{-1}\mathbf{K}$ is the transition rate matrix of a continuous-time Markov process where the effect of raising $\mathbf{P}$ to the tth power is equivalent to propagating the process by a time t. Let $\{\lambda_l, \phi_l\}$ be the set of eigenvalues and eigenvectors of $\mathbf{P}$ for $0 \leq l \leq N-1$ where N is the number of vertices. Since $\mathbf{P}$ is normalized, the largest eigenvalue λ_0 is associated with a constant eigenvector ϕ_0. Discarding λ_0 and ϕ_0, the diffusion coordinates are defined as

$$\mathbf{\Phi}_{t,i} = \left[\lambda_1^t \phi_1(i), \ldots, \lambda_{N-1}^t \phi_{N-1}(i)\right] \tag{5}$$

and encode a distribution starting as a Dirac delta distribution on vertex i and diffusing for a time t. The diffusion distance between vertices i and j is then defined as

$$d_{ij,t} = \|\mathbf{\Phi}_{t,i} - \mathbf{\Phi}_{t,j}\| \tag{6}$$

where $\|\cdot\|$ is the Euclidean norm. Note that the significance of the time t depends on the diffusion rate implicit in the kernel $\mathbf{K}$. Since the eigenspectrum of $\mathbf{P}^t$ decays very quickly for sufficiently large t, the diffusion distance $d_{ij,t}$ can often be accurately approximated using only the first few eigenvalues and eigenvectors for a

substantial time and memory savings [15]. Details about this approximation are provided in App. B.

B. Mixing Time

Thermodynamics is generally concerned with equilibrium or quasiequilibrium systems, where there is no net redistribution of matter or energy and thermodynamic observables are time independent. From the standpoint of statistical thermodynamics, this means that the time average of any thermodynamic observable over the system's trajectory through the phase space does not depend on the initial microstate provided that the averaging is performed over a sufficiently long time interval; the shortest such interval is known as the mixing time.

Let Ω be the phase space of a thermodynamic system, $\boldsymbol{q}$ and $\boldsymbol{p}$ be the canonical coordinates and momenta, and $\mu(\boldsymbol{q}, \boldsymbol{p}, t)$ be the probability distribution of microstates on Ω at time t. Suppose that the system has an arbitrary initial condition $\mu_0(\boldsymbol{q}, \boldsymbol{p}) = \mu(\boldsymbol{q}, \boldsymbol{p}, 0)$. Briefly setting aside the specifics of the time evolution of $\mu(\boldsymbol{q}, \boldsymbol{p}, t)$, the steady-state distribution $\mu_\infty(\boldsymbol{q}, \boldsymbol{p}) = \lim_{t\to\infty} \mu(\boldsymbol{q}, \boldsymbol{p}, t)$ defines the equilibrium condition on Ω, and the mixing time could in principle be defined by the approach of $\mu(\boldsymbol{q}, \boldsymbol{p}, t)$ to $\mu_\infty(\boldsymbol{q}, \boldsymbol{p})$. That said, the literature does not seem to precisely define the necessary conditions for the two distributions to be effectively indistinguishable, nor the effect of the initial condition $\mu_0(\boldsymbol{q}, \boldsymbol{p})$ on the resulting mixing time. A likely reason for this is the historical emphasis on equilibrium states for which, by hypothesis, the observation time can be made as long as necessary for there to be complete mixing.

Classically, the time evolution of $\mu(\boldsymbol{q}, \boldsymbol{p}, t)$ is governed by the Liouville equation [44, 45]. Liouville's theorem states that the resulting convective derivative of the probability distribution is zero, or that the flow of probability density resembles that of an incompressible fluid. This raises questions relating to the convergence of probability distributions that are closely related to those about the origins of irreversibility, and that continue to be discussed in the literature. This discussion is avoided by simply supposing that $\mu(\boldsymbol{q}, \boldsymbol{p}, t)$ evolves by the diffusion equation; there is evidence in the literature that the relaxation behavior of similar systems is independent of the choice of microscopic dynamics up to an overall multiplicative constant [46–48], meaning that the use of diffusive dynamics likely does not substantially change the relative values of the ϵ-mixing time as defined below as a function of packing fraction. One consequence of this supposition is that the marginal probability distributions on the configuration and momentum subspaces evolve independently, though only the marginal probability distribution on the configuration space has a limiting distribution (the momentum subspace is unbounded). This suggests that the mixing time be defined by the convergence of the marginal probability distribution $\nu(\boldsymbol{q}, t)$ on the configuration space Ω_q to the limiting distribution
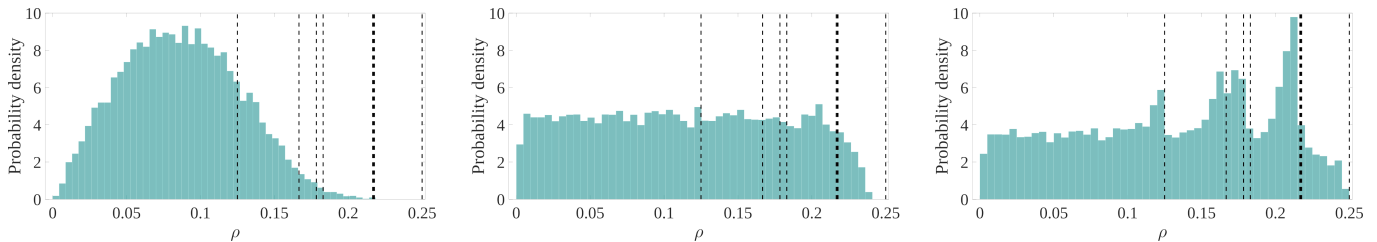
FIG. 7: Distributions of tautological function values for point clouds on $\Lambda(4)$ sampled by three different procedures, with the dashed lines indicating the radii of previously-identified critical points. Sampling points uniformly gives the distribution on the left. The distribution in the middle uses importance sampling and is more uniform. The distribution on the right is for a point cloud that both uses importance sampling and samples additional points in the neighborhoods of the critical points.

