Title
Statistical Topology of the Configuration Spaces of Hard Disks and Spheres

Author
Ericok, Ozan Burak

Publication Date
2022

Peer reviewed|Thesis/dissertation

Statistical Topology of the Configuration Spaces of Hard Disks and Spheres

By

OZAN BURAK ERICOK
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Materials Science and Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Jeremy Mason, Chair

Sabyasachi Sen

Roland Faller

Committee in Charge

2022

In dedication to my lovely wife Ezgi and our wonderful son Gece.

Contents

Statistical Topology of the Configuration Spaces of Hard Disks and Spheres

Abstract

As phenomena that necessarily emerge from the collective behavior of interacting particles, phase transitions continue to be difficult to predict using statistical thermodynamics. A phase is generally defined as a part of a thermodynamic system with uniform physical properties. For example, the liquid and vapor forms of water are distinct phases that can be mechanically separated. One question to be asked is whether the transition between them always involves a discontinuous change in properties though. The short answer to this question is that it depends on the path that the system takes. It is known that there is a first-order transition if the path crosses the liquid-vapor line in phase diagram of water, but there is no transition if the path goes around the critical point. One can then ask whether there is an entirely unambiguous definition of a phase, or only of a phase transition. Further, setting aside the question of whether a phase can be precisely defined, can a phase transition be predicted on the basis of statistical thermodynamics?

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 associates first-order phase transitions with spontaneous symmetry breaking as quantified by an appropriately-constructed order parameter. This work instead suggests that phase transitions are often associated with dramatic changes 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 in the configuration spaces of hard disk and hard sphere systems of increasing size. These systems are often used to model simple fluids as they are one of the simplest systems that can be computationally studied and yet are complicated enough to undergo phase transitions. Initially, a framework to geometrically construct the symmetry invariant configuration spaces of hard disks and spheres is developed and applied to systems of a few hard disks and spheres. Particularly, a metric is defined on the base configuration space and various quotient

spaces that respects the desired symmetries. This is used to construct explicit triangulations of the configuration spaces as α-complexes. The critical configurations are found to be associated with geometric changes to the configuration space that connect previously distant regions and reduce the configuration space diameter. The number of such critical configurations around the packing fraction of the solid-fluid phase transition increases exponentially with the number of disks and spheres, suggesting that the onset of the first-order phase transition in the thermodynamic limit is associated with a discontinuity in the configuration space diameter.

The implication is that a phase transition in the thermodynamic limit is associated with a discontinuous change in the probability distribution on the phase space, and this concept is initially explored for finite non-equilibrium systems. Definitions are proposed for all of the extrinsic variables of the fundamental thermodynamic relation that are consistent with existing results in the equilibrium thermodynamic limit. The probability density function on the phase space is interpreted as a subjective uncertainty about the microstate, and the Gibbs entropy formula is modified to allow for entropy creation without introducing additional physics or modifying the phase space dynamics. Resolutions are proposed to the mixing paradox, Gibbs' paradox, Loschmidt's paradox, and Maxwell's demon thought experiment. Finally, the extrinsic variables of the fundamental thermodynamic relation are evaluated as functions of time and space for a diffusing ideal gas, and the initial and final values are shown to coincide with the expected equilibrium values when interpreted in a classical context.

Acknowledgments

The people acknowledged below and many others have influenced the way I approach my research problem in different ways. I sincerely believe that I can not thank them enough.

First, I would like to thank my advisor Prof. Jeremy Mason. You always treated me with respect, and provided an excellent environment for research. I thank you for trusting me on this project and for having patience while I was trying to find my way on the project. I am grateful that we had a lot of scientific discussions, and I will truly miss meeting with you in your office and talking about research.

Second, I would like to thank my dear friend Erman Cineli for his help. You were always there whenever I needed help for advanced mathematical concepts. I did not only learn new mathematical concepts from you, but also how a mathematician approaches a particular problem. I truly miss the nights that we gathered together and had scientific discussions for hours.

Third, I would like to thank the current and former members of the Mason research group. I learned a lot of interesting research ideas from them over the last four years. I would also like to thank Prof. Sabyasachi Sen and Prof. Roland Faller for accepting to be in my dissertation committee. I gratefully acknowledge NSF grant #1839370, which partially supported this work.

Finally, I would like to thank my lovely wife Ezgi for her extraordinary patience and support along this journey. She was always the first person who listened my weird ideas. Thank you for encouraging me whenever I needed it.

CHAPTER 1

Introduction*

While statistical thermodynamics is generally considered to provide a satisfactory explanation of equilibrium states, the features of the configuration space that regulate the appearance of a phase transition are not yet entirely clear. Consider n particles in a spatial domain X. Let Γ_n be the phase space of the system such that a single point in Γ_n completely specifies the microscopic state, i.e., the positions and momenta of all the particles. Given an initial microstate, the system's time evolution corresponds to a continuous trajectory in the phase space. A thermodynamic property measurement effectively averages the value of this property along the portion of the trajectory corresponding to the measurement time interval. The ergodic hypothesis [**19**] suggests that this time average can be replaced by a weighted average over the entire phase space if the time required to explore the phase space is substantially less than that of the measurement; the weight is a canonical measure that indicates the probability of observing a microstate at a generic instant in time. Finally, the Ehrenfest classification of phase transitions defines the transition order as the degree of the lowest derivative of the free energy that is discontinuous [**70**]. If ensemble averages are used to calculate these derivatives, and the canonical measure and the thermodynamic properties are smooth functions on the phase space, then from where do the discontinuities at a phase transition arise? If one objects that discontinuities should only be expected in the thermodynamic limit, there should nevertheless be incipient discontinuities in finite systems that develop into proper discontinuities in the limit. From where do these incipient discontinuities arise?

The topological hypothesis [**23**, **52**] roughly suggests that the essence of the difficulty is the assumption that ergodicity holds close to a phase transition. That is, averages over the phase space should only be performed over the region that is effectively accessible from the initial microstate. If the extent of this accessible region changes discontinuously as a function of some thermodynamic variable, then the region over which the averages are computed would change discontinuously as

*The content of this chapter has previously been published in Refs. [**45**, **47**, **48**].

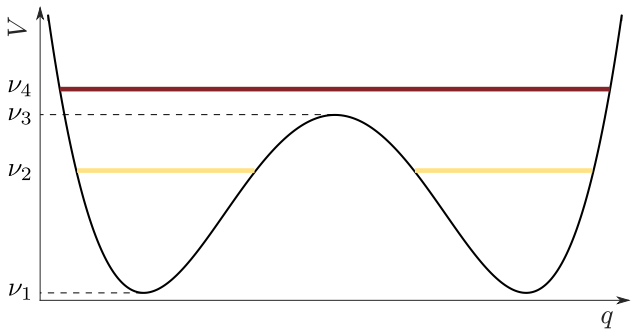

FIGURE 1.1. The sublevel sets of the potential energy surface Σ_ν consist of two disconnected components for energy levels $\nu_1 < \nu < \nu_3$. The topology of the sublevel set changes at $\nu = \nu_3$ where the two disconnected components merge.

well, and could result in discontinuous thermodynamic properties. Slightly more formally, the topological hypothesis more specifically asserts that discontinuous changes to the accessible part of the configuration space with increasing potential energy are associated with topological changes in the equipotential energy submanifolds; the configurations associated with these topological changes are known as critical points (a name that unfortunately conflicts with that for the end point of a phase equilibrium curve). Intuitively, the critical points could be the highest energy points along paths between two previously disconnected regions, or shortcuts between two previously distant regions. Franzosi et al. [**53**, **54**, **60**] claimed to prove that a topological change in 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 where $V(\bar{q}_1, \ldots, \bar{q}_N)$ is the potential energy of the particles at positions $\bar{q}_i$. Kastner and Mehta [**78**] subsequently studied a two-dimensional Φ^4 system and observed a second-order phase transition occurring in an energy interval that didn't contain any critical points, disproving the claim. Gori et al. [**60**] responded by refining the hypothesis, suggesting that the phase transition in the Φ^4 system is actually related to a divergence in the time required for the system to transition between two parts of the configuration space.

As evidence that the topology and geometry of the accessible part of the configuration space should be functions of the thermodynamic control variables, consider the one-dimensional potential energy surface in Fig. 1.1. The three distinct critical points of V are to the two minima associated with the basins and the one saddle point associated with the barrier. A sublevel set of the potential

energy surface Σ_ν refers to the subset of the configuration space where the potential energy is less than or equal to ν, and coincides with the potentially accessible region. Observe that Σ_ν is empty for energies $\nu < \nu_1$, but changes for $\nu = \nu_1$ to the union of the two points at the minima. The disconnected components in the two basins grow for $\nu_1 < \nu < \nu_3$, remaining separated until $\nu = \nu_3$ where the saddle point appears and merges the previously disconnected components. Finally, the sublevel set grows as a single component for $\nu > \nu_3$. Suppose now that a random walker is initially positioned in the left basin with the energy $\nu = \nu_2$ as shown in yellow in Fig. 1.1. Since the sublevel set at this energy level consists of two disconnected components, the random walker cannot transition to the right basin and can only explore the part of the configuration space connected to its initial position. Raising the energy to $\nu > \nu_3$ discontinuously changes the connectivity of the space and the region accessible to the random walker; such discontinuous changes to the connectivity of the accessible region are referred to as topological changes. By contrast, the ergodic assumption suggests that the averages should be computed over both basins for all values of the internal energy, and is unreasonable from the standpoint of possible trajectories of the microstate.

The literature and the discussions above do suggest that there is a relationship between the distribution of critical points and the onset of a phase transition, but a less direct relationship than initially supposed. Perhaps the natural signature of a phase transition is a discontinuity in the mixing time required for the system to adequately explore the configuration space geometry, and such a discontinuity is generally (but not always) associated with topological changes and critical points. The significance of the configuration space geometry is also supported by a recent study of the topological and geometric properties of the two-dimensional XY model by Bel-Hadj-Aissa et al. [**9**]; they computed the mean geometric 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 these as background, the fundamental conjecture of this study is:

CONJECTURE 1.0.1. *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, or whether this will always remain a vague quantity significant only in that it is necessarily shorter than the time scales of quasiequilibrium processes. It is hypothesized 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 [**29**,**118**] 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.

The conjecture is tested 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 [**36**], 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 [**2**] 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 [**98**, **128**]. 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 recently was claimed to contain a hexatic phase [**63**, **133**]. It is significant that there is still not agreement in the literature about the order

of the transitions and the phases involved [**10**, **44**, **98**, **128**] even for this simple system. Similar studies of hard spheres in three dimensions include those by Isobe and Krauth [**69**] and Pieprzyk et al. [**110**] 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. [**25**] explored the topology of the configuration space of five hard disks in a square box by regularizing the potential energy surface and using classical Morse theory [**97**, **102**]. Baryshnikov et al. [**6**] 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. Ritchey [**113**] studied configuration spaces of hard disk in the hexagonal torus 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 recent studies of Ericok and Mason [**45**, **46**, **47**, **48**] constituting the building blocks of this dissertation also consider the configuration spaces of hard disks and spheres from a geometric point of view. They proposed a rigorous procedure to explicitly triangulate the configuration spaces of hard disks and to study their geometric properties.

One immediate purpose of this study is to develop a rigorous framework to practically test Conj. 1.0.1 using hard disk systems. The overall purpose of this work, however, is to understand the very nature of the configuration space, and how its topology and geometry affect the evolution of non-equilibrium thermodynamic properties (e.g., entropy) of thermodynamic systems. Chapter 2 provides the mathematical formulation of the configuration space of hard disks and describes some of the necessary mathematical concepts that are used to study configuration spaces; namely, a brief introduction to Morse theory, the numerical procedure to identify critical points of the configuration space of hard disks, the definitions of the symmetries present in the configuration spaces, geometric representation and properties of the configuration spaces of hard disks. Chapter 3 and 4 respectively discuss how various symmetry groups affect the topological and geometric properties of the configuration space of hard disks and spheres. Explicit triangulations of the configuration spaces for small numbers of disks and spheres are constructed, and the relationship of the critical points to the configuration space topology and geometry is explored. Chapter 5 introduces the concepts of the ϵ-mixing time and studies the configuration space diameter as a

function of disk radius for increasing numbers of disks and spheres. Finally, Chapter 6 begins building the foundations for a finite, non-equilibrium statistical thermodynamics that is consistent with existing results in the equilibrium thermodynamic limit, and that introduces a minimum of arbitrary constants and additional physics.

CHAPTER 2

Preliminaries*

2.1. Tautological function

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

$$\Lambda(n) = \{\mathbf{x} = (\bar{x}_1 \dots \bar{x}_n) \,|\, \bar{x}_i \in T^d\}. \tag{2.1}$$

$\Lambda(n)$ will often be called the base configuration space in the following since all the other spaces in this work are derived from it. A two-dimensional torus T^2 for hard disks is obtained by identifying opposite edges of a regular square or hexagon, whereas a three-dimensional torus T^3 for hard spheres is obtained by identifying opposite faces of a rhombic dodecahedron. The justification for the choices of the hexagonal torus and the rhombic dodecahedron is that the densest packings in the corresponding dimensions can be constructed by repeated translations of the fundamental cells with their inscribed sphere. Figure 2.1 shows these domains with some of the periodic images of the fundamental unit cells. The center to center distance to the neighboring cells is one in all cases. The formulation below applies for both hard disks and spheres, and both systems are called a hard disk system in the following.

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

$$\tau(\mathbf{x}) = \min_{1 \le i < j \le n} r_{ij} \tag{2.2}$$

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

$$\Gamma(n, \rho) = \tau^{-1}([\rho, \infty)) \tag{2.3}$$

*The content of this chapter has previously been published in Refs. [**45**, **47**, **48**].

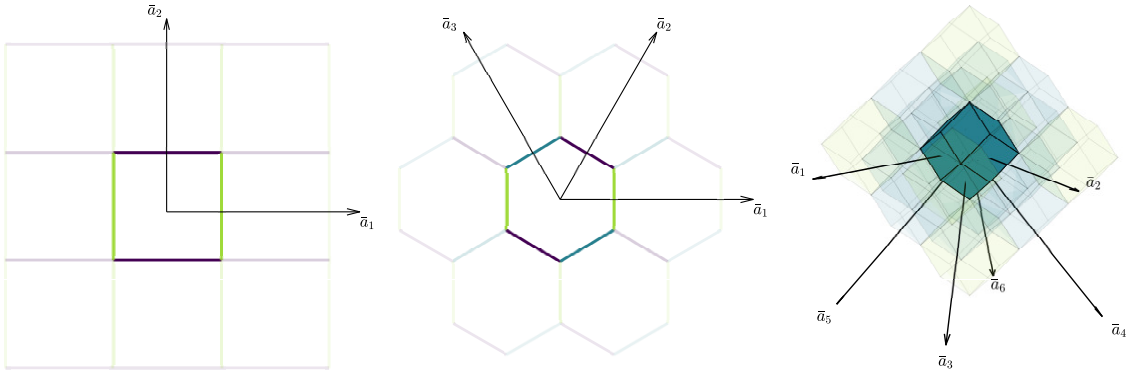

FIGURE 2.1. A two-dimensional square torus (left), two-dimensional hexagonal torus (middle) and a three-dimensional torus (right) are shown with some of their periodic images. The unit vector $\bar{a}_1$ points in the x-direction in both cases. All the unit vectors pass through the centers of the corresponding faces.

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

2.2. Morse theory

Morse theory [**97**, **102**] 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 vanishes, and is associated with 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. It is remarkable that while the choice of the function f is nearly arbitrary, the topological information gained by examining the critical points of f is a property of the manifold and is therefore independent of that choice.

Let $M_a = \{x \in M \,|\, f(x) < a\}$ 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 $D^1 \times D^1$. In three dimensions, a 0-handle and a 3-handle are solid balls $D^0 \times D^3$ and $D^3 \times D^0$ whereas a 1-handle and a 2-handle are solid cylinders $D^1 \times D^2$ and $D^2 \times D^1$.