$\nu_\infty(\boldsymbol{q})$.

The Kullback-Leibler divergence [16] is often called the relative entropy, and is a standard way to quantify how much a probability distribution ν differs from a reference probability distribution ν_∞:

$$I(t; \nu_0) = \int_\Omega \nu(\boldsymbol{q}, t) \log \frac{\nu(\boldsymbol{q}, t)}{\nu_\infty(\boldsymbol{q})} \mathrm{d}\boldsymbol{q}. \qquad (7)$$

We propose that the condition for $\nu(\boldsymbol{q}, t)$ to have converged to $\nu_\infty(\boldsymbol{q})$ be that $I(t; \nu_0) \leq \epsilon$, where ϵ is an adjustable parameter analogous to a conventional convergence threshold.

Now suppose that ν_0 is a Dirac delta distribution centered at $\boldsymbol{q}_0$ (the initial configuration is precisely known), and that $t_\epsilon(\boldsymbol{q}_0)$ is the minimum time required for this initial condition to converge to the limiting distribution $\nu_\infty(\boldsymbol{q})$ on Ω_q in the sense above. The ϵ-mixing time $\langle t \rangle_\epsilon$ is defined as the weighted average of t_ϵ over all possible choices of the initial condition, or

$$\langle t \rangle_\epsilon = \int_\Omega t_\epsilon(\boldsymbol{q}) \nu_\infty(\boldsymbol{q}) d\boldsymbol{q} \qquad (8)$$

where the measure of integration is taken to be $\nu_\infty(\boldsymbol{q})$ in the absence of a natural alternative; the purpose of averaging over all possible choices of the initial condition is to make $t_\epsilon(\boldsymbol{q}_0)$ independent of this choice. The ϵ-mixing time is employed as a precisely-defined proxy for the thermodynamic mixing time on the configuration space in the following. While the reader is cautioned that the ϵ-mixing time and the thermodynamic mixing time should differ by at least an overall multiplicative constant [46–48], this should not affect the detection of packing fractions where either mixing time changes rapidly.

This leaves only the definition of a diffusive process occurring on a connected graph G instead of on a continuous space Ω_q. Let $\boldsymbol{\nu}(t)$ be the probability masses on the vertices of G at time t, and define the graph Laplacian as $\mathbf{L} = \mathbf{D} - \mathbf{K}$ [49]. The governing equation for a diffusive process is

$$(\partial/\partial t + \mathbf{L})\boldsymbol{\nu}(t) = 0. \qquad (9)$$

Let $\boldsymbol{\Lambda}$ be the diagonal matrix of the eigenvalues $\lambda_0 = 0 < \lambda_1 \cdots \leq \lambda_{N-1}$ of $\mathbf{L}$, and $\boldsymbol{\Phi}$ be the matrix of the cor-

responding eigenvectors where the first column is a constant vector. Since $\mathbf{L}$ is a symmetric matrix, the columns of $\boldsymbol{\Phi}$ form an orthogonal basis and the solution to the diffusion equation is

$$\boldsymbol{\nu}(t) = \boldsymbol{\Phi} e^{-t\boldsymbol{\Lambda}} \boldsymbol{\Phi}^T \boldsymbol{\nu}_0 \qquad (10)$$

for the initial condition $\boldsymbol{\nu}_0$. The steady-state distribution is readily calculated using the fact that all of the diagonal elements of $e^{-t\boldsymbol{\Lambda}}$ go to zero in the limit of long time except the first term which goes to one. This suggests that the steady-state distribution ν_∞ is always the uniform distribution over all vertices of G.

The right panel of Fig. 1 shows the different behaviors of the diffusion distance and the mixing time in practice. Observe that while both exhibit a discontinuous jump at the filtration value where the two components of the space merge, the diffusion diameter has an additional peak at the filtration value of the bottleneck. The initial rise is attributed to the probability mass having difficulty diffusing to the point of the bottleneck through a small number of paths. Since the space is constructed by means of k-nearest neighbor graphs, more paths spanning the neck appear with further increases in the filtration value, subsequently reducing the diffusion diameter. That is, the peak in the diffusion diameter is effectively an artifact of the construction of the space and the precise sequence in which the edges appear. By comparison, the average mixing time $\langle t \rangle_\epsilon$ is unaffected by the bottleneck since the small volume of that region reduces its contribution to the average in Eq. 8. More details about the calculation of the ϵ-mixing time for this system are given in App. C.

IV. RESULTS

The hard disk equation of state usually appears in the literature as a function of packing fraction η [50, 51]. The radius ρ of the disks appearing in Eq. 3 can be converted to a packing fraction using $\eta = n\pi\rho^2/A$ for hard disks and $\eta = 4n\pi\rho^3/(3V)$ for hard spheres, where $A = \sqrt{3}/2$ and $V = \sqrt{2}/2$ are respectively the area of the fundamental hexagon and the volume of the fundamental rhombic dodecahedron in Fig. 2.

The authors previously constructed quotient spaces of two hard disks [36] and two hard spheres [37] and studied the topological and geometric properties of both the original and quotient spaces as functions of packing fraction. They observed that while the geometric and topological features of these spaces changed dramatically with the choice of symmetry groups by which to quotient, the general behavior of the diffusion diameter (and significantly the locations of the discontinuities) as a function of packing fraction did not. This suggests that the quotient space has the advantage of a substantially reduced volume while retaining the essential geometric and topological features of the full configuration space. If the behavior of the ϵ-mixing time is similarly robust to the choice of symmetry groups by which to quotient, this would allow Conj. 1 to be tested more easily and for systems with larger numbers of disks.

Figure 8 shows the ϵ-mixing times of the quotient spaces $\Gamma(2,\eta)/\mathcal{T}$ and $\Gamma(2,\eta)/\{\mathcal{T},\mathcal{P},\mathcal{L}\}$ for hard disks (top) and hard spheres (bottom). Since only relative changes in the ϵ-mixing time are significant, all the results in this section are normalized to their maximum value for ease of comparison. Initially consider the hard disk results in the top row. There are only two critical points, one index-0 and one index-1. The volumes and ϵ-mixing times of both spaces initially grow slowly with decreasing η, with a discontinuity appearing at precisely the packing fraction of the index-1 critical point; the discontinuity is stronger for $\Gamma(2,\eta)/\mathcal{T}$ since the geometric change is more severe, though at the price of greatly increased computational cost. The bottom row shows the results for hard spheres, for which there are two distinct index-0, one index-1, and one index-2 critical points. Since having distinct index-0 critical points results in the space having multiple disconnected components for certain intervals of η, the ϵ-mixing times for each component are shown with an opacity that indicates the fraction of vertices participating in that component. As before, there is a discontinuity in the ϵ-mixing time at the packing fraction of the index-1 critical point, in this situation indicating that the disconnected components are joined. Notice that a similar discontinuity does not occur at the packing fraction of the index-2 critical point, nor indeed at any other critical point of the tautological function. Since the behavior of the diffusion diameter was previously found to be similarly robust to the choice of symmetry groups by which to quotient, the quotient spaces $\Gamma/\{\mathcal{T},\mathcal{P},\mathcal{L}\}$ will be used exclusively in the following.

Figure 9 shows the diffusion diameters and ϵ-mixing times of the quotient spaces $\Gamma/\{\mathcal{T},\mathcal{P},\mathcal{L}\}$ for $n = 3\ldots7$ hard disks as functions of packing fraction, with dashed lines indicating the packing fractions of the critical points. The first observation is that not all of the critical points correspond to substantial geometric changes, at least not ones to which the diffusion distance and ϵ-mixing time are sensitive. The second is that the diffusion diameter is generally much noisier than the ϵ-mixing
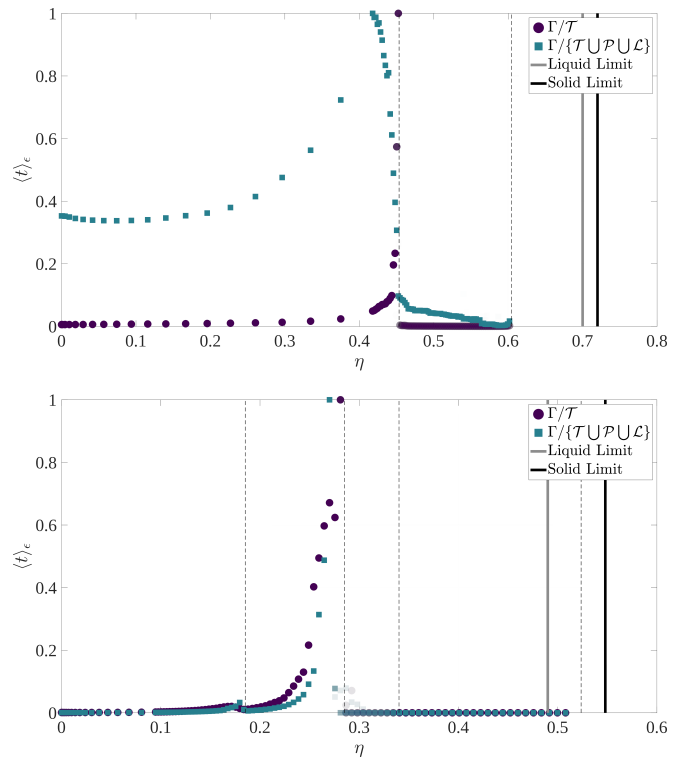

FIG. 8: ϵ-mixing times of the quotient spaces $\Gamma(2,\eta)/\mathcal{T}$ and $\Gamma(2,\eta)/\{\mathcal{T},\mathcal{P},\mathcal{L}\}$ for hard disks (top) and hard spheres (bottom). Dashed lines represent the packing fractions of the critical points.

time, and while the structure in the signal could perhaps be analyzed for further geometric information that is not the purpose of our study. The third is that while the discontinuities in the diffusion distance and ϵ-mixing time do not always occur at the same packing fractions, the largest discontinuities generally do.