Equation 2.3 represents the configuration space of hard disks $\Gamma(n, \rho)$ by means of the superlevel sets of τ. This should allow a Morse-type theory to be used with the the critical points of τ to identify changes in the configuration space topology. Baryshnikov et al. [**6**] began to develop a min-type Morse theory specifically for hard disk configuration spaces, and proved that the critical points of τ are precisely the mechanically-balanced configurations. However, the difficulty with using the functions τ or the hard disk potential in this context is that they are not differentiable. This is resolved by regularizing the hard disk potential, resulting in an infinitely-differentiable potential energy. More specifically, the hard disk potential is replaced by an exponential repulsion where ρ and w govern the onset and the strength of the repulsion, respectively:

$$E = \sum_i \sum_{j>i} \exp[-w(r_{ij} - \rho)].$$

This can be seen to converge to the hard disk potential in the limit of large w in Fig. 2.2. Moreover, the critical points of E can be related to those of τ (despite τ not being differentiable) in the following way. Observe that the radius ρ of the spheres could be increased by as much as $\tau - \rho = \min_{i<j}(r_{ij} - \rho)$ without the spheres overlapping. The soft-min approximation replaces the minimum by the logarithm of a sum of exponentials, or

$$\tau - \rho \approx -\frac{1}{w} \log \Big\{ \sum_i \sum_{j>i} \exp[-w(r_{ij} - \rho)] \Big\}, \tag{2.4}$$

where the sum is generically dominated by a single term corresponding to the minimum in the limit of large w. Observe that the critical points of the right side of Eq. 2.4 are the same as those of E for any finite value of w since the logarithm and multiplication by a constant are strictly monotone transformations, and the critical points of the left hand side are the same as those of τ since ρ is a constant. Equation 2.4 then suggests that the critical points of τ be defined to be those of E in the large w limit. It is for this reason that the critical points of the hard disk potential energy surface and the critical points of τ discussed in this work both effectively refer to the critical points of the differentiable function E in the $w \to \infty$ limit. The critical index is calculated as the number of negative eigenvalues of the Hessian matrix of E evaluated at the critical configuration for large

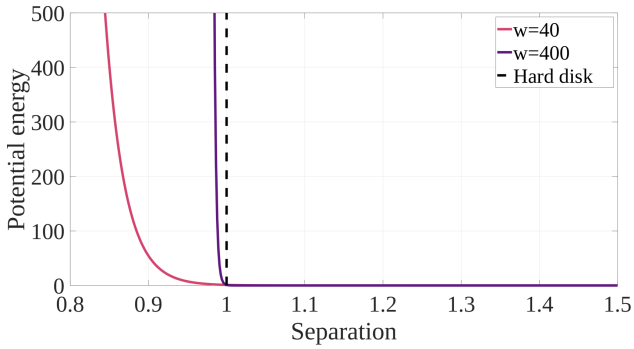

FIGURE 2.2. The hard disk potential (black dashed line) is infinite if the separation of two sphere centers is less than the sum of their radii and zero otherwise. Observe that E converges to the hard disk potential as $w \to \infty$.

but finite w. Ritchey [**113**] provided an alternative definition of the index suitable for min-type functions that coincides with the above definition for every known non-degenerate critical point.

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 $n = 1 \ldots 12$ hard disks[1] and spheres[2] are readily available in the interactive databases.

2.3. Symmetries

The specific relationship of the topology to the geometry depends not only on the potential energy function, but also on the symmetries considered to construct the configuration space. Initially consider fixing a coordinate system to identify points in a spatial region X, assigning labels to each of n particles in this region, and representing every possible configuration of the system by a point in the product space X^n. One possible issue with this approach is that assigning labels to the particles in two different ways would map a single configuration to two different points in the configuration space. The implications of removing this redundancy by forgetting the labels can be seen by considering a system of two hard disks in the hexagonal torus for which the critical points

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

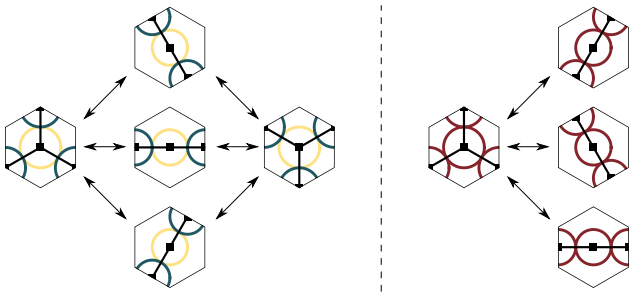

FIGURE 2.3. The critical points of the translation-invariant configuration spaces of two hard disks on a hexagonal torus (labeled on the left, unlabeled on the right). The configurations where the disks have three connections are minima of the hard disk potential, and those with two connections are saddle points.

of the hard disk potential are shown in Fig. 2.3. For a disk radius ρ, the accessible part of the configuration space is a connected component where every pair of disk centers is separated by at least 2ρ. The accessible region manifestly changes with ρ, being empty in the limit of large disk radius and the configuration space of points in the limit of small disk radius. 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 subspaces cannot transition 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.

More generally, redundant points in the configuration space are introduced by one or more symmetry groups acting on a given configuration of disks. The redundancies can be removed by mapping the set of points in the base configuration space that are related by the specified symmetry group to a single point in a new space, known as the quotient space, that preserves sensible notions of configuration similarity. Whereas a point in the base configuration space specifies a configuration of disks, a point in the quotient space instead specifies a class of equivalent configurations from the standpoint of the specified symmetry group. The map from the base configuration space to the quotient space is known as a quotient map, and this process is often referred to as quotienting the

base configuration space by the specified symmetry group. The standard approach in the literature seems to be to quotient the configuration space by all possible symmetries, e.g., the use of correct Boltzmann counting in statistical mechanics is generally regarded as being equivalent to quotienting the configuration space by permutations of the particle labels [**85**], and the homogeneity of space encourages the use of center of mass coordinates in classical mechanics [**32**]. However, the example in Fig. 2.3 suggests that quotienting by such symmetries could affect the geometry and topology of the configuration space in unexpected ways. One of the purposes of this study is to begin the exploration of what these effects could be.

The hard disk system defined in Eq. 2.3 has a variety of symmetries that are first studied by Ritchey [**113**], or physical operations that make multiple points in Λ physically indistinguishable. These include continuous translations $\mathcal{T}$, permutations of the sphere labels $\mathcal{P}$, inversions about the origin $\mathcal{I}$, and the symmetries of the tessellation $\mathcal{L}$. For example, exchanging the labels of two spheres changes the point in Λ but does not change the physical configuration. Formally, a configuration $\mathbf{q}$ is a translation of $\mathbf{p}$ by $\bar{t}$ if $\bar{q}_i = \bar{p}_i + \bar{t}$ where, e.g., $\bar{q}_i$ are the coordinates of the ith disk. Given a permutation $\pi \in \mathcal{P}$, $\mathbf{q}$ is a permutation of $\mathbf{p}$ if $\bar{q}_i = \bar{p}_{\pi(i)}$ for all i. A configuration $\mathbf{q}$ is the inverse of $\mathbf{p}$ if $\bar{q}_i = -\bar{p}_i$ for all i. Finally, for any symmetry element $L \in \mathcal{L}$ with representation $\bar{\bar{L}}$, a configuration $\mathbf{q}$ is a symmetric copy of $\mathbf{p}$ if $\bar{q}_i = \bar{\bar{L}}\bar{p}_i$ for all i.

2.4. Geometric representation

One of the computationally convenient way to study topological and geometric properties of a manifold is to use an explicit triangulation of it by means of a simplicial complex. Intuitively, a k-simplex is the generalization of a triangle to an arbitrary dimension. More formally, a k-simplex is the convex hull of $k+1$ affinely independent points in R^d for $0 \leq k \leq d$. In this sense, 0-simplex is a point, 1-simplex is a line segment, 2-simplex is a triangle with its interior included and 3-simplex is a tetrahedron with its interior included etc. Given a set of simplices, a simplicial complex is the collection of these simplices without improper intersections. An interested reader can find the precise definitions in Ref. [**104**]. Although there exists different simplicial complexes in the literature [**15**, **26**, **65**, **126**], the definitions of the simplicial complexes used in this work to geometrically represent the configuration spaces of hard disks are given below.

2.4.1. k-nearest neighbor graph. A k-nearest neighbor graph is a graph that contains edges between vertices p and q if the distance between them is among the k-th smallest distances from p to the rest of the vertices [**111**]. In the most general setting, the vertices might represent any object and the edges between them might represent any relation between the vertices.

The top row of Fig. 2.4 presents two k-nearest neighbor graphs of the same point cloud with $k = 5$ and $k = 20$. Observe that the structures of the graphs are significantly affected by the choice of k. However, there is not a universal rule to select an appropriate k value, and the choice strictly depends on the problem. Roughly speaking, a typical procedure usually experiments with different k values ranging from small to large, and the desired features of the graph are plotted as a function of k. An optimum (or near-optimum) value is then selected among the values of k that produces persistent and meaningful features in the plots.

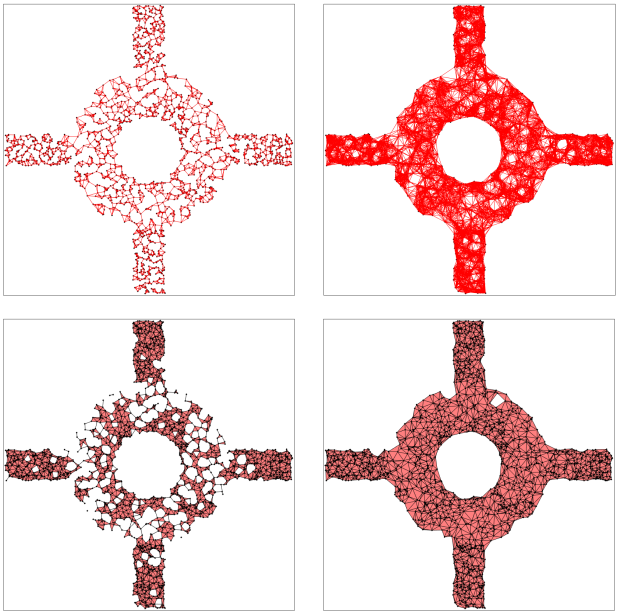

FIGURE 2.4. k-nearest neighbor graphs with $k = 5$ and $k = 20$ (top). α-complexes with $\alpha = 0.05$ and $\alpha = 0.2$ (bottom). Observe that the α-complexes include the interiors of the triangles.

2.4.2. α-complex. Although graphs capture the essential information of the data, graphs are usually considered as the 1-skeleton of a simplicial complex meaning that a graph can not contain any object other than vertices and edges. In other words, a simplicial complex can be thought

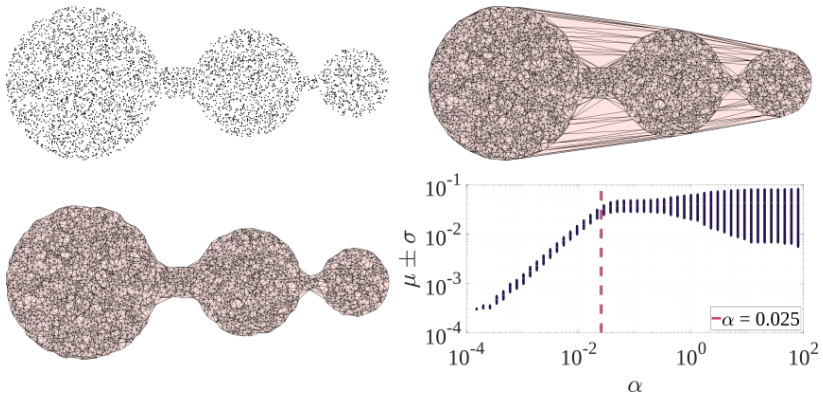

FIGURE 2.5. A schematic showing the construction of the α-complex. The point cloud in the top left is comprised of points that are uniformly sampled inside the three circles of radius 1, 0.7 and 0.45, and has the Delaunay triangulation on the top right. The proposed length scale analysis is on the bottom right, and the α-complex with $\alpha = 0.025$ is on the bottom left.

of as generalization of graphs that captures more information about the data by including higher dimensional simplices in the analysis. Among other simplicial complexes such as Vietoris-Rips complex [**126**] and Cech complex [**65**], this work considers the α-complex [**40**]. More formally, an α-complex $C_\alpha(X)$ is a subcomplex of a Delaunay triangulation $DT(X)$ of a point cloud X in R^d containing the simplices in $DT(X)$ whose circumradii is less than or equal to a given α.

The bottom row of Fig. 2.4 shows two α-complexes of the same point cloud with $\alpha = 0.05$ and $\alpha = 0.2$. The choice of α affects the structure of the complex and depends on the problem. While the standard approach in the literature involves persistent homology [**24**, **39**], a simple length scale analysis is likely sufficient for our purposes. Let μ be the average length of an edge in the α-complex and σ be the standard deviation of the edge lengths. When α is very small, the complex contains many disconnected components and poorly represents the geometry of the underlying manifold. For these values, μ increases with α as successively longer edges appear. When α is very large, the complex contains simplices that extend outside of the underlying manifold and does not adequately constrain paths through the space. For these values, inclusion of the longest edges increases σ substantially. There is generally an interval of intermediate α where μ is relatively stable and σ has not yet increased for which the complex is a good approximation to the underlying manifold. Figure 2.5 shows the procedure to construct the α-complex for an example point cloud along with the proposed length scale analysis.

The fundamental reason why α-complex is preferred over Vietoris-Rips complex or Cech complex is that it respects the dimensionality of the data. Unlike α-complexes, a typical Vietoris-Rips complex or a Cech complex might contain individual simplices whose dimensions are larger the dimension of the problem. Even though this is mathematically convenient, a physical justification is not so obvious. In an α-complex, the largest dimension of the simplices is bounded above by the dimension of the problem. Moreover, α-complex is stable to small perturbations meaning that a small change in the value of α does not dramatically affects the structure of the complex.

2.4.3. Filtrations of spaces. Given that our concern is with the geometry of the configuration space, it is convenient to construct a triangulation that represents this geometry explicitly. This is done by first constructing a triangulation of the configuration space of points $\Lambda(n)$ as a k-nearest neighbor graph or an α-complex. Each simplex of the triangulation is associated with a value of τ equal to the smallest value of any of the bounding vertices. A triangulation of the configuration space of hard disks $\Gamma(n, \rho)$ is then constructed as the subcomplex containing all simplices with associated values of τ greater than or equal to ρ, i.e., all triangles corresponding to locations of the sphere centers such that spheres of radius ρ do not overlap. This allows the geometry of $\Gamma(n, \rho)$ to be monitored for changes as a function of ρ. For brevity, $\Gamma(n, \rho)$ is said to be a filtration of $\Lambda(n)$ by τ.

Changing the filtration value not only affects the triangulations of the spaces, but their topological invariants as well. One of these invariants is the so-called merge tree associated to a function $f : M \to R$ [**100**]. Intuitively, a merge tree tracks the way components merge or separate as the sublevel (or superlevel) value changes. Formally, two points x and y are considered equivalent if they belong to the same level set $f(x) = f(y) = \ell$ and belong to the same component of the sublevel set $f^{-1}((-\infty, \ell])$; quotienting M by this equivalence relation gives the merge tree. There is a very similar object in the chemical physics literature known as a disconnectivity graph [**7**]. The merge tree can also be thought as a tree whose leaf nodes are minima (index-0), internal nodes are saddle points (index-1) and the root vertex is the global maximum [**100**].

As an example, Fig. 2.6 shows filtrations of the space in Fig. 2.5 by the height function with $-1 < y < 1$. For $y < -1$ the y-sublevel set is empty. As y is increased above -1 the bottom part of the largest disk appears first, with the bottom parts of the second and third disks following as y

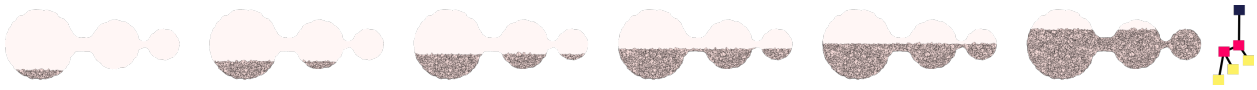

FIGURE 2.6. The evolution of the example α-complex by the height filtration. Observe that the topology only changes when the filtration reaches to the heights of the channels. The merge tree of the α-complex is shown on the right where the yellow squares are the local minima, the pink squares are saddles, and the dark blue is the global maximum.

increases further. When the y-sublevel set reaches the bottom of the first channel, a narrow channel appears that merges the first and second components. Further increasing y causes this channel to slowly grow, and when y reaches the bottom of the second channel all of the components merge into a single component. Higher values of y fill out the space further but no longer change the topology. The merge tree of the example α-complex is shown on the right in Fig. 2.6.