Figure 10 shows the corresponding results to Fig. 9 for $n = 3\ldots6$ hard spheres. While the number of critical points is greatly increased relative to the hard disk systems in Fig. 9, the number of discontinuities in the diffusion diameter and ϵ-mixing time are approximately the same. This suggests that there is perhaps a small class of critical points associated with substantial geometric changes to the configuration space, and that the distribution of these critical points is the most relevant to the underlying hypothesis. Notice particularly that the packing fraction of the largest discontinuities appears to be approaching the packing fraction of the lower end of the phase-coexistence interval with increasing n. Unfortunately, even using all of the techniques in Sec. II C, the memory requirements increase so rapidly with n that we could not realistically examine the spaces for $n \geq 7$. Even for five and six hard spheres, the volume of the space increases so rapidly with decreasing packing fraction that we cannot report diffusion diameters and ϵ-mixing times over the entire domain. This does not substantially affect our conclusions though, since the sampled domain of η
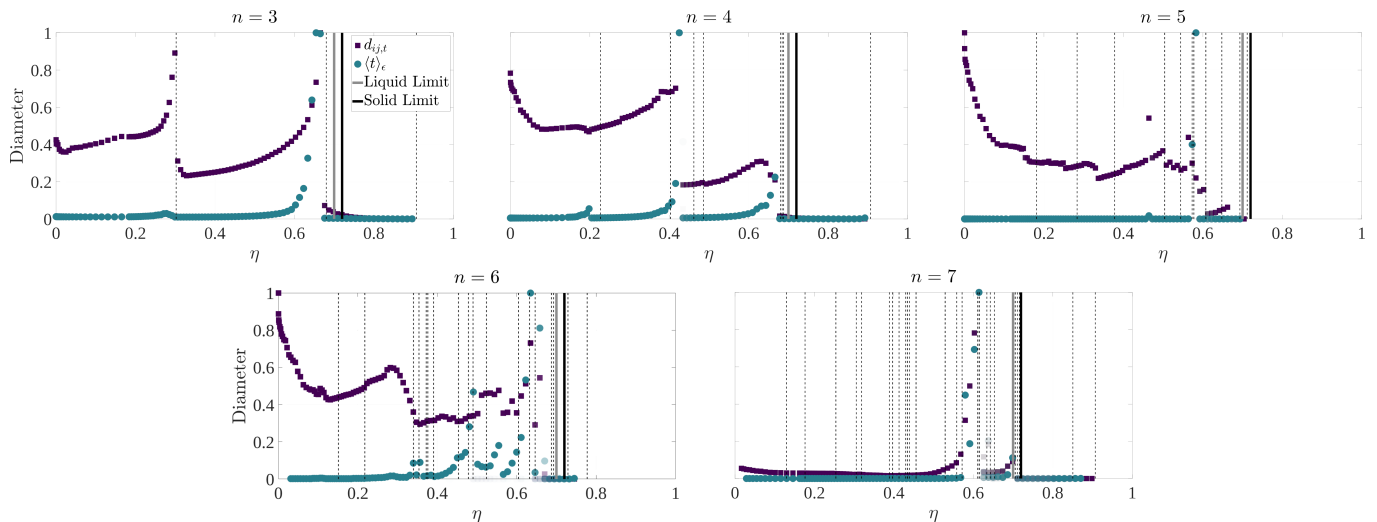
FIG. 9: Diffusion diameters and ϵ-mixing times of the spaces $\Gamma/\{\mathcal{T}, \mathcal{P}, \mathcal{L}\}$ for $n = 3 \ldots 7$ hard disks as functions of packing fraction. Dashed lines shows the packing fractions of the critical points.

already extends well below the coexistence interval, and notably includes all of the index-1 critical points.

There is evidence that, at least for the hard disk systems, the substantial geometric changes leading to discontinuities in the ϵ-mixing time are often associated with the lowest-packing-fraction index-1 critical point. Figure 11 shows the packing fractions of the largest discontinuities in Figs. 9 and 10 as black squares, with the number of hard disks (top) and hard spheres (bottom) increasing on the vertical axes. The packing fractions of all known critical points are also shown, with index-0 critical points in purple, index-1 critical points in blue, and all others in green. While ϵ-mixing time data is only available for $n \le 7$ hard disks and $n \le 6$ hard spheres, populations of critical points for $n \le 12$ hard disks and hard spheres have been sampled using established techniques that are described elsewhere [35–37]. Apart from the increasing concentration of low-index critical points around the phase coexistence interval with increasing n, the most striking aspect of the figure is that the largest discontinuities almost always occur close to the packing fraction of the last index-1 critical point to appear with decreasing packing fraction (each of the black squares comes slightly before an index-1 critical point due to finite sampling). This is not entirely unexpected, since each index-1 critical point either joins previously disconnected components or add a new class of closed paths to the space. Supposing that discontinuities in the ϵ-mixing time are associated with the former, it also makes sense that the largest discontinuity would be observed at lower packing fractions where the disconnected components being joined had the opportunity to grow to substantial volumes. The increasing concentration of low-index critical points around the phase coexistence interval draws the discontinuity closer to the liquid limit with increasing n, though that is admittedly a noisy trend for the small numbers of disks considered here. Nevertheless, this does provide evidence that something along the lines of Conj. 1 could be true for hard sphere systems, and perhaps for thermodynamics systems as well.

V. CONCLUSION

A phase transition is necessarily related to a discontinuous change in the probability density function describing the distribution of the system's microstates on the phase space. The question of how this could occur as the thermodynamic control variables are continuously varied has not been conclusively answered in the literature. One proposal called the topological hypothesis [7, 8] suggests that topological changes to the accessible region of the configuration space is a necessary condition for a phase transition. This paper instead suggests that a substantial change to the geometry of the accessible region is a necessary condition for a phase transition, and that such a change is often (but not always) associated with a topological change. More specifically, Conj. 1 proposes that a discontinuity in the mixing time on the configuration space is a necessary condition for a first-order phase transition in the thermodynamic limit. Our main result is a preliminary test of this conjecture for hard disk and hard sphere systems with few enough disks that the configuration space geometry can be explicitly studied.

The configuration spaces of hard disks and hard spheres are represented as graphs with vertices representing specific configurations of disks and edges indicating the distances between the configurations. The vertices of the graph are sampled more densely around critical points, or locations where the topology of the configuration space is known to change as a function of disk radius, to more accurately capture the topological and geometric changes in those regions. The diffusion diameters and
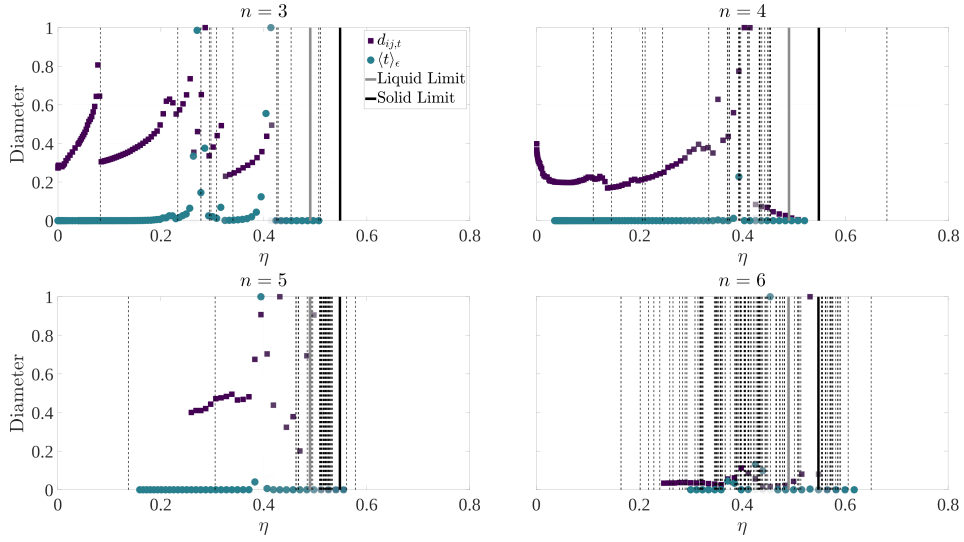
FIG. 10: Diffusion diameters and ϵ-mixing times of the spaces $\Gamma/\{\mathcal{T}, \mathcal{P}, \mathcal{L}\}$ for $n = 3 \ldots 6$ hard spheres as functions of packing fraction. Dashed lines shows the packing fractions of the critical points.
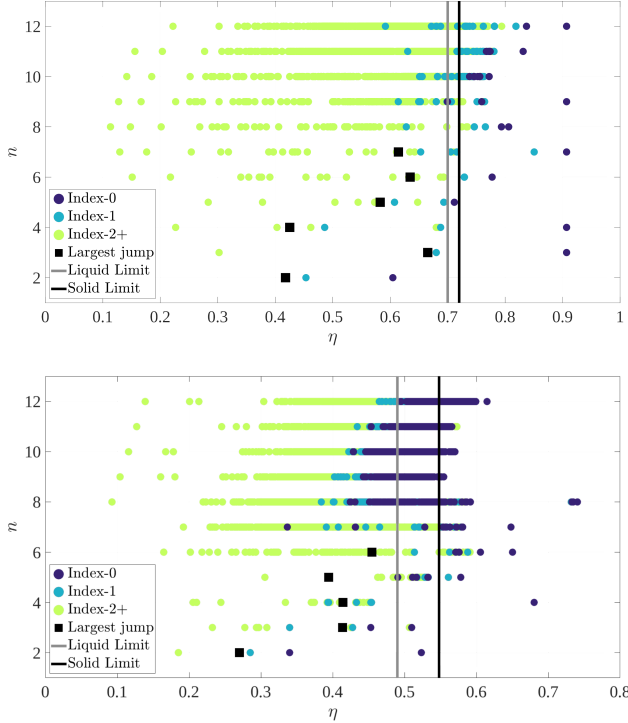

FIG. 11: The largest discontinuous jumps (black squares) observed in Figs. 9 (top) and 10 (bottom) along with the packing fractions of all known critical points, colored by their indices (index-0 in purple (dark circles), index-1 in blue (medium circles), all others in green (light circles)).

the ϵ-mixing times of the resulting graphs are calculated as functions of packing fraction for $n \leq 7$ and $n \leq 6$ hard spheres; the ϵ-mixing time is proposed here, and is designed to be a proxy to the thermodynamic mixing time that can be explicitly evaluated.

The geometric signals obtained for the diffusion diameter and the ϵ-mixing time are consistent in the sense that their discontinuities generally occur at the same packing fractions. Apart from relating more directly to the content of Conj. 1, the ϵ-mixing time is a much smoother function of packing fraction than the diffusion diameter. The discontinuities in the ϵ-mixing time are found to occur at comparatively few critical points; for the hard disk and hard sphere systems at least these are predominantly index-1 critical points. Along with the observation that the low-index critical points are increasingly concentrated around the phase coexistence interval with the number of disks, this suggests that a discontinuous change in the ϵ-mixing time could indeed coincide with the first-order phase transitions in the hard disk and hard sphere systems in the thermodynamic limit.