2.5. Diffusion distance

Conjecture 1.0.1 roughly suggests that a first order phase transition occurs when there is a singularity in the geometry of the configuration space. A filtration of the configuration space like the one discussed above allows us to track the topological and geometric properties of a configuration space as a function of disk radius. The geometric property considered in this study is the configuration space diameter measured by either the diffusion distance or ϵ-mixing time. Since ϵ-mixing time is discussed in Ch. 5 in great detail, this section provides the definition of the diffusion distance.

Given a weighted graph G, the diffusion distance requires the choice of a kernel k_{ij} that provides a notion of similarity between vertices i and j, with large values indicating that the points are close in the underlying manifold. The kernel should be (i) strictly positive such that $k_{ij} \geq 0$, and (ii) symmetric such that $k_{ij} = k_{ji}$. This work uses the standard Gaussian kernel $k_{ij} \propto \exp[-s_{ij}^2/(2\sigma^2)]$ where s_{ij} is the geodesic distance between vertices i and j and σ is a length scale representing the locality of the neighborhood; a small σ means that only the interactions between the closest neighbors significantly contribute. Unless otherwise specified, σ is equal to the radius of a ball with volume equal to 25% that of the filtered space.

The calculation of the diffusion distance begins with the $n \times n$ kernel matrix $\bar{\bar{K}}$, containing the kernel values k_{ij} for every pair of vertices i and j. The transition matrix $\bar{\bar{P}}$ is then defined as

$\bar{\bar{P}} = \bar{\bar{D}}^{-1}\bar{\bar{K}}$ where $\bar{\bar{D}}$ is a diagonal matrix with elements $d_{ii} = \sum_j k_{ij}$. Intuitively, $\bar{\bar{P}}$ contains the transition probabilities for a Markov chain process, and propagating this process forward by t steps is equivalent to raising $\bar{\bar{P}}$ to the t^{th} power. Rather than calculating the matrix power explicitly though, let λ_l and $\bar{\phi}_l$ be the eigenvalues and eigenvectors of $\bar{\bar{P}}$ for $l \geq 0$. Since $\bar{\bar{P}}$ is normalized, the largest eigenvalue λ_0 is always 1 with a constant eigenvector $\bar{\phi}_0$. Discarding the first eigenvalue and eigenvector, the diffusion coordinates are defined using $(n-1)$ of the eigenvalues and eigenvectors as

$$\bar{\Phi}_t(i) = \left[\lambda_1^t \bar{\phi}_1(i), \ldots, \lambda_{n-1}^t \bar{\phi}_{n-1}(i)\right].$$

Finally, the diffusion distance $d_{ij,t}^d$ at time step t is defined as

$$d_{ij,t}^d = \|\bar{\Phi}_t(i) - \bar{\Phi}_t(j)\| \tag{2.5}$$

where $\|\cdot\|$ is the Euclidean norm. The time step t is usually chosen here to make the product σt approximately equal to the length of the shortest path between the two most distant points in the complete space. Note that the significance of the time t depends on the diffusion rate implicit in the kernel $\bar{\bar{K}}$. Since the eigenspectrum of $\bar{\bar{P}}^t$ decays very quickly for sufficiently large t, the diffusion distance $d_{ij,t}^d$ can often be accurately approximated using only the first few eigenvalues and eigenvectors for a substantial time and memory savings [**118**].

The core of the diffusion distance lies in the calculation of the distance s_{ij} between vertices i and j. This work calculates s_{ij} using Dijsktra's algorithm when configuration spaces are represented as k-nearest neighbor graphs and using the algorithm proposed by Bhattacharya [**11**] when configuration spaces are represented as α-complexes. The essential difference is that a random walker on a graph is only allowed to travel along the edges of the graph, whereas a walker in a general simplicial complex is allowed to travel through the interiors of the simplices. The potential significance of this is clear if one considers a unit n-dimensional cube; a path between opposite corners along the cube edges is length n, whereas a path along the body diagonal is length $\sqrt{n}$. The algorithm proposed by Bhattacharya [**11**] calculates the geodesic distances between vertices along paths that are allowed to pass through the interiors of simplices. The algorithm is effectively a modification of Dijsktra's algorithm, adding an additional step that allows paths to pass through the interior of a simplex with only one vertex whose distance from a central vertex is unknown. Figure 2.7 shows

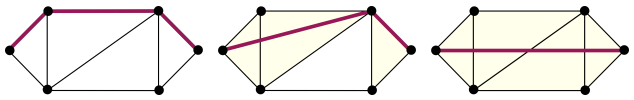

FIGURE 2.7. Example simplicial complexes (black points, black lines, and yellow triangles) and shortest paths (bold purple lines). The left complex consists of six 0-simplices and nine 1-simplices. The middle complex additionally has three 2-simplices (yellow triangles). The right complex has one more 2-simplex. The shortest path length between two vertices can decrease when higher-dimensional simplices are included.

several examples of the shortest paths between points as calculated by this method. On the left, the interiors of the triangles are not included and the simplicial complex is only comprised of vertices and edges. The distance between any two vertices can be found by Dijkstra's algorithm [**34**]. Triangle interiors shown in yellow are included in the simplicial complexes in the middle and the right, allowing shorter paths through the space.

Figure 2.8 shows the diameter of the spaces in Fig. 2.6 as a function of y for the diffusion distance. The similarity kernel $\bar{\bar{K}}$ is the standard Gaussian kernel with a standard deviation of 0.1. Observe that the three disconnected components appear as y is increased. The discontinuous jumps in the diffusion diameter correspond to the sublevel set values for which previously disconnected components of the space merge, making the accessible region of the space larger. These are precisely the locations where the topology of the space changes, i.e., the saddle points marked as pink squares in the merge tree on the top of Fig. 2.8. This provides preliminary evidence that the diffusion distance could be a reasonable choice for the distance in Conj. 1.0.1; changes in the value of the diffusion diameter appear to indicate corresponding changes in the topology and the extent of the accessible region, and therefore in the mixing time of a thermodynamic system. While the index-1 critical points (saddles) can be thought of as channels connecting index-0 critical points (minima), the effect of the higher index critical points is more complicated. Recalling the discussion of the distance between cube corners above, an index-k critical point could correspond to being able to pass through the diagonal of a k-dimensional cube, providing shortcuts between previously distant regions.

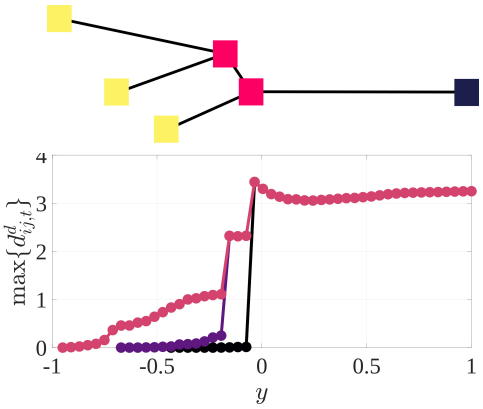

FIGURE 2.8. The diameter of the sublevel sets as measured the diffusion distance at $t = 50$ for the filtration in Fig. 2.6. Point color is associated with a particular component of the space. The discontinuities in the diameter coincide with the merging of components.

CHAPTER 3

Configuration Spaces of Hard Disks*

This chapter studies the definitions of quotient maps and quotient spaces for the configuration spaces of hard disks. Configuration spaces of hard disks have been studied before [**6**, **25**]. Ritchey [**113**] specifically studied the configuration spaces of hard disks on the hexagonal torus. They precisely defined the critical points and the associated critical indices on the configuration space of hard disks in the context of Morse theory (previously explained in Secs. 2.1 and 2.2), and considered the equivalence classes critical points generated by translations, permutations and discrete lattice symmetries. A high density of critical points around the packing fraction of the solid-liquid transition indicated a rapidly-changing topology of the potential energy surface there. This is suggestive of the 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. One area that they did not extensively explore is the effect that quotienting the configuration space by the symmetry groups that they identified has on the number and distribution of the critical points.

As far as the author know, explicit triangulations of the configuration spaces of hard disks, quotiented by symmetry groups or otherwise, have not been generated before. One of our purposes 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.

*The content of this chapter has previously been published in Ref. [**45**].

More specifically, this chapter 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 of the hard disk potential. Constructing explicit triangulations of the configuration space and the various quotient spaces is not trivial even for two disks, and is sufficient to identify 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 [**40**], and the isometric feature mapping (ISOMAP) algorithm [**119**] is used for dimensionality reduction.

Section 3.1 presents the critical points found after millions different initialization of the procedure described in Sec. 2.2. Section 3.2 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 3.3 defines a procedure to map a hard disk configuration into a space with coordinates that are invariant to the desired symmetry groups. Section 3.4 presents and discusses the explicit triangulations of the quotient spaces as a function of disk radius. Finally, Secs. 3.5 and 3.6 respectively present the proofs that the proposed metric and descriptors satisfy certain mathematical relations.

3.1. Critical configurations

This chapter considers the configuration space of hard disks in the square and hexagonal toruses shown in Fig. 2.1; 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. 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 inherent features of the thermodynamic system.

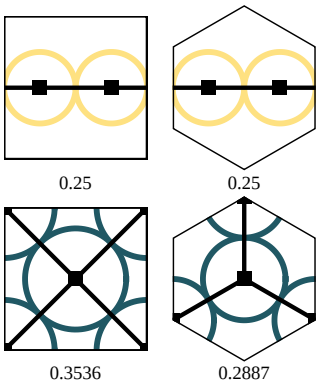

FIGURE 3.1. Representatives of the equivalence classes of critical points for two disks on the square and hexagonal toruses. The bottom (top) row corresponds to index-0 (index-1) critical points. The disk radius is reported below each configuration.

Representatives of the equivalence classes of the critical points in both domains found after millions of initializations for $n = 2$ are shown in Fig. 3.1. These are the critical configurations that changes the topology of the configuration space. Ritchey [**113**] 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.

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

$$d_\Lambda(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} \|\bar{p}_i - \bar{q}_i\| \tag{3.1}$$

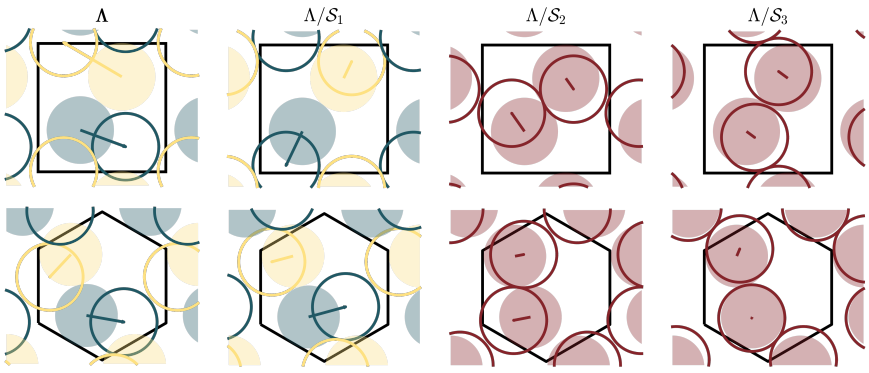

FIGURE 3.2. Distances between two configurations in the square and hexagonal toruses. The two configurations are indicated by filled and empty circles, and colors indicate the labeling of the disks. Table 3.1 lists the symmetry groups used to construct the quotient spaces.

where $\|\bar{p}_i - \bar{q}_i\|$ is the geodesic distance between the two positions of the ith disk. Figure 3.2 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 (identity of indiscernibles, symmetry, and the triangle inequality) with proofs provided in Sec. 3.5.

The configuration space of points Λ equipped with the metric d_Λ therefore constitutes a metric space (Λ, d_Λ). Given a metric space and an equivalence relation $\sim$ (usually deriving from a symmetry group), there is a natural induced metric $d_{\Lambda/\sim}$ on the quotient space $\Lambda/\sim$ [**22**]. When the equivalence relation additionally derives from a group of isometries $\mathcal{S}$, then the metric $d_{\Lambda/\mathcal{S}}$ on the quotient space $\Lambda/\mathcal{S}$ can be written as

$$d_{\Lambda/\mathcal{S}}(\mathbf{p}, \mathbf{q}) = \inf_{S \in \mathcal{S}} \{d_\Lambda[\mathbf{p}, S(\mathbf{q})]\} \tag{3.2}$$

where $\inf(\cdot)$ indicates the infimum. Along with Eq. 3.1, 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. As already discussed in Sec. 2.3, the operations belonging to all of these groups are isometric as is required to use Eq. 3.2. There

TABLE 3.1. Isometric symmetry groups applied to the configuration space, and the corresponding distances between the configurations in Fig. 3.2. $\mathcal{T}$, $\mathcal{P}$, $\mathcal{I}$ and $\mathcal{L}$ are the groups of translations, permutations, the inversion, and symmetries of the tiling.

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

are three quotient spaces considered in this chapter, all derived from the base configuration space of points Λ. The first $\Lambda/\mathcal{T}$ quotients out translation symmetries induced by the periodic boundary conditions and the homogeneity of space, and is conceptually derived by fixing the first disk at the origin.The second $\Lambda/\{\mathcal{T}\cup\mathcal{P}\cup\mathcal{I}\}$ also quotients out the inversion about the origin and permutations of the disk labels. The third $\Lambda/\{\mathcal{T}\cup\mathcal{P}\cup\mathcal{I}\cup\mathcal{L}\}$ also quotients out the discrete symmetries induced by the choice of the domain geometry. For simplicity of notation, Tab. 3.1 indicates the use of the symbols $\mathcal{S}_i$ to represent the symmetry groups by which the base configuration space Λ is quotiented. Table 3.1 also shows the distances between the configurations shown in Fig. 3.2 in the base configuration space and in the three quotient space considered.

Practically, the distances $d_{\Lambda/S}$ are computed by fixing the first configuration and generating all copies of the second configuration that only differ by the action of the discrete symmetry elements $\mathcal{S}/\mathcal{T}$. 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 [**27**, **58**]. $d_{\Lambda/\mathcal{S}}$ is reported as the minimum of these distances for all $S \in \mathcal{S}/\mathcal{T}$.

The left column of Fig. 3.2 and the first row of Tab. 3.1 show the distance between configurations in the base configuration space Λ. The distance in $\Lambda/\mathcal{S}_1$ 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 $\Lambda/\mathcal{S}_2$ is additionally minimized over permutations of the disk labels (indicated by the uniform disk color) and inversion about the origin. The distance in $\Lambda/\mathcal{S}_3$ is additionally minimized over the symmetries of the tiling, i.e., the symmetries of the square and hexagon. Observe that the distance between two configurations cannot increase (and generally decreases) as more symmetries are included.

3.3. Descriptors

As stated previously, the configuration space in Eq. 2.3 contains redundant information. Specifically, every configuration is equivalent to multiple other configurations related by the symmetry operations discussed by Ritchey [**113**]. Quotienting the space by these symmetry operations not only removes the redundancy, but usually gives a quotient space that 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 points is the product space $T^2 \times T^2$. Fixing the first point at the origin effectively quotients the space by the translation group, making $\Lambda/\mathcal{S}_1$ 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

$$x' = (R + r\cos\theta)\cos\phi$$

$$y' = (R + r\cos\theta)\sin\phi$$

$$z' = r\sin\theta$$

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

All other quotient spaces are initially embedded in an infinite-dimensional space resembling the space of Fourier coefficients, and a numerical approach is used to estimate the minimum number of descriptors necessary to maintain the embedding. Given a configuration of n points (disk centers), a distribution f is defined as a sum of Dirac-delta distributions $\delta(\bar{a}_j)$ located at the points $\bar{a}_j$ in the a_1a_2-coordinate system in Fig. 2.1. This distribution is expanded in a Fourier series, or

$$f(\bar{a}) = \sum_{j=1}^{n} \delta(\bar{a}_j) = \sum_{\bar{k}} c_{\bar{k}} \exp\left(2\pi i \bar{k} \cdot \bar{a}\right) \tag{3.3}$$

where $c_{\bar{k}}$ are the complex coefficients of the 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

$$c_{\bar{k}} = \sum_{j=1}^{n} \exp\left(-2\pi i \bar{k} \cdot \bar{a}_j\right). \tag{3.4}$$

The $c_{\bar{k}}$ respect the periodicity of the lattice and are invariant to permutations of the disk labels due to the commutative property of the summation in Eq. 3.3. 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

$$z_{\bar{k}} = \sqrt{c_{\bar{k}}^* c_{\bar{k}}} \tag{3.5}$$

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. 3.12 in Sec. 3.6. 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 $\Lambda/\mathcal{S}_2$.