Future studies along these lines would likely need to extend the analysis to larger numbers of disks to make the trends in the approach to the thermodynamic limit clearer. Given the rate of increase in the computational requirements with number of disks, this would require at a minimum more efficient algorithms to search for critical points and calculate the distances between configurations. The computational requirements would also be reduced by using a further-reduced representation of the spaces by means of witness graphs [52], or by restricting to particular intervals of disk radius close to the phase coexistence region. Further development of a suitable min-type Morse theory would also be helpful [34, 53].
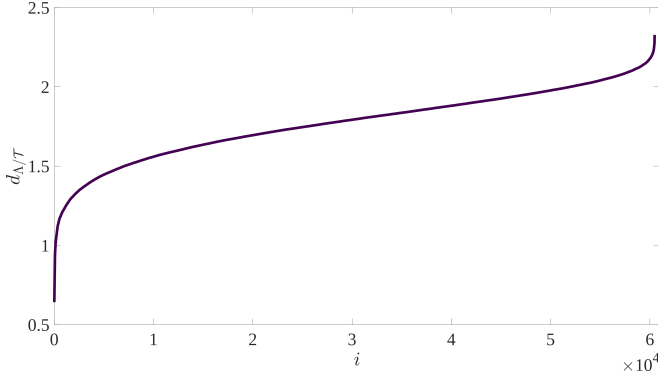
FIG. 12: The sorted true distances $d_{\Lambda/\mathcal{T}}$ calculated for a fixed generic configuration and the $7! \times 12$ symmetric versions of a second generic configuration for 7 hard disks.

Appendix A: Heuristic for the distance

Let $\mathcal{S} = \{\mathcal{T}, \mathcal{P}, \mathcal{L}\}$ in Eq. 4. Since $\mathcal{T}$ is a continuous group and $\{\mathcal{P}, \mathcal{L}\}$ is discrete, Eq. 4 can be rewritten as

$$d_{\Lambda/\mathcal{S}}(\mathbf{x}, \mathbf{y}) = \min_{S \in \{\mathcal{P}, \mathcal{L}\}} d_{\Lambda/\mathcal{T}}[\mathbf{x}, S(\mathbf{y})]$$

$$d_{\Lambda/\mathcal{T}}[\mathbf{x}, S(\mathbf{y})] = \inf_{\boldsymbol{t} \in \mathcal{T}} d_\Lambda[\mathbf{x}, S(\mathbf{y}) + \boldsymbol{t}]$$

$$d_\Lambda[\mathbf{x}, S(\mathbf{y}) + \boldsymbol{t}] = \sum_{i=1}^{n} \|\boldsymbol{x}_i - S(\boldsymbol{y}_i) - \boldsymbol{t}\|.$$

Evaluating $d_{\Lambda/\mathcal{T}}[\mathbf{x}, S(\mathbf{y})]$ involves solving a global optimization problem over rigid translations for a fixed discrete symmetry operation S. Since evaluating $d_{\Lambda/\mathcal{S}}(\mathbf{x}, \mathbf{y})$ involves solving $n! \times O(\mathcal{L})$ of these global optimization problems, the computational cost could be significantly reduced if a suitable approximation for $d_{\Lambda/\mathcal{T}}[\mathbf{x}, S(\mathbf{y})]$ could be found to quickly reject some of the S.

Such an approximation is constructed by, for a fixed number of iterations m, randomly sampling a random translation $\mathbf{t}$, calculating the distance $d_\Lambda[\mathbf{x}, S(\mathbf{y}) + \mathbf{t}]$ for that translation, and accepting that translation only if it reduces $d_\Lambda[\mathbf{x}, S(\mathbf{y}) + \mathbf{t}]$ [54]. The resulting approximations for $d_{\Lambda/\mathcal{T}}[\mathbf{x}, S(\mathbf{y})]$ are sorted by increasing magnitude, and the full optimization problem is solved only for the first M symmetry operations in the sorted list. The rationale for this procedure is that calculating the approximations using a moderate m is less expensive than solving the global optimization problem.

For two generic configurations of 7 hard disks there are $7! \times 12 = 60\,480$ symmetric versions of the second configuration due to the actions of $\mathcal{P}$ and $\mathcal{L}$. Figure 12 shows the sorted true distances $d_{\Lambda/\mathcal{T}}$, and $d_{\Lambda/\mathcal{S}}$ is the minimum of this set. Figure 13 shows the same configurations sorted by the approximate values of $d_{\Lambda/\mathcal{T}}[\mathbf{x}, S(\mathbf{y})]$ for $m = 1$, 10, 50, and 100 plotted with the true values of $d_{\Lambda/\mathcal{T}}[\mathbf{x}, S(\mathbf{y})]$ on the vertical axis. Observe that with increasing m these curves should converge to the one in Fig. 12, and that even for small m the true minimum distance should appear within the first $M \ll 60\,480$ symmetry operations. Numerical experiments suggest that $m = 50$ and $M = 200$ are generally sufficient for an approximation with a relative error on the order of 10^{-4}.

Appendix B: Eigenspectrum analysis

The diffusion coordinates in Eq. 5 involve the tth powers of the eigenvalues of $\mathbf{P}$ sorted by decreasing magnitude. This means that for sufficiently large t, the computational expense of calculating the diffusion distance can be considerably reduced with negligible loss of accuracy by truncating the eigenspectrum and only considering the first few diffusion coordinates [15]. Figure 14 shows the tth powers of the eigenvalues of $\mathbf{P}$ for $\Lambda(4)/\{\mathcal{T}, \mathcal{P}, \mathcal{L}\}$ at times $t = 1$, 5, and 10. Observe that even for the relatively small $t = 5$ most of the diffusion coordinates will be numerically indistinguishable from zero. Only the first 200 diffusion coordinates are used here unless otherwise specified, changing the diffusion diameter to a relative error of less than 10^{-8} and making this source of error negligible compared to other sources.