Constructing an embedding of $\Lambda/\mathcal{S}_3$ further requires the descriptors to be invariant to the symmetries of the plane tiling. This is done explicitly by averaging the resulting descriptors, or

$$\hat{z}_{\bar{k}} = \frac{1}{O(\mathcal{L})} \sum_{L \in \mathcal{L}} z_{\bar{k}}^L \tag{3.6}$$

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.

Section 3.6 provides a proof that not all of these descriptors are independent. The invariance of the $z_{\bar{k}}$ 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 $\hat{z}_{\bar{k}}$ to the symmetries of the plane tiling results in more complicated relationships that are fully described in Sec. 3.6. 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 3.3 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 (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. 3.2 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 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

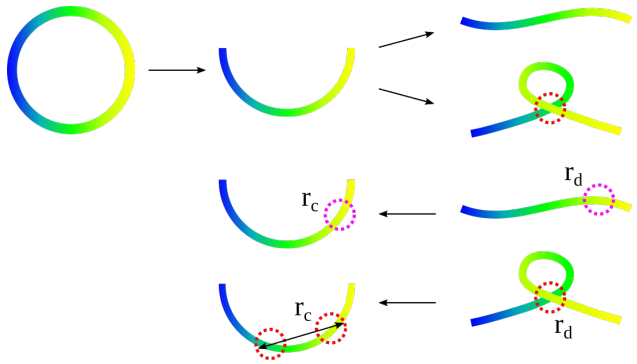

FIGURE 3.3. An illustration of the self-intersection search. The full circle on the left represents the base configuration space, with points related by a symmetry operation in the same color. The middle half-circle represents the space quotiented by the symmetry group, and on the right are possible 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.

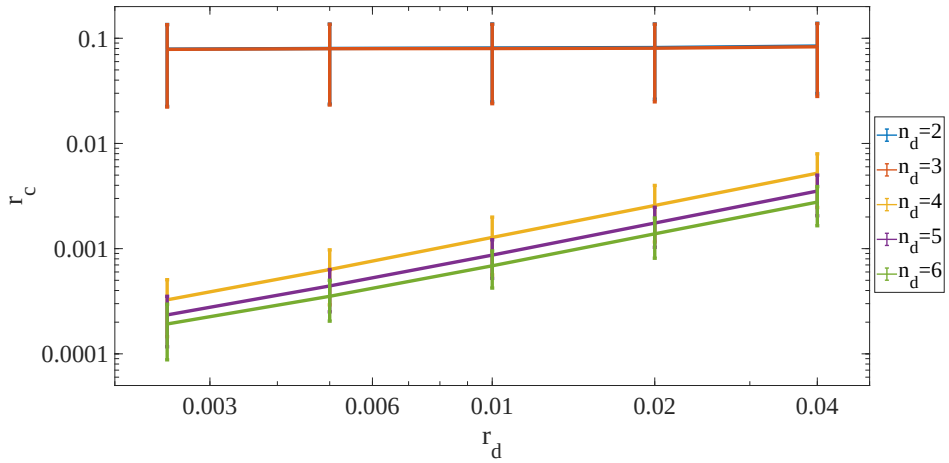

FIGURE 3.4. The inverse analysis for $n_d = 2 \ldots 6$ with different r_d values. The mean and standard deviation of r_c approach zero as r_d decreases for $n_d \geq 4$, suggesting that $n_d = 4$ is sufficient to embed the quotient space $\Lambda/\mathcal{S}_2$.

of the quotient space is likely self-intersecting as shown in Fig. 3.3, n_d is increased by one, and the process is repeated. Figure 3.4 shows the results of this analysis for the quotient space $\Lambda/\mathcal{S}_2$ and $n_d = 2 \ldots 6$. 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 $n_d \leq 3$. It can be concluded that four descriptors are sufficient to embed the quotient space $\Lambda/\mathcal{S}_2$.

3.4. 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 [**81**]. Repeated sampling of the Jacobian matrix for the quotient space $\Lambda/\mathcal{S}_2$ 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. 3.4 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 [**119**] 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. As briefly discussed in Sec. 2.4, the configuration spaces of hard disks are geometrically represented as α-complexes in this chapter. 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 that is previously described in Sec. 2.4.2. Let μ and σ respectively be the mean and standard deviation of the edge lengths. Figure 3.5 shows the result of this length scale analysis for the quotient space $\Lambda/\mathcal{S}_1$, 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 [**115**]. This study always samples $n_p = 10^4$ points, giving $n_t \approx 10^6$ tetrahedra in the full Delaunay triangulation of a 2-manifold embedded in R^3. Assuming that the volume of the convex hull of $\Lambda/\mathcal{S}_1$ for two disks is cove7red by uniform equilateral tetrahedra would give $\alpha_e = 2^{1/6}(6V/n_t)^{1/3}$ for the tetrahedral edge length where V is the manifold's volume. Since the space for $\Lambda/\mathcal{S}_1$ is constructed using the rectangle $[0, 1] \times [0, 2]$, the lower bound is $\alpha_e = 0.0111$. As seen in Fig. 3.5, this estimate is conservative.

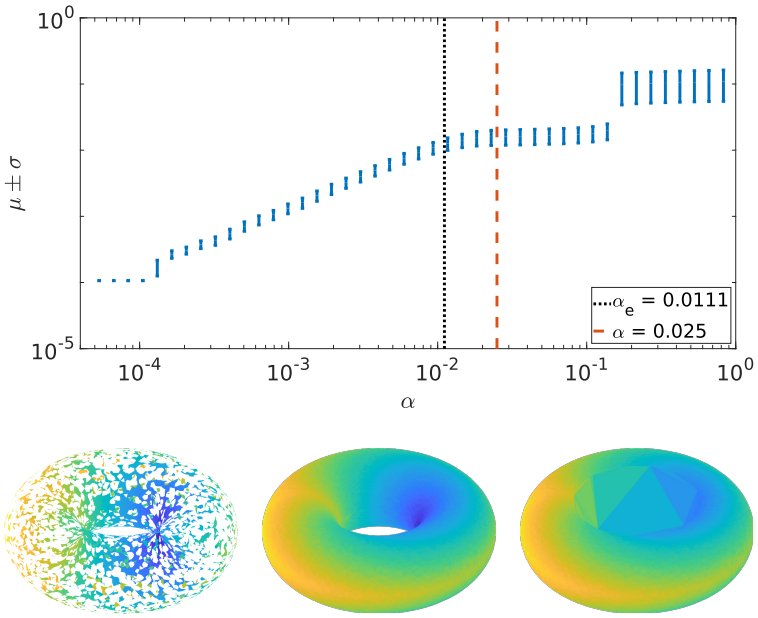

FIGURE 3.5. The length scale analysis for $\Lambda/\mathcal{S}_1$ for the square torus. μ and σ denote the mean and standard deviation of the edge lengths of the α-complex. The black-dotted line shows the lower bound estimate α_e, whereas the red-dashed line shows the α value actually used to construct the α-complex. α-complexes for increasing α values $\{0.0001, 0.025, 0.5\}$ are shown at the bottom.

3.4.1. 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. 3.6. 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 $\Lambda/\mathcal{S}_1$. 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 disk labels. That is, the submanifold that is contracted to a point 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 $\Lambda/\mathcal{S}_1$ contain $n!$ elements, despite this being widely assumed (perhaps because each of these equivalence classes does contain $n!$ elements in Λ). Moreover, changing the

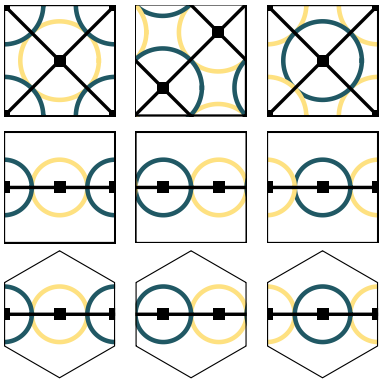

FIGURE 3.6. Critical points can be related by both translational and permutation symmetries. This happens for both critical points of the square, but only for the index-1 critical point of the hexagon.

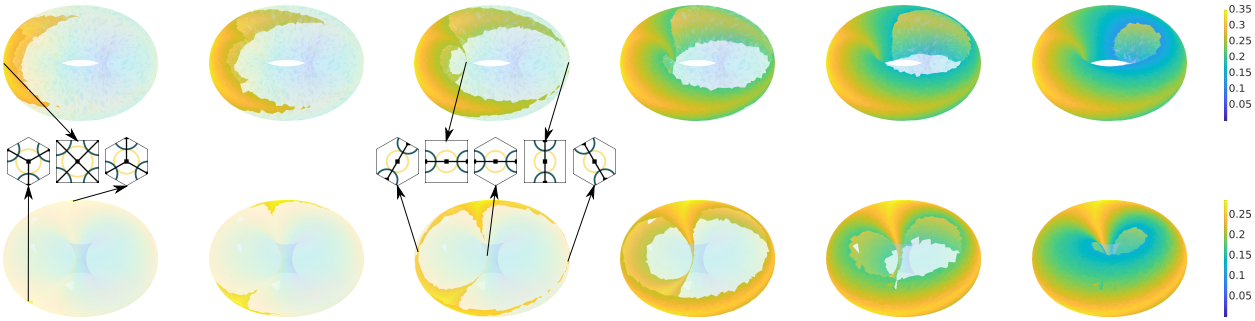

FIGURE 3.7. The evolution of the translation invariant configuration space, or the quotient space $\Gamma(2,\rho)/\mathcal{S}_1$, for the square torus (top) and for the hexagonal torus (bottom) with $\rho = \{0.28, 0.26, 0.25, 0.21, 0.17, 0.12\}$. The locations of the critical points in this space are indicated by arrows.

domain of an integral from $\Lambda/\mathcal{S}_1$ to $\Lambda/\mathcal{S}_2$ 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 [**61**, **107**].

Figure 3.7 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. 3.6 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 $\rho \to 0$ 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 3.7 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 $\Lambda/\mathcal{S}_1$ is that of a torus for both the square and the hexagon, as expected.

3.4.2. Adding permutation and inversion invariance. The descriptors $z_{\bar{k}}$ defined in Sec. 3.3 are by construction invariant to rigid translations, inversions about the origin, and permutations of disk labels. One way to construct the quotient space $\Lambda/\mathcal{S}_2$ is then to use the $z_{\bar{k}}$ as coordinates. Figure 3.4 suggests that four of these are sufficient for a proper embedding of $\Lambda/\mathcal{S}_2$. 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. 3.8 and Fig. 3.9 use the Euclidean distance using the corresponding descriptors, whereas the bottom rows use the distance defined in Eq. 3.2. 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. 3.2 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. 3.1 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 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

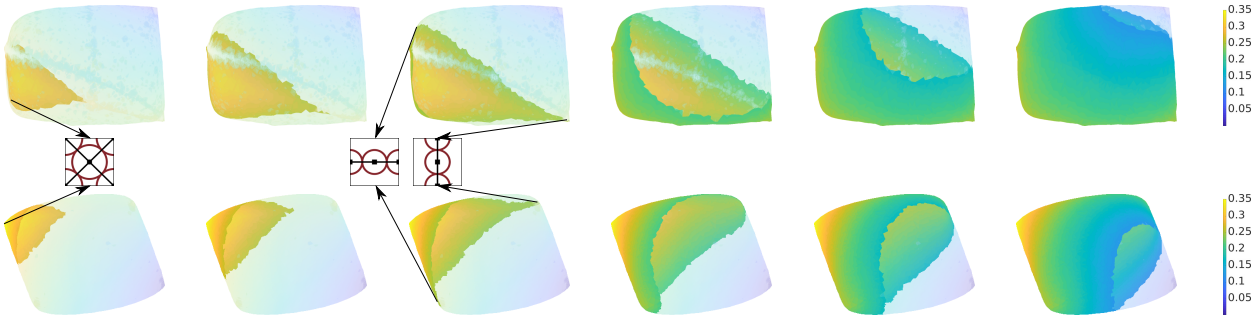

FIGURE 3.8. The evolution of the translation, permutation and inversion invariant configuration space, or the quotient space $\Gamma(2,\rho)/\mathcal{S}_2$, for the square torus constructed with the standard Euclidean distance (top) and the distance in Eq. 3.2 (bottom) with $\rho = \{0.28, 0.26, 0.25, 0.21, 0.17, 0.12\}$. The locations of the critical points are indicated by arrows.

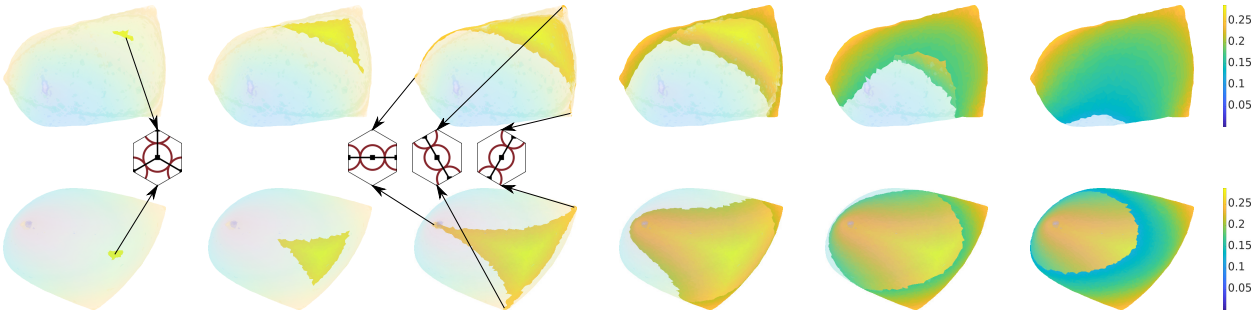

FIGURE 3.9. The evolution of the translation, permutation and inversion invariant configuration space, or the quotient space $\Gamma(2,\rho)/\mathcal{S}_2$, for the hexagonal torus constructed with the standard Euclidean distance (top) and the distance in Eq. 3.2 (bottom) with $\rho = \{0.28, 0.26, 0.25, 0.21, 0.17, 0.12\}$. The locations of the critical points are indicated by arrows.

the space as measured by the diffusion distance [**29**]. As ρ is further decreased, the spaces continue to grow and eventually close up, indicating that the topology of the quotient space $\Lambda/\mathcal{S}_2$ 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.

3.4.3. Adding lattice invariance. The descriptors $\hat{z}_{\bar{k}}$ defined in Sec. 3.3 are additionally invariant to the symmetries of the plane tiling, and are used as coordinates for the embedding of the quotient space $\Lambda/\mathcal{S}_3$. As before, dimensionality reduction is performed with the ISOMAP algorithm. The top rows of Fig. 3.10 and Fig. 3.11 use the Euclidean distance among the descriptors, whereas

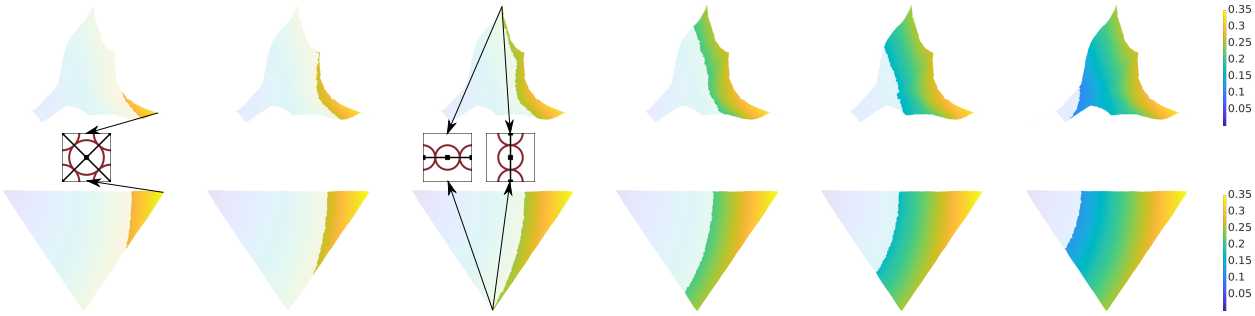

FIGURE 3.10. The evolution of the translation, permutation, inversion and lattice symmetry invariant configuration space, or the quotient space $\Gamma(2,\rho)/\mathcal{S}_3$, for the square torus constructed with the standard Euclidean distance (top) and with the distance in Eq. 3.2 (bottom) with $\rho = \{0.28, 0.26, 0.25, 0.21, 0.17, 0.12\}$. The locations of the critical points are indicated by arrows.

the bottom rows use the distance defined in Eq. 3.2. The two versions of $\Gamma(2,\rho)/\mathcal{S}_3$ are topologically identical as before. That said, the one using Eq. 3.2 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. 3.8 and Fig. 3.9. 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. 3.2 are the smallest symmetric regions of their corresponding domains; the bottom row of Fig. 3.10 is 1/8 of the square torus, whereas that of Fig. 3.11 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, $\Lambda/\mathcal{S}_3$ has a boundary and is topologically equivalent to a disk, in contrast to $\Lambda/\mathcal{S}_2$ having the topology of a sphere and $\Lambda/\mathcal{S}_1$ that of a torus.

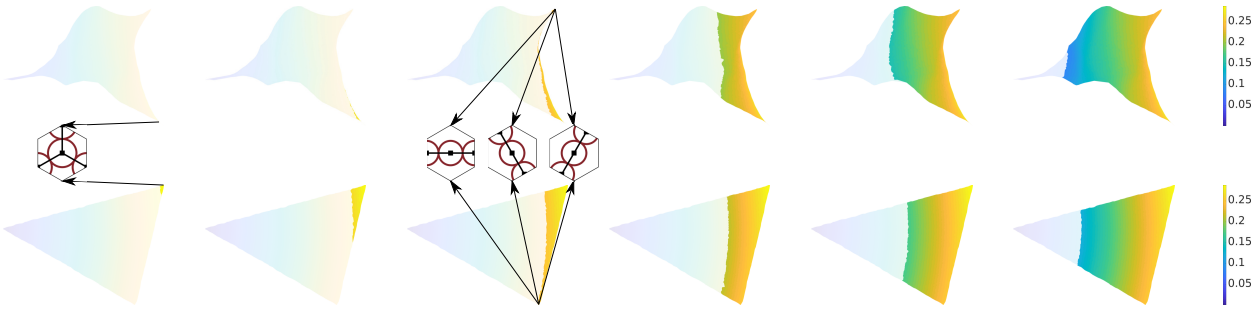

FIGURE 3.11. The evolution of the translation, permutation, inversion and lattice symmetry invariant configuration space, or the quotient space $\Gamma(2,\rho)/\mathcal{S}_3$, for the hexagonal torus constructed with the standard Euclidean distance (top) and with the distance in Eq. 3.2 (bottom) with $\rho = \{0.28, 0.26, 0.25, 0.21, 0.17, 0.12\}$. The locations of the critical points are indicated by arrows.

3.5. Proof that d_Λ is a metric

This section proves that the distance function in Eq. 3.1 is a metric. Let $\bar{x}$ be a column vector in R^2 and $\bar{\bar{P}}$ be a projection matrix whose rows contain the unit vectors of the square or hexagonal domain as in Fig. 2.1. 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 $\bar{w} = \bar{\bar{P}}\bar{x}$. 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 ith disk, and drop the index in the following.

For the square torus in Fig. 2.1, the geodesic distance reduces to

$$\|\bar{p} - \bar{q}\| = \sqrt{a^2 + b^2}$$

$$a = \min\left(|p_x - q_x|, 1 - |p_x - q_x|\right)$$

$$b = \min\left(|p_y - q_y|, 1 - |p_y - q_y|\right)$$

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 $p_x = q_x$.

TABLE 3.2. 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	$p_x + 1/2$	$p_y + \sqrt{3}/2$	$t_1 + 1/2$	$t_2 + 1$	$t_3 + 1/2$

$b = 0$ 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 $d_\Lambda(\mathbf{p}, \mathbf{q}) = 0 \Longrightarrow \mathbf{p} = \mathbf{q}$.

For the hexagonal torus in Fig. 2.1, $\bar{\bar{P}}_h = [1, 0; 1/2, \sqrt{3}/2; -1/2, \sqrt{3}/2]$ and the geodesic distance reduces to

$$\|\bar{p} - \bar{q}\| = \{\min[a^2 + b^2, (1-a)^2 + b^2, (a - 0.5)^2 + (b - \sqrt{3}/2)^2]\}^{1/2}$$

$$a = |p_x - q_x|$$

$$b = |p_y - q_y|.$$

Let $\bar{\bar{P}}_h \bar{p} = \bar{t}$ and $\bar{\bar{P}}_h \bar{q} = \bar{w}$. 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. 3.2. 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 $d_\Lambda(\mathbf{p}, \mathbf{q}) = 0 \Longrightarrow \mathbf{p} = \mathbf{q}$.

□

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 $d_\Lambda(\mathbf{p}, \mathbf{r}) \leq d_\Lambda(\mathbf{p}, \mathbf{q}) + d_\Lambda(\mathbf{q}, \mathbf{r})$.

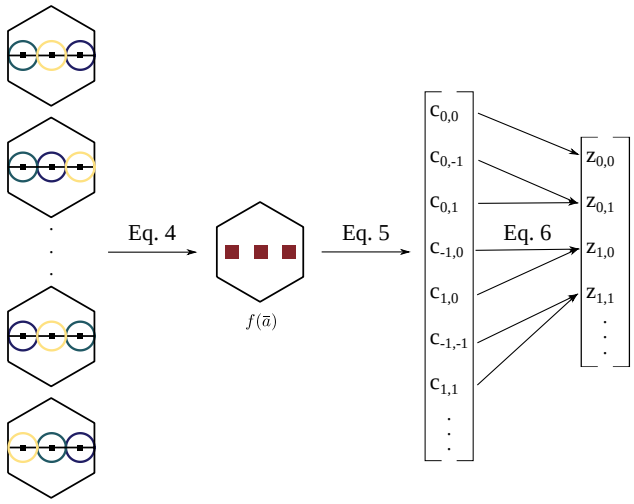

FIGURE 3.12. The construction of the $z_{\bar{k}}$. Permuting the disk labels does not change the function $f(\bar{a})$ in Eq. 3.3. The Fourier coefficients $c_{\bar{k}}$ are calculated with Eq. 3.4. The descriptors $z_{\bar{k}}$ are defined as the magnitude of the $c_{\bar{k}}$ as in Eq. 3.5.

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 ith term in each of the sums, and that this is true since $\|\cdot\|$ is the geodesic distance. □

3.6. Relations between the descriptors

This section provides proofs and supporting material relating to the descriptors $z_{\bar{k}}$ and their properties. The procedure to calculate the $z_{\bar{k}}$ is shown in Fig. 3.12. For a disk configuration $\mathbf{p}$ where the coordinates of the ith disk center $\bar{p}_i$ are given in the a_1a_2-coordinate system, the first step is to define the function $f(\bar{a})$ using Eq. 3.3. By construction, the result is invariant to the choice of the $n!$ permutations of the disk labels. The Fourier coefficients $c_{\bar{k}}$ of $f(\bar{a})$ are calculated using Eqs. 3.3 and Eq. 3.4, but change phase when the underlying disk configuration is translated. This is the motivation for taking the modulus of the $c_{\bar{k}}$ using Eq. 3.5, since the resulting $z_{\bar{k}}$ are invariant to translations.

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

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

□

Since inverting the underlying disk configuration has the effect of taking the complex conjugate of the $c_{\bar{k}}$, constructing the $z_{\bar{k}}$ as in Eq. 3.5 additionally makes them invariant to inversions.

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

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

□

The final step is to sort the $z_{\bar{k}}$ with respect to the distance of $\bar{k}$ to the origin and to remove any redundant values. For example, the Fourier coefficients $c_{\bar{k}}$ and $c_{-\bar{k}}$ have the same magnitudes $z_{\bar{k}}$ for any configuration.

PROOF. Let $\mathbf{p}$ be any configuration.

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

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

□

Other redundancies are introduced by the symmetries $\mathcal{L}$ of the domain. While it is straightforward to write down the action of a particular symmetry operator $L \in \mathcal{L}$ as a matrix $\bar{\bar{L}}$ that acts on a vector of coordinates in the xy-coordinate system, the calculation of the $z_{\bar{k}}$ is naturally done in the a_1a_2-coordinate system. For this reason, it is useful to be able to convert vectors of coordinates or transformation matrices written using one coordinate system to the other. This is accomplished by means of a coordinate transformation matrix $\bar{\bar{T}}$ such that $\bar{a} = \bar{\bar{T}}\bar{x}$ and $\bar{x} = \bar{\bar{T}}^{-1}\bar{a}$, where $\bar{x}$ and $\bar{a}$ are vectors of coordinates of the same point in the xy- and a_1a_2-coordinate systems. This also allows the matrix $\bar{\bar{L}}$ to be converted into a matrix $\bar{\bar{U}} = \bar{\bar{T}}\bar{\bar{L}}\bar{\bar{T}}^{-1}$ that performs the equivalent action in the a_1a_2-coordinate system. This can be seen by the following calculation, where $\bar{x}'$ and $\bar{a}'$ are vectors of coordinates of a point related by the symmetry operator L to $\bar{x}$ and $\bar{a}$:

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

Given the ability to calculate the matrix $\bar{\bar{U}}$ for any $L \in \mathcal{L}$, the calculation of the redundant $z_{\bar{k}}$ specifically for the square and hexagonal domains proceeds as below.

For the square torus, $\bar{\bar{T}} = [1, 0; 0, 1]$ and $\mathcal{L}$ is the dihedral symmetry group of order eight (D_4). Equation 3.6 is

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

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

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

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

$$[p, q] \sim [-p, q] \sim [q, -p] \sim [-q, -p] \sim [-p, -q] \sim [p, -q] \sim [-q, p] \sim [q, p].$$

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

$$[p, q] \sim [q, q-p] \sim [q-p, -p] \sim [-p, -q] \sim [-q, p-q] \sim [p-q, p] \sim [-p, q-p] \sim [-q, -p] \sim$$
$$[p-q, -q] \sim [p, p-q] \sim [q, p] \sim [q-p, q].$$

CHAPTER 4

Configuration Spaces of Hard Spheres*

The purpose of this chapter is to develop a procedure to practically test Conj. 1.0.1 for hard sphere systems. By hypothesis, the mixing time of a thermodynamic system depends on the connectivity of the space and the volumes of the accessible regions. This suggests that changes to the accessible domain and the expected time to explore the domain be analyzed by a distance function that is sensitive to variations in the length and number of paths connecting pairs of points. Although there are other distances defined on graphs with the desired properties like the commute time distance [**51**, **127**], this chapter considers the diffusion distance [**29**, **118**]. Given a weighted graph G, the diffusion distance roughly measures the difference in concentration profiles of a species diffusing on the manifold when starting from Dirac delta functions at vertices i and j. It decreases with decreasing length of paths and increasing number of paths connecting vertices i and j. A formal definition of the diffusion distance with a toy example is already provided in Sec. 2.5.

The diameter of a space is defined as the diffusion distance between the two most distant points, and is expected to increase with the mixing time of a thermodynamic system. The diffusion distance is closely related to the mixing time, roughly defined as the time required for the probability distribution of microstates to converge to the Boltzmann distribution (setting aside Loschmidt's paradox) and for the preparation protocol to be forgotten. As such, any discontinuities in the mixing time as a function of ρ should coincide with discontinuities in the diameter of $\Gamma(n, \rho)$ as measured by the diffusion distances. While it requires the space to have the property that a path connects each pair of points, the distances can still be calculated separately for each component of a disconnected space.

Discontinuous jumps in the diameter are expected to occur when two or more previously disconnected components merge, and therefore to be associated with one or more critical configurations.

*The content of this chapter has previously been published in Ref. [**48**], and is used here with the permission of American Physical Society.

This motivates first identifying critical points of the configuration spaces for $n = 1 \ldots 12$ hard spheres to predict the radius intervals where geometric changes will occur. Section 4.1 describes the known critical configurations of systems with $n = 1 \ldots 12$ hard spheres on a rhombic dodecahedron shown in Fig. 2.1 with periodic boundary conditions, and the relevance of the distribution of critical points to the phase coexistence interval. After analyzing the properties of these critical points, explicit triangulations for $n = 2$ spheres are constructed to verify that the appearance of critical points can substantially affect the geometric properties (i.e., the diameter) in the expected way. If so, then it is reasonable to conjecture that the same could happen for systems with $n > 2$ spheres, or even in the thermodynamic limit. For this purpose, the distances and descriptors proposed in Ch. 3 have been extended to the hard sphere systems. Section 4.3 gives approximate triangulations of the configuration spaces for small numbers of spheres, and explores the relationship of the critical configurations to the space geometry. Finally, Section 4.4 reviews these results in the context of Conj. 1.0.1, particularly for the configuration spaces of two hard spheres.

4.1. Critical configurations

As explained in Sec. 2.2, the critical points of τ are identified with those of E in the limit of large w. The critical points of E are defined by the condition $\nabla E = 0$, and are therefore minima of the nonnegative function $|\nabla E|^2$. The locations of the critical points can therefore be sampled by repeatedly initializing a search with a random initial configuration and using any standard minimization algorithm (e.g., conjugate gradient) with $|\nabla E|^2$ as the objective function. While this does not provide an exhaustive enumeration of critical points, the results below show that tens of thousands of distinct critical points have already been identified in this way. It should be mentioned that there is an additional complication when $n < 9$, since for small values of n it is possible for a sphere to be in contact with multiple periodic images of another sphere. The consequence of this is that all of the spheres in all 19 images of the fundamental rhombic dodecahedron shown in Fig. 2.1 need to be considered in such situations. Where $n \geq 9$, multiple contacts are disallowed geometrically and only the spheres in the fundamental cell need to be considered.

Ritchey [**113**] suggested that the critical points could be grouped into equivalence classes related by symmetry operations, including rigid translations, permutations of the sphere labels, and

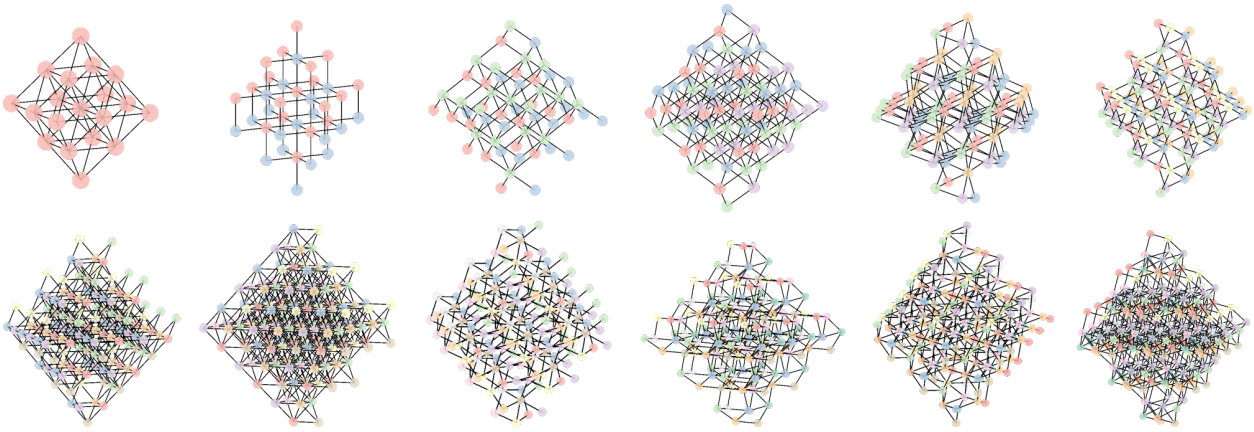

FIGURE 4.1. The densest configurations found for $n = 1 \ldots 12$. The rest of the critical configurations can be visualized using the interactive database described in the text. Some of the critical configurations correspond to the well-known crystal structures, e.g., the first and the eighth are face-centered cubic, the second is simple cubic and the fourth is body-centered cubic.

symmetries of the tessellation. An interactive database of all known critical points has been created for $n = 1 \ldots 12$ spheres in the RD using a Colab Notebook and is freely available online[1]. The densest known configurations for $n = 1 \ldots 12$ are shown in Fig. 4.1.