Appendix C: Mixing time of an example system

Figure 15 shows the ϵ-mixing times $t_\epsilon(\boldsymbol{q}_0)$ as a function of $\boldsymbol{q}_0$ for the example system in Fig. 1. The k-nearest neighbor graph is constructed with $k = 7$, the length of an edge is given by the Euclidean distance between the corresponding vertices, and $\epsilon = 0.001$ is used since the steady-state probability mass at each vertex is 0.0005 (the inverse of the number of vertices). Points closer to the center have smaller ϵ-mixing times since the initial Dirac-delta distribution can spread to the upper and lower regions more easily as compared to distributions starting near the corners. The average mixing time $\langle t \rangle_\epsilon$ for this example is around 15 000.

[1] L. Landau, Nature **138**, 840–841 (1936).
[2] L. D. Landau and E. M. Lifshitz, *Statistical Physics: Volume 5*, vol. 5 (Elsevier, 2013).
[3] L. E. Boltzmann, Wiener Berichte **63**, 679–711 (1871).
[4] P. Ehrenfest and T. Ehrenfest, *The conceptual foundations of the statistical approach in mechanics* (Courier Corporation, 1990).
[5] G. D. Birkhoff, Proceedings of the National Academy of Sciences **17**, 656–660 (1931).
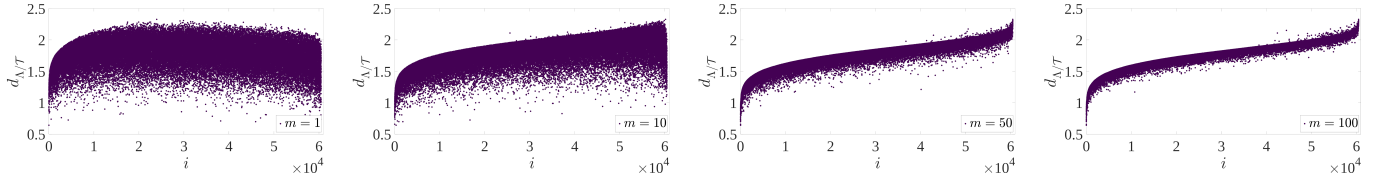[6] J. v. Neumann, Proceedings of the National Academy of

FIG. 13: The same configurations as in Fig. 12 are sorted by the approximate values of $d_{\Lambda/\mathcal{T}}[\mathbf{x}, S(\mathbf{y})]$ for $m = 1, 10, 50,$ and 100 and are plotted with the true values of $d_{\Lambda/\mathcal{T}}[\mathbf{x}, S(\mathbf{y})]$ on the vertical axis.
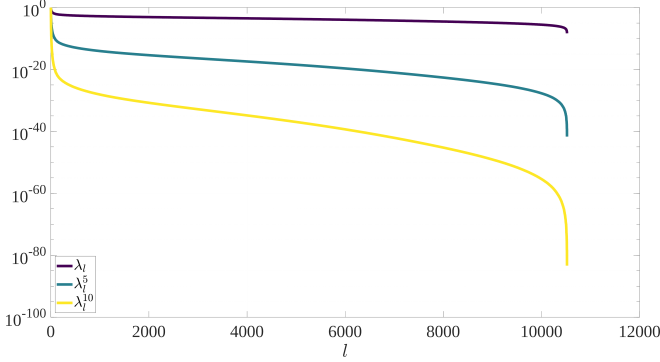

FIG. 14: The tth powers of the eigenvalues of $\mathbf{P}$ for $\Lambda(4)/\{\mathcal{T}, \mathcal{P}, \mathcal{L}\}$ at times $t = 1, 5,$ and 10. The values generally decay rapidly, and are often numerically indistinguishable from zero for sufficiently large t.
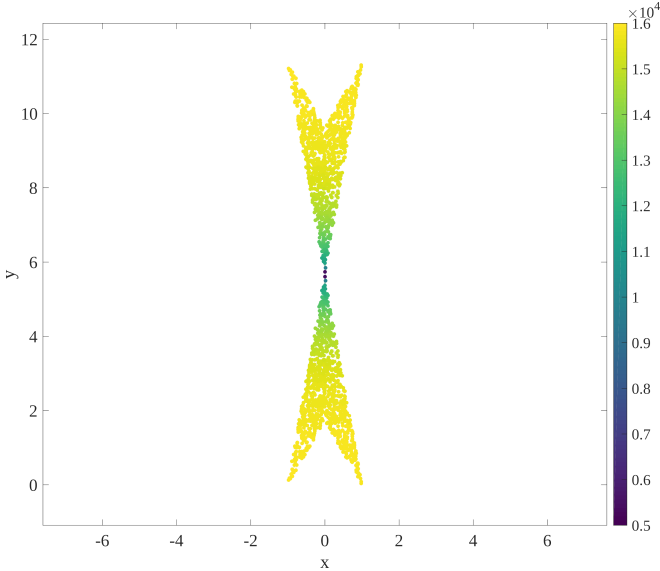

FIG. 15: Individual mixing times $t_\epsilon(\boldsymbol{q}_0)$ of the example system shown in Fig. 1. The average mixing time $\langle t \rangle_\epsilon$ is around 15 000.

Sciences **18**, 70–82 (1932).

[7] L. Caiani, L. Casetti, C. Clementi, and M. Pettini, Physical review letters **79**, 4361 (1997).

[8] R. Franzosi and M. Pettini, Physical review letters **92**, 060601 (2004).

[9] R. Franzosi, M. Pettini, and L. Spinelli, Nuclear Physics B **782**, 189–218 (2007).