A critical configuration is associated with a vector $\mathbf{x}$ of coordinates of the sphere centers, an adjacency matrix $\bar{\bar{A}}$, a sphere radius ρ, and a critical index p. For a given labeling of the spheres, $\bar{\bar{A}}$ is an integer matrix where the element a_{ij} is the number of contacts between spheres i and j; the main diagonal is always zero since a sphere cannot be in contact with itself (except for $n = 1$). A sphere is called a rattler if it is not in contact with any other spheres.

This description of a critical point is clearly not the same for all elements of a given equivalence class, e.g., a translation adds a structured vector to $\mathbf{x}$, and a permutation reorders the rows and columns of $\mathbf{x}$ and $\bar{\bar{A}}$. This is handled by using the program nauty [**95**] to construct a canonical labeling of the spheres that is the same for every configuration in a given equivalence class, and translating the first sphere to the origin. After applying the canonical labeling, one can decide whether a candidate critical point belongs to an equivalence class that has already been observed on the basis of $\bar{\bar{A}}$, ρ and p.

[1]The database of critical points is available at `https://github.com/burakericok/hard_sphere_crits`.

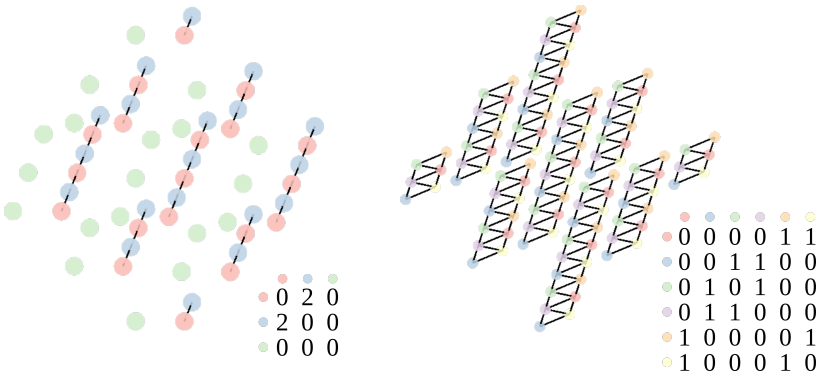

FIGURE 4.2. Examples of critical configurations for $n = 3$ (left) and $n = 6$ (right), with adjacency matrices indicated in the insets. The left one is a non-degenerate index-2 critical point with $\rho = 0.25$, whereas the right one is a degenerate index-8 critical point with $\rho = 0.167$. Observe that the edges between adjacent columns on the right are degenerate, and these edges are not reflected in the adjacency matrix.

Previously, Baryshnikov et al. [**6**] showed that the critical points of the configuration spaces of hard spheres are precisely the mechanically-balanced configurations. Let $\bar{e}_{ij}$ be the vector pointing from the center of sphere i to that of sphere j, and w_{ij} be the force along or weight on this edge of the contact graph. A configuration is said to be mechanically balanced if there is a nontrivial set of repulsive forces along the edges such that the sum of forces on every sphere vanishes; that is, if there is a nontrivial set of non-negative weights such that $\sum_{j\sim i} w_{ij}\bar{e}_{ij} = 0$ for every fixed vertex i, where the sum is performed over all vertices adjacent to i. A critical configuration is said to be non-degenerate if, for each edge, there exists at least one mechanically-balanced set of weights such that the weight on that edge is positive.

Figure 4.2 provides several example critical points to illustrate these definitions. The critical point on the left is a non-degenerate index-2 critical point with $n = 3$ spheres and $\rho = 0.25$, and the green sphere is a rattler since it has no contacts with the other spheres. The reason this is an index-2 critical point is that there are two independent directions that the blue sphere can be moved relative to the red one while increasing the radius to second order; any possible motions of the red sphere relative to the blue one can be made equivalent to these by a translation. The critical point on the right is an index-8 critical point with $n = 6$ spheres and $\rho = 0.1667$. The edges between two adjacent columns are all degenerate edges; it is not possible to assign positive weights to these edges without breaking mechanical balance. As Fig. 4.3 shows, the majority of the critical

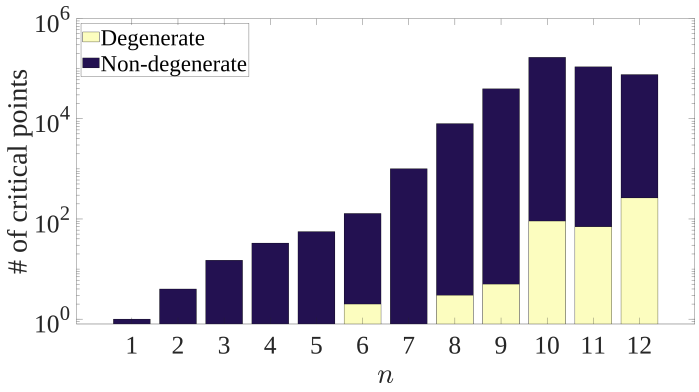

FIGURE 4.3. The number of equivalence classes of non-degenerate and degenerate critical points found by the algorithm described in Sec. 4.1 for ten million random initial configurations for $n = 1 \ldots 10$ and three million for $n = 11 \ldots 12$.

points already found are non-degenerate. It is also significant that, even though the algorithm used to generate Fig. 4.3 does not identify all critical points, the number of known critical points seems to increase exponentially in n on the basis of the well-sampled interval $n = 6 \ldots 10$.

If the distribution of critical points is indeed related to the hard sphere phase transition, then it is reasonable to suppose that there will be a significant number of critical points at densities close to the liquid and solid limits of the infinite system, even when the systems are finite. Figure 4.4 shows the distributions of the critical points for $n = 9 \ldots 12$ spheres as a function of their index and packing fraction η (the fraction of volume occupied by spheres). Almost all of the critical points occur around the coexistence interval in each case, with the lower index critical points slightly below the solid limit and the higher index critical points closer to the liquid limit. Moreover, the distributions of the critical points change with increasing n, becoming more concentrated and shifting closer to the solid side of the coexistence interval. While this could be a result of biased sampling of the critical points for 11 and 12 spheres, closer study of the distributions for 9 and 10 spheres suggests that the distribution of lower index critical points either remains centered in the middle of the coexistence interval or narrows and shifts toward the solid limit with increasing n. Further observe that the volume of the configuration space and the number of critical points both increase exponentially with n (Fig. 4.3), suggesting that the density of critical points in the configuration space does not change as rapidly as either one. If these two trends hold for all n, then any involvement of the critical point distribution in the appearance of a phase transition in

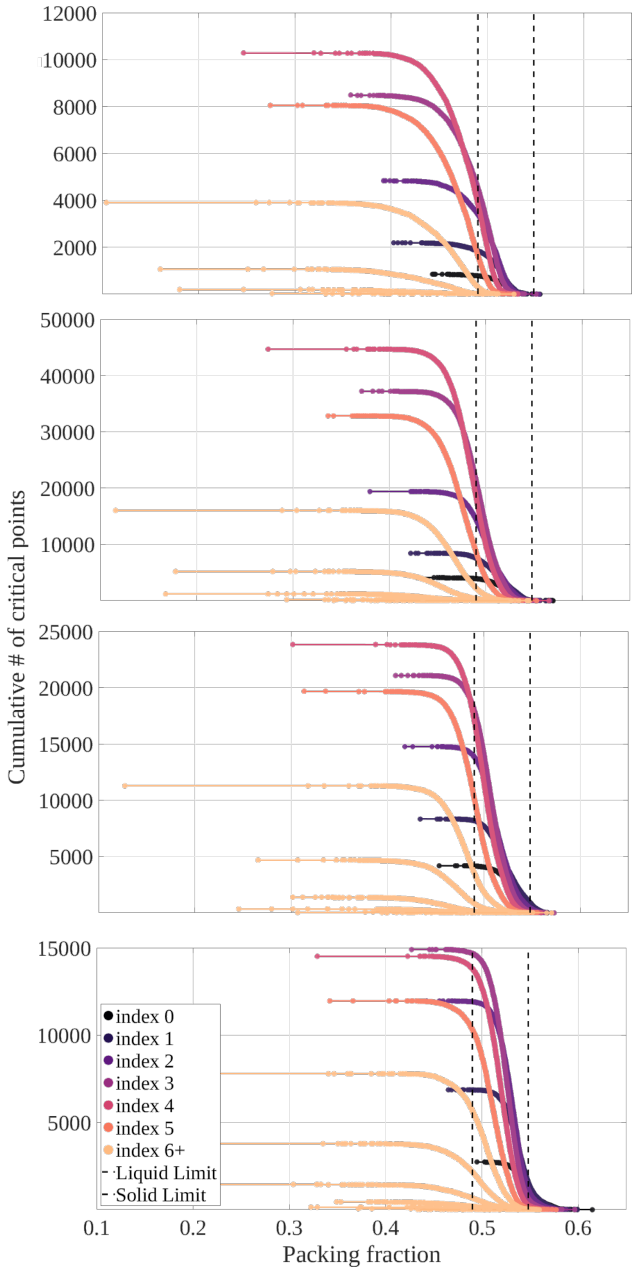

FIGURE 4.4. Distribution of equivalence classes of critical points for $n = 9 \ldots 12$ (from top to bottom) as a function of index and packing fraction. The critical points are sorted by packing fraction, and the cumulative distributions are constructed for decreasing packing fraction (i.e., right to left).

the thermodynamic limit would likely remain unchanged even for systems with small n. This is the line of reasoning that motivates the detailed study of the $n = 2$ sphere system in the following sections.

4.2. Descriptors

This section extends the definitions of the descriptors first proposed in Ref. [**45**] and presented in Sec. 3.3 to the hard sphere model. Consider n hard spheres in three dimensional torus $\mathbb{T}^3$. The derivation starts by defining a new density function $f(\bar{a})$ as the sum of Dirac-delta distributions located at the sphere centers $\bar{a}_j$ with respect to the $a_1a_2a_3$-coordinate system shown in Fig. 2.1, or

$$f(\bar{a}) = \sum_{j=1}^{n} \delta(\bar{a}_j). \tag{4.1}$$

Observe that the density function is invariant under the permutations of the sphere labels due to the commutative property of summation. Expanding $f(\bar{a})$ as a Fourier series yields

$$f(\bar{a}) = \sum_{\bar{k}} c_{\bar{k}} \mathrm{e}^{2\pi \mathrm{i} \bar{k} \cdot \bar{a}} \tag{4.2}$$

where $c_{\bar{k}}$ are the complex coefficients of the expansion and $\bar{k} = [u, v, w]$ for integers u, v and w. By orthogonality of the complex exponentials, the infinite set of coefficients can be calculated as

$$c_{\bar{k}} = \sum_{j=1}^{n} \mathrm{e}^{-2\pi \mathrm{i} \bar{k} \cdot \bar{a}_j}. \tag{4.3}$$

The complex coefficients respect the periodicity of the lattice. Moreover, translating the configurations as rigid bodies only changes the phase of the coefficients, keeping their moduli constant. Hence, a new set of real-valued descriptors that are additionally invariant under translations is defined as

$$z_{\bar{k}} = \sqrt{c_{\bar{k}}^* c_{\bar{k}}} \tag{4.4}$$

where * is the complex conjugate. It is easy to show that $z_{\bar{k}}$ is also invariant under the inversion symmetry. Finally, descriptors invariant under the lattice symmetries (octahedral symmetry O_h) can be defined as

$$\hat{z}_{\bar{k}} = \frac{1}{48} \sum_{t=1}^{48} z_{\bar{k}}^t \tag{4.5}$$

where $z_{\bar{k}}^t$ are the descriptors $z_{\bar{k}}$ calculated for the configuration generated by applying the tth symmetry element in O_h to the configuration.

Not all $\hat{z}_{\bar{k}}$ are independent. Let $\bar{x}$ and $\bar{a}$ denote the coordinates of a given point in the xyz- and $a_1a_2a_3$-coordinate systems, respectively. The relationship of the two coordinate systems is given by the transformation matrix $\bar{\bar{T}}$, i.e., $\bar{a} = \bar{\bar{T}}\bar{x}$ and $\bar{x} = \bar{\bar{T}}^{-1}\bar{a}$. Let $\bar{\bar{O}} \in O_h$ be a symmetry element written as a matrix in the xyz-coordinate system. A symmetric copy $\bar{x}'$ of a vector $\bar{x}$ can be written as $\bar{x}' = \bar{\bar{O}}\bar{x}$. Then,

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

where $\bar{\bar{U}} = \bar{\bar{T}}\bar{\bar{O}}\bar{\bar{T}}^{-1}$. Observe that $\bar{\bar{U}}$ corresponds to the symmetry matrix $\bar{\bar{O}}$ written in the $a_1a_2a_3$-coordinate system.

Suppose $z_{\bar{k}}^t$ are computed for the tth symmetry element in O_h. Then,

$$z_{\bar{k}}^t = \left\| \sum_{j=1}^{n} \exp\left(-2\pi \mathrm{i} \bar{k} \cdot \bar{\bar{U}}_t \bar{a}_j\right) \right\| = \left\| \sum_{j=1}^{n} \exp\left(-2\pi \mathrm{i} \bar{k}' \cdot \bar{a}_j\right) \right\|$$

where $\bar{k}' = \bar{k} \cdot \bar{\bar{U}}_t$. Computing $\bar{k}'$ for all elements of O_h yields the equivalent indices presented in Tab. 4.1.

4.3. Configuration Spaces

As previously discussed in Sec. 2.3, the configuration spaces of hard spheres defined in Eq. 2.3 has a variety of symmetries, or physical operations that make multiple points in Γ physically indistinguishable. The symmetry groups $\mathcal{S}_i$ and the quotient spaces $\Gamma/\mathcal{S}_i$ considered in this chapter are indicated in Tab. 4.2.

As in Ch. 3, there are two methods to sample points in the quotients of the configuration spaces of hard spheres by the desired symmetries. The first proposes an infinite set of real-valued descriptors defined in Sec. 4.2 that are invariant to the desired symmetries, mapping the full configuration space into an infinite-dimensional Euclidean space. The isometric feature embedding (ISOMAP) [**119**] algorithm is used to reduce the dimension of the image as much as possible while preserving the local geometry as defined by the edge lengths in the k-nearest neighbor graphs.

TABLE 4.1. The equivalent indices where k'_i denotes the components of $\bar{k}'$.

k'_1	k'_2	k'_3	k'_1	k'_2	k'_3
u	v	w	$w-v$	$u-w$	u
$u+v-w$	$-w$	$v-w$	$u-w$	$w-v$	$-v$
$-w$	$u+v-w$	$u-w$	w	$w-v-u$	$w-v$
u	$-v$	$u-w$	$w-v-u$	$-w$	$-v$
$w-u$	$w-v$	$w-v-u$	u	v	$u+v-w$
$w-v-u$	w	$w-v$	v	u	w
$u-w$	$v-w$	$-w$	w	$u+v-w$	u
$w-v$	$w-u$	w	$-u$	v	$w-u$
$-u$	$-v$	$-w$	$v-w$	$w-u$	$-u$
$-v$	$-u$	$w-v-u$	$w-u$	$v-w$	v
$-u$	v	$v-w$	$-w$	$u+v-w$	$v-w$
$w-v-u$	$-w$	$-u$	$u+v-w$	w	v
$u-w$	$v-w$	$u+v-w$	$-u$	$-v$	$w-v-u$
v	u	$u+v-w$	$-v$	$-u$	$-w$
$w-u$	$w-v$	w	$-w$	$w-v-u$	$-u$
$v-w$	$u-w$	$-w$	$u+v-w$	w	u
$w-v-u$	w	$w-u$	v	$-u$	$v-w$
w	$u+v-w$	v	$-v$	u	$u-w$
$w-v$	$u-w$	$-v$	$w-u$	$v-w$	$-u$
u	$-v$	$w-v$	$v-w$	$u-w$	$u+v-w$
$u+v-w$	$-w$	$u-w$	$u-w$	$w-v$	u
$-w$	$w-v-u$	$-v$	v	$-u$	$w-u$
$v-w$	$w-u$	v	$-v$	u	$w-v$
w	$w-v-u$	$w-u$	$w-v$	$w-u$	$w-v-u$

TABLE 4.2. The spaces considered in this work. $\mathcal{T}$, $\mathcal{P}$, $\mathcal{I}$ and $\mathcal{L}$ are the groups of translations, permutations, inversions, and proper symmetries of the tiling.

Space	Symmetries
$\Gamma/\mathcal{S}_1$	$\mathcal{T}$
$\Gamma/\mathcal{S}_2$	$\mathcal{T}\bigcup\mathcal{P}\bigcup\mathcal{I}\bigcup\mathcal{L}$

The second method uses the same distance functions Eqs. 3.1 and 3.2 with the appropriate extensions to hard sphere systems. Once again, the computation of $d_{\Lambda/\mathcal{S}}$ is formulated as a global optimization problem and solved by the tabu search algorithm [**27**,**58**]. After the pairwise distances $d_{\Lambda/\mathcal{S}}$ between the sampled configurations are computed, the ISOMAP algorithm is used to construct an approximate embedding into a Euclidean space.

Both methods initially sample configurations of n points in the rhombic dodecahedron (RD), map them into a quotient configuration space, use the appropriate distance function to metrize this

space, and triangulate the resulting point cloud as an α-complex [**40**]. While both methods give quotient spaces with the same topology, the second method better preserves the quotient spaces geometry.

Although the Eq. 3.2 can be extended to the configuration spaces of hard spheres without difficulty, the computational cost of the algorithm is an issue. The computation of $d_{\Lambda/\mathcal{S}}$ involves a global optimization problem, and constructing the pairwise distance matrix of m configurations for a naive implementation would require solving $\mathcal{O}(m^2)$ of these problems. Instead, the symmetry-invariant descriptors for configurations of hard spheres (derived in Sec. 4.2) are used in a preprocessing step to reduce the computational cost. Recall that the ISOMAP algorithm only requires the lengths of the edges of the k-nearest neighbor graphs as inputs. These graphs can be approximated using the distances between configurations as calculated by the descriptors, i.e., the k'-nearest neighbors found using the descriptors should include all of the k-nearest neighbors that would be found using $d_{\Lambda/\mathcal{S}}$ for sufficiently large $k' > k$. The pairwise distances $d_{\Lambda/\mathcal{S}}$ then only need to be calculated for the k'-nearest neighbors, reducing the computational cost to $\mathcal{O}(mk')$. Finally, the desired k-nearest neighbor graphs with edge lengths determined by $d_{\Lambda/\mathcal{S}}$ are extracted and supplied to the ISOMAP algorithm.

One of the weaknesses of the above approach is that it does not reveal the minimum dimension of the Euclidean space required for the quotient map to be an embedding. For example, an embedding of the configuration space of two spheres requires more than six dimensions, but the ISOMAP algorithm does not protest when asked to construct an isometric embedding in four dimensions even though this results in self intersections.

Whether the ISOMAP algorithm in fact generates an embedding is checked by the following procedure. The map of the quotient space into a Euclidean space should preserve the local geometry isometrically, i.e., all the distances between a configuration and its k-nearest neighbors before and after the transformation should be preserved. This suggests that a simple scatter plot of the distances before and after the transformation could indicate something about the success of the operation. The top panel in Fig. 4.5 represents a schematic for the inverse analysis. After the new coordinates (squares) are constructed, a k-nearest neighbor graph of each point in the new space is found, and the pairwise Euclidean distances d_E in the new space are compared to the corresponding

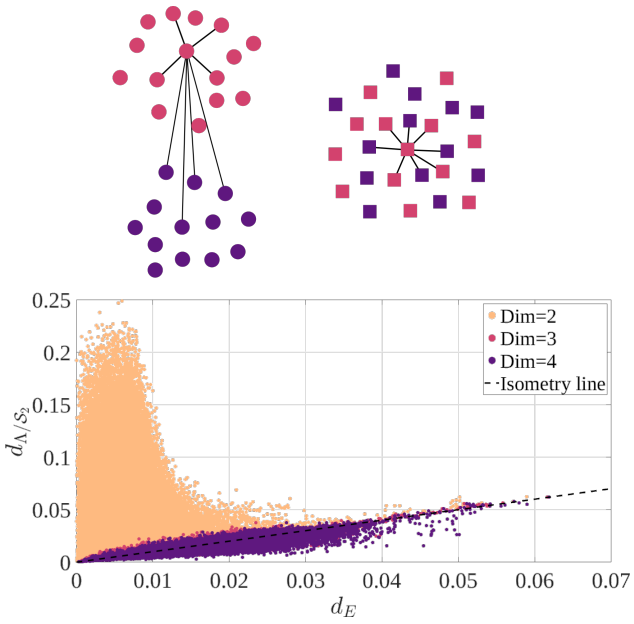

FIGURE 4.5. Top panel shows an improper embedding, i.e., two separate regions in $\Lambda/\mathcal{S}$ (left) are merged in the new Euclidean space (right). Given a point in the new space, the pairwise distances to its neighbors are small whereas the corresponding pairwise distances in the original space are large. The bottom panel shows the result of this analysis for $\Lambda/\mathcal{S}_2$ for $n = 2$ spheres.

$d_{\Lambda/\mathcal{S}}$ in the old space. If the embedding is proper, then these distances should follow the isometry line with little distortion. If it is not, then some of the short edges in the new space will map back to long edges in the old space. The bottom panel shows the result of this analysis for $\Lambda/\mathcal{S}_2$ for $n = 2$ spheres, and provides clear visual evidence that at least three coordinates are needed for a proper embedding of $\Lambda/\mathcal{S}_2$.

The ideas developed so far are applied to the configuration spaces of $n = 2$ spheres because it is the most complicated system that can still be visualized. The summary of the critical points sampled by the procedure in Sec. 2.2 for the $n = 2$ case are given in Tab. 4.3, and these configurations can be visualized using the online database. Following Ref. [**45**], it is observed that the 0-radius configuration acts like an index-3 critical point, and is added to the list even though it is not listed as a critical point in the online database.

TABLE 4.3. The critical points found after millions of random initializations for $n = 2$. Observe that the number of critical points changes dramatically as the space is quotiented by an increasing number of symmetries.

Index	Radius	Packing fraction	# in $\Lambda/\mathcal{S}_1$	# in $\Lambda/\mathcal{S}_2$
0	0.3560	0.5236	1	1
0	0.3062	0.3401	2	1
1	0.2887	0.2850	8	1
2	0.2500	0.1851	6	1
3	0	0	1	1

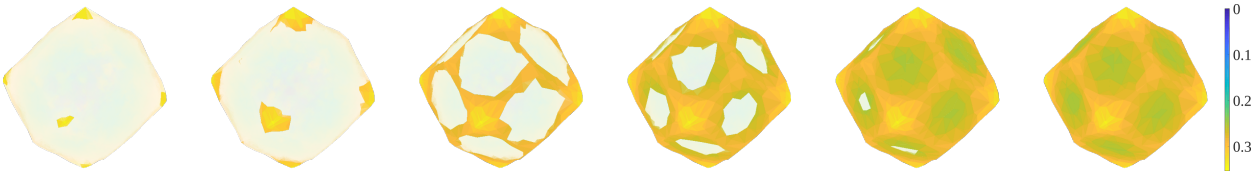

FIGURE 4.6. The evolution of the configuration space quotiented by translations $\Gamma(2, \rho)/\mathcal{S}_1$ (periodic boundary conditions are imposed by identifying opposite faces) with $\rho = \{0.31, 0.29, 0.27, 0.26, 0.25, 0.21\}$.

4.3.1. Adding translation invariance. For a system of $n = 2$ spheres, fixing the first sphere at the origin effectively quotients the configuration space by translations. Since the full configuration space for $n = 2$ is the product space $\mathbb{T}^3 \times \mathbb{T}^3$, the translation invariant configuration space $\Lambda/\mathcal{S}_1$ is in fact equivalent to a 3-dimensional torus $\mathbb{T}^3$ as expected. Figure 4.6 shows the evolution of this space as a function of sphere radius. $\Gamma(2, 0.31)/\mathcal{S}_1$ is comprised of six 0-handles corresponding to the first index-0 critical point. It should be noted that since each of these handles is shared by six rhombic dodecahedra, there is only one index-0 critical point with $\rho = 0.3560$ in this space. The 0-handles corresponding to the second index-0 critical point just appear in $\Gamma(2, 0.29)/\mathcal{S}_1$. There are eight of these handles and they are each shared by four RD, resulting in two index-0 critical points with $\rho = 0.3062$. Observe that for $\rho > 0.2887$ the space is comprised of three disconnected handles, accounting for periodicity. These disconnected handles are connected by 24 1-handles corresponding to index-1 critical points, with each 1-handle shared by three RD for a total of eight index-1 critical points. The index-2 critical points appear in the middle of the rhombic faces when $\rho = 0.25$ and are each shared by two RD for a total of six index-2 critical points. Finally, the space continues to fill for $\rho < 0.25$, and the 0-radius configuration acts like an index-3 critical point that closes the space. The topology of $\Lambda/\mathcal{S}_1 = \Gamma(2, 0)/\mathcal{S}_1$ is that of a 3-dimensional torus.

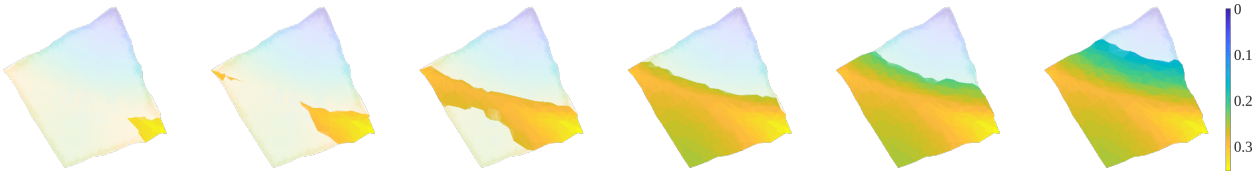

FIGURE 4.7. The evolution of the configuration space quotiented by translations, permutations, inversions and lattice symmetries $\Gamma(2,\rho)/\mathcal{S}_2$ with $\rho = \{0.31, 0.29, 0.27, 0.25, 0.21, 0.15\}$.

4.3.2. Adding permutation and lattice invariance. Figure 4.5 suggests that three coordinates are needed to represent the configuration space quotiented by translations, permutations, inversions, and lattice symmetries. The evolution of this space is shown in Fig. 4.7 as a function of sphere radius. Observe that the first index-0 critical point has already appeared when $\rho = 0.31$, and $\Gamma(2, 0.31)$ is comprised of a single connected component, the 0-handle. When $\rho = 0.29$, $\Gamma(2, 0.29)$ has two 0-handles as disconnected components. The index-1 critical point connects these when $\rho = 0.2887$, and all the spaces for $\rho < 0.2887$ have a single connected component. Further decreasing ρ causes the index-2 critical point to appear when $\rho = 0.25$ and continues to fill the space.

4.3.3. Paths in configuration space. Although the triangulation of a configuration space allows us to study its topological and geometric properties in detail, it is sometimes desirable to have a simpler representation. As previously mentioned in Sec. 2.4.3, the merge tree summarizes how the connected components of a space merge and separate. While useful, this retains no information about the higher index critical points.

Consider a graph with vertices corresponding to the critical points. An edge is placed between two critical points if and only if there exists a path in the space from one to the other such that the value of τ along the path is monotone. This ensures that the path will always be within the superlevel set constrained by the radii of the given critical points, and the path goes directly from one critical point to the other.

Finding such a path is not an easy task. Fortunately, there are well-established algorithms to find minimum-energy paths (MEPs) in the chemical physics literature. The first step to find such a path is to construct the straight line path connecting the two critical points. This path is then relaxed using the zero-temperature string method [**129**, **130**], aligning the path with the gradient

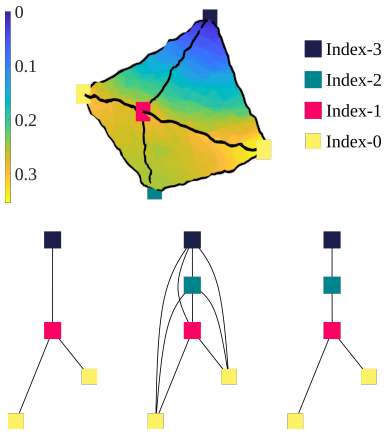

FIGURE 4.8. The top row shows the configuration space $\Lambda/\mathcal{S}_2$ quotiented by permutations, translations, inversions and symmetries of the tiling whereas the bottom row shows its merge tree (left), MEP graph (middle) and the sparsified MEP graph. The black curves on $\Lambda/\mathcal{S}_2$ are the actual MEPs connecting the critical points.

field of the potential energy (the nudged-elastic band method [**66**, **75**] accomplishes the same task). The resulting graph is referred to as the MEP graph.

The top row of Fig. 4.8 shows the configuration space quotiented by permutations, translations, inversions and lattice symmetries $\Lambda/\mathcal{S}_2$. The MEPs between the critical points found with the procedure above are shown as black curves. Though not visible, there exists a path connecting the index-2 and index-3 critical points. The bottom of Fig. 4.8 shows the merge tree (left) and the MEP graph (middle); notice that the merge tree doesn't contain the index-2 critical point since the number of components doesn't change there, implying that the MEP graph in the middle contains strictly more information about the space. The two index-0 critical points are not connected by an edge because there is no way to go from one to the other along a path where τ is monotone. The MEP between them instead passes through the index-1 critical point, acting as a bridge to connect two disconnected portions of the space. This is exactly the location where significant changes to the topological and geometric properties are expected to occur. Finally, the graph on the bottom right in Fig. 4.8 is the sparsified MEP graph where only the edges between critical points with adjacent indices are included. This graph is visually simpler, and can be shown to include precisely the same information as the full MEP graph.

4.4. Diffusion Diameter

Although the MEPs have a clear physical meaning, a thermodynamic system doesn't always follow these paths from one state to another. A concept like the diffusion distance instead effectively considers all the paths that the system could take, and provides a clearer connection to the mixing time of the system.

The diameters of the configuration spaces $\Gamma(2,\eta)/\mathcal{S}_1$ (left) and $\Gamma(2,\eta)/\mathcal{S}_2$ (right) as measured by the diffusion distance are plotted on the bottom of Fig. 4.9 as a function of packing fraction η. The top shows the sparsified MEP graphs of both spaces. The vertical lines indicate the corresponding packing densities of the critical configurations. In both cases, the disconnected regions of the space corresponding to index-0 critical configurations start growing slowly at the corresponding densities. When the index-1 critical configuration appears at $\eta = 0.285$, the disconnected components get connected and discontinuous jumps in the diffusion distance are observed in both cases (the small gap in the left figure is due to the finite resolution of the α-complex). From a thermodynamic standpoint, this is the location where the available portion of the configuration space and the domain of integration abruptly changes. Further decreasing the density decreases the diffusion distance since more paths are available to the system. Notice that the diffusion distance on the right starts to increase after the index-2 critical point appears. The zero density (or zero radius) configuration corresponds to a sharp corner, making it difficult for a random walker to diffuse into that region and increasing the diffusion distance.