[10] R. Franzosi and M. Pettini, Nuclear Physics B **782**, 219–240 (2007).

[11] M. Kastner and D. Mehta, Physical review letters **107**, 160602 (2011).

[12] M. Gori, R. Franzosi, and M. Pettini, Journal of Statistical Mechanics: Theory and Experiment **2018**, 093204 (2018).

[13] G. Bel-Hadj-Aissa, M. Gori, R. Franzosi, and M. Pettini, Journal of Statistical Mechanics: Theory and Experiment **2021**, 023206 (2021).

[14] R. R. Coifman and S. Lafon, Applied and computational harmonic analysis **21**, 5–30 (2006).

[15] R. Talmon, I. Cohen, S. Gannot, and R. R. Coifman, IEEE signal processing magazine **30**, 75–86 (2013).

[16] S. Kullback and R. A. Leibler, The annals of mathematical statistics **22**, 79–86 (1951).

[17] C.-N. Yang and T.-D. Lee, Physical Review **87**, 404 (1952).

[18] T.-D. Lee and C.-N. Yang, Physical Review **87**, 410 (1952).

[19] I. Bena, M. Droz, and A. Lipowski, International Journal of Modern Physics B **19**, 4269–4329 (2005).

[20] J. C. Dyre, Journal of Physics: Condensed Matter **28**, 323001 (2016).

[21] B. Alder and T. Wainwright, Physical Review **127**, 359 (1962).

[22] H. Weber, D. Marx, and K. Binder, Physical Review B **51**, 14636 (1995).

[23] A. C. Mitus, H. Weber, and D. Marx, Physical Review E **55**, 6855 (1997).

[24] E. P. Bernard and W. Krauth, Physical review letters **107**, 155704 (2011).

[25] M. Engel, J. A. Anderson, S. C. Glotzer, M. Isobe, E. P. Bernard, and W. Krauth, Physical Review E **87**, 042134 (2013).

[26] A. L. Thorneywork, J. L. Abbott, D. G. A. L. Aarts, and R. P. A. Dullens, Phys. Rev. Lett. **118**, 158001 (2017).

[27] B. I. Halperin and D. R. Nelson, Physical Review Letters **41**, 121 (1978).

[28] A. Young, Physical Review B **19**, 1855 (1979).

[29] M. Isobe and W. Krauth, The Journal of chemical physics **143**, 084509 (2015).

[30] S. Pieprzyk, M. N. Bannerman, A. C. Brańka, M. Chudak, and D. M. Heyes, Physical Chemistry Chemical Physics **21**, 6886–6899 (2019).

[31] G. Carlsson, J. Gorham, M. Kahle, and J. Mason, Physical Review E **85**, 011303 (2012).

[32] M. Morse, *The calculus of variations in the large*, vol. 18 (American Mathematical Soc., 1934).

[33] J. Milnor, *Morse theory.(AM-51)*, vol. 51 (Princeton university press, 2016).

[34] Y. Baryshnikov, P. Bubenik, and M. Kahle, International

Mathematics Research Notices **2014**, 2577–2592 (2014).

[35] K. Ritchey, Ph.D. thesis, Ohio State University, Electronic Thesis or Dissertation (2019).

[36] O. B. Ericok and J. K. Mason, Granular Matter **24**, 1–15 (2022).

[37] O. B. Ericok, K. Ganesan, and J. K. Mason, Phys. Rev. E **104**, 055304 (2021).

[38] D. Burago, I. D. Burago, Y. Burago, S. Ivanov, and S. A. Ivanov, *A course in metric geometry*, vol. 33 (American Mathematical Soc., 2001).

[39] F. Glover, ORSA Journal on computing **1**, 190–206 (1989).

[40] R. Chelouah and P. Siarry, European journal of operational research **123**, 256–270 (2000).

[41] D. R. Sheehy, Discrete & Computational Geometry **49**, 778–796 (2013).

[42] M. Buchet, F. Chazal, S. Y. Oudot, and D. R. Sheehy, Computational Geometry **58**, 70–96 (2016).

[43] H. Kahn and T. E. Harris, National Bureau of Standards applied mathematics series **12**, 27–30 (1951).

[44] J. Liouville, Journal de Mathématiques Pures et Appliquées **3**, 342–349 (1838).

[45] J. W. Gibbs, *Elementary principles in statistical mechanics: Developed with especial reference to the rational foundations of thermodynamics* (C. Scribner's sons, 1902).

[46] T. Gleim, W. Kob, and K. Binder, Physical review letters **81**, 4404 (1998).

[47] G. Szamel and E. Flenner, Europhysics Letters **67**, 779 (2004).

[48] L. Berthier and W. Kob, Journal of Physics: Condensed Matter **19**, 205130 (2007).

[49] F. R. Chung and F. C. Graham, *Spectral graph theory*, 92 (American Mathematical Soc., 1997).

[50] A. Mulero, C. Galán, M. Parra, and F. Cuadros, in *Theory and Simulation of Hard-Sphere Fluids and Related Systems* (Springer, 2008), p. 37–109.

[51] J. Tian, H. Jiang, and A. Mulero, Molecular Physics **118**, e1687948 (2020).

[52] B. Aronov, M. Dulieu, and F. Hurtado, Computational Geometry **46**, 894–908 (2013), ISSN 0925-7721, euroCG 2009.

[53] F. Mori and M. Salvetti, Mathematical Research Letters **18**, 39–57 (2011).

[54] M. T. Wells, G. Casella, and C. P. Robert, in *A Festschrift for Herman Rubin* (Institute of Mathematical Statistics, 2004), p. 342–347.