Although the geometric analysis in this section is performed for $n = 2$ spheres, it is conjectured that there will be a corresponding signature of a first-order phase transition around the melting packing fraction in the thermodynamic limit. The strongest support for this conjecture is the relationship of the critical point distribution (especially the index-1 critical points) to the coexistence interval in Fig. 4.4. One of the main actions of index-1 critical points is to connect basins of attraction and thereby to introduce discontinuities in the space diameter. That the number and perhaps concentration of index-1 critical points within the coexistence interval both increase with n suggests that there could be a catastrophic geometric event coinciding with the appearance of the liquid in the thermodynamic limit, and that this could be associated with a similarly catastrophic

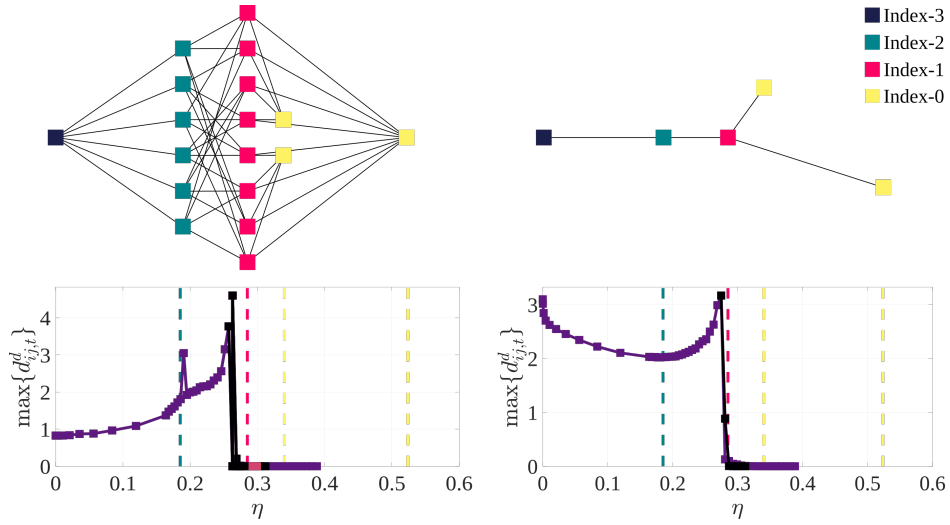

FIGURE 4.9. The sparsified MEP graphs of the configuration spaces quotiented by the translations (top left) and all the symmetries listed in Tab. 4.2 (top right). The corresponding diffusion distances at $t = 8$ and $t = 20$ are plotted below the sparsified MEP graphs. Observe that the discontinuous jumps coincide with the merging of the components; the small gap observed between the packing fraction of the critical points and the discontinuity in the left plot is due to finite resolution of the α-complex.

discontinuity in the space diameter. Many of the present results are intended to be the foundations of a future inquiry into whether this is indeed so.

CHAPTER 5

A Geometric Conjecture about Phase Transitions*

This chapter further investigates the geometric nature of a phase transition using the hard disk and sphere systems as a model. The application of the ideas discussed in the previous chapters is shown in Fig. 5.1 using a toy model. 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 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 it is reasonable to suppose that the magnitude of the discontinuities and their location with respect to any control variables would vary systematically with size in a more realistic situation.

Referring back to the discussions in Ch. 1, the fundamental conjecture of this work is:

*The content of this chapter has previously been published in Ref. [**46**].

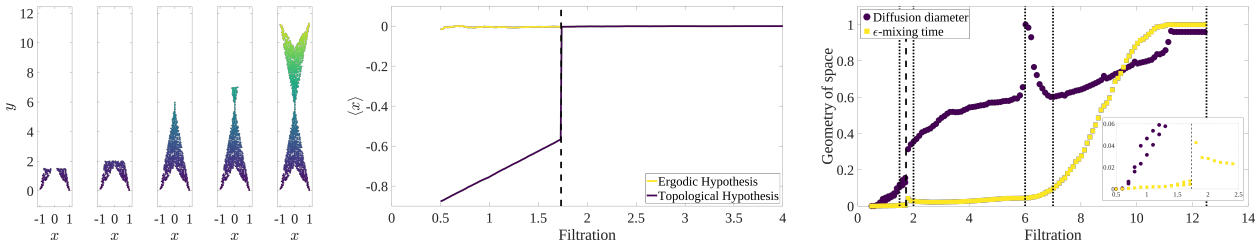

FIGURE 5.1. The left panel shows the evolution of a model configuration space with points colored by their y-coordinate. The filtration value is an upper bound for the y-coordinate of any point on the system's trajectory. The middle panel shows the average x-coordinate $\langle x \rangle$ calculated 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 mixing time, is also sensitive to the bottleneck. The five dotted lines show the filtration values used in the left panel.

CONJECTURE 1.0.1. *A necessary condition for a first-order phase transition is a discontinuity in the mixing time on the configuration space.*

The preliminary findings in the previous chapters suggest that the critical points in the configuration space are in fact related to the observed phase transitions. The technical contribution of this chapter is a set of techniques described in Sec. 5.2 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. 5.3) of these spaces are computed as functions of particle diameter and particle number, and in Sec. 5.4 significant discontinuities are observed at the packing fractions of critical points close to the phase coexistence intervals. Along with the observation [**48**] 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. 1.0.1.

5.1. Critical Configurations

This chapter respectively studies the configuration spaces of hard disks and spheres in the hexagonal torus and rhombic dodecahedron shown in Fig. 2.1. The procedure to identify the

critical points is previously presented in Sec. 2.2. Interactive databases of the critical points 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. [**48**, **113**] and in Fig. 4.4 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.

5.2. Configuration Spaces

The procedure followed previously in the previous chapters or in Refs. [**45**, **48**] 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, dramatically reducing the 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 [**45**, **48**] or in Ch. 3, 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, the distance functions in Eqs. 3.1 and 3.2 are used for this purpose. While algorithms like Tabu search [**27**, **58**] can be used to calculate these distances, the number of problems to be solved increases rapidly with the number of isometry groups considered. For example, when $\mathcal{S} = \mathcal{T} \cup \mathcal{P} \cup \mathcal{L}$, the global optimization over rigid translations needs to be solved $n! \times O(\mathcal{L})$ times to evaluate Eq. 3.2 where $O(\mathcal{L})$ is the order of $\mathcal{L}$.

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

The second technique reduces the number of optimization problems that need to be solved in Eq. 3.2. 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 Sec. 5.5. This reduces the computational cost enough to be able to evaluate a small number of distances $d_{\Lambda/\mathcal{S}}$, but not all the pairwise distances in the point cloud when identifying nearby points to reconstruct $\Lambda/\mathcal{S}$.

The third techniques 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 $\mathcal{L}$ [**45**, **48**]. 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

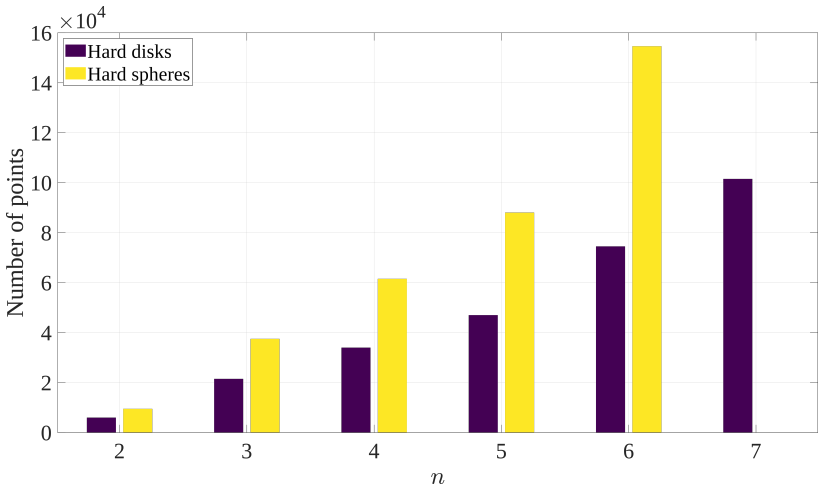

FIGURE 5.2. Number of points used in the graph representations for hard disks and spheres.

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. 5.2) 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 [**45**, **48**]. 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 [**21**, **117**]. Fortunately, the only geometric information required to calculate the diffusion diameter and the ϵ-mixing time are geodesic distances, and these can be reasonably approximated from the k-nearest neighbor graph alone. Figure 5.3 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

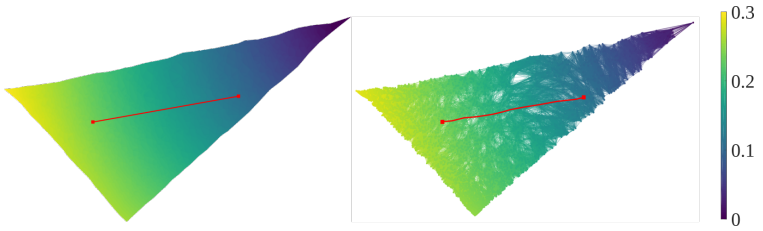

FIGURE 5.3. The quotient space $\Lambda(2)/\{\mathcal{T}, \mathcal{P}, \mathcal{L}\}$ of two points on T^2 represented as an α-complex (left) and a graph (right), with color indicating the value of the tautological function. The red path connecting two points is a straight line on the left, but is slightly longer on the right.

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 5.4 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 [**76**] 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.

5.3. Mixing Time

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

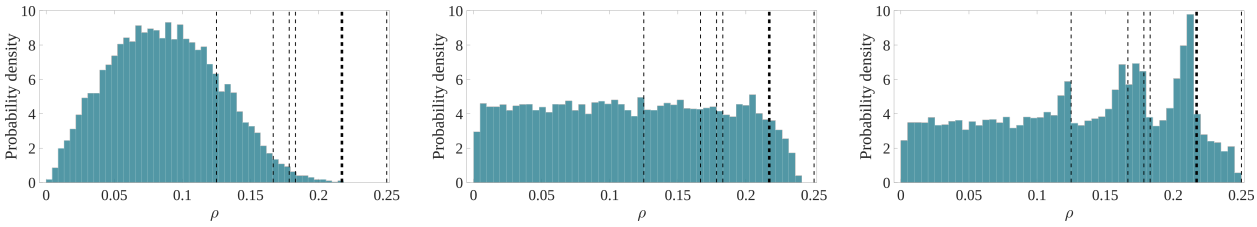

FIGURE 5.4. Distributions of tautological function values for point clouds on $\Lambda(4)$ sampled by three different procedures, with the dashed lines indicating the radii of previously-identified critical points. Sampling points uniformly gives the distribution on the left. The distribution in the 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.

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. This chapter studies the diameter of configuration space as measured by the diffusion distance and the mixing time as a function of packing fraction (or equivalently, the disk radius). Since the diffusion distance is previously defined in Sec. 2.5, this section begins to develop the notion of the 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, $\bar{q}$ and $\bar{p}$ be the canonical coordinates and momenta, and $\mu(\bar{q}, \bar{p}, t)$ be the probability distribution of microstates on Ω at time t. Suppose that the system has an arbitrary initial condition $\mu_0(\bar{q}, \bar{p}) = \mu(\bar{q}, \bar{p}, 0)$. Briefly setting aside the specifics of the time evolution of $\mu(\bar{q}, \bar{p}, t)$, the steady-state distribution $\mu_\infty(\bar{q}, \bar{p}) = \lim_{t\to\infty} \mu(\bar{q}, \bar{p}, t)$ defines the equilibrium condition on Ω, and the mixing time could in principle be defined by the approach of $\mu(\bar{q}, \bar{p}, t)$ to $\mu_\infty(\bar{q}, \bar{p})$. 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 $\mu_0(\bar{q}, \bar{p})$ 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 $\mu(\bar{q}, \bar{p}, t)$ is governed by the Liouville equation [**57**, **89**]. 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 widely discussed in the literature. This discussion is avoided by simply supposing that $\mu(\bar{q}, \bar{p}, t)$ evolves by the diffusion equation, while not endorsing this approach generally. 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 $\nu(\bar{q}, t)$ on the configuration space Ω_q to the limiting distribution $\nu_\infty(\bar{q})$.

The Kullback-Leibler divergence [**82**] 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_∞:

$$I(t; \nu_0) = \int_\Omega \nu(\bar{q}, t) \log \frac{\nu(\bar{q}, t)}{\nu_\infty(\bar{q})} \mathrm{d}\bar{q}. \tag{5.1}$$

It is proposed that the condition for $\nu(\bar{q}, t)$ to have converged to $\nu_\infty(\bar{q})$ 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 $\bar{q}_0$, and that $t_\epsilon(\bar{q}_0)$ is the minimum time required for this initial condition to converge to $\nu_\infty(\bar{q})$ 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

$$\langle t \rangle_\epsilon = \int_\Omega t_\epsilon(\bar{q}) \nu_\infty(\bar{q}) d\bar{q} \tag{5.2}$$

where the measure of integration is taken to be $\nu_\infty(\bar{q})$ in the absence of a natural alternative. The ϵ-mixing time is regarded as a precisely-defined proxy for the thermodynamic mixing time on the configuration space in the following.

This leaves only the definition of a diffusive process occurring on a graph G instead of on a continuous space Ω_q. Let $\bar{\nu}(t)$ be the probability masses on the vertices of G at time t, and define the graph Laplacian as $\bar{\bar{L}} = \bar{\bar{D}} - \bar{\bar{K}}$ [**28**]. The governing equation for a diffusive process is

$$(\partial/\partial t + \bar{\bar{L}})\bar{\nu}(t) = 0. \tag{5.3}$$

Let $\bar{\bar{\Lambda}}$ be the diagonal matrix of the eigenvalues $\lambda_0 = 0 < \lambda_1 \cdots \leq \lambda_{N-1}$ of $\bar{\bar{L}}$, and $\bar{\bar{\Phi}}$ be the matrix of the corresponding eigenvectors where the first column is a constant vector. Since $\bar{\bar{L}}$ is a symmetric matrix, the columns of $\bar{\bar{\Phi}}$ form an orthogonal basis and the solution to the diffusion equation is

$$\bar{\nu}(t) = \bar{\bar{\Phi}} e^{-t\bar{\bar{\Lambda}}} \bar{\bar{\Phi}}^T \bar{\nu}_0 \tag{5.4}$$

for the initial condition $\bar{\nu}_0$. The steady-state distribution is readily calculated using the fact that all of the diagonal elements of $e^{-t\bar{\bar{\Lambda}}}$ 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. 5.1 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. 5.2. More details about the calculation of the ϵ-mixing time for this system are given in Sec. 5.7.

5.4. Results

The hard disk equation of state usually appears in the literature as a function of packing fraction η [**103**, **121**]. The radius ρ of the disks appearing in Eq. 2.3 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. 2.1.

Chapters 3 and 4 previously constructed quotient spaces of two hard disks and two hard spheres, and studied the topological and geometric properties of both the original and quotient spaces as functions of packing fraction. It is observed that while the geometric and topological features of the 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. 1.0.1 to be tested more easily and for systems with larger numbers of disks.

Figure 5.5 shows the ϵ-mixing times of the quotient spaces $\Gamma(2,\eta)/\mathcal{T}$ and $\Gamma(2,\eta)/\{\mathcal{T}\cup\mathcal{P}\cup\mathcal{L}\}$ 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

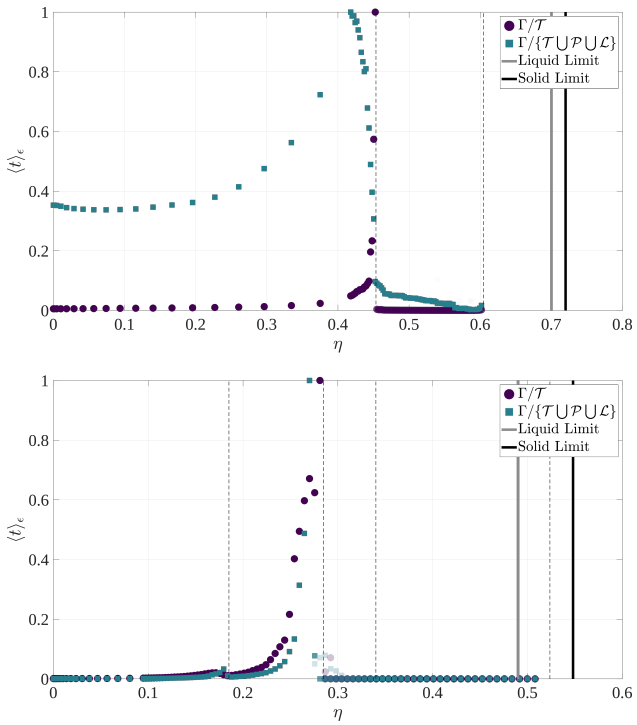

FIGURE 5.5. ϵ-mixing times of the quotient spaces $\Gamma(2,\eta)/\mathcal{T}$ and $\Gamma(2,\eta)/\{\mathcal{T}\cup\mathcal{P}\cup\mathcal{L}\}$ for hard disks (top) and hard spheres (bottom). Dashed lines represent the packing fractions of the critical points.

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 $\Gamma/\{\mathcal{T}\cup\mathcal{P}\cup\mathcal{L}\}$ will be used exclusively in the following.

Figure 5.6 shows the diffusion diameters and ϵ-mixing times of the quotient spaces $\Gamma/\{\mathcal{T}\cup\mathcal{P}\cup\mathcal{L}\}$ 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 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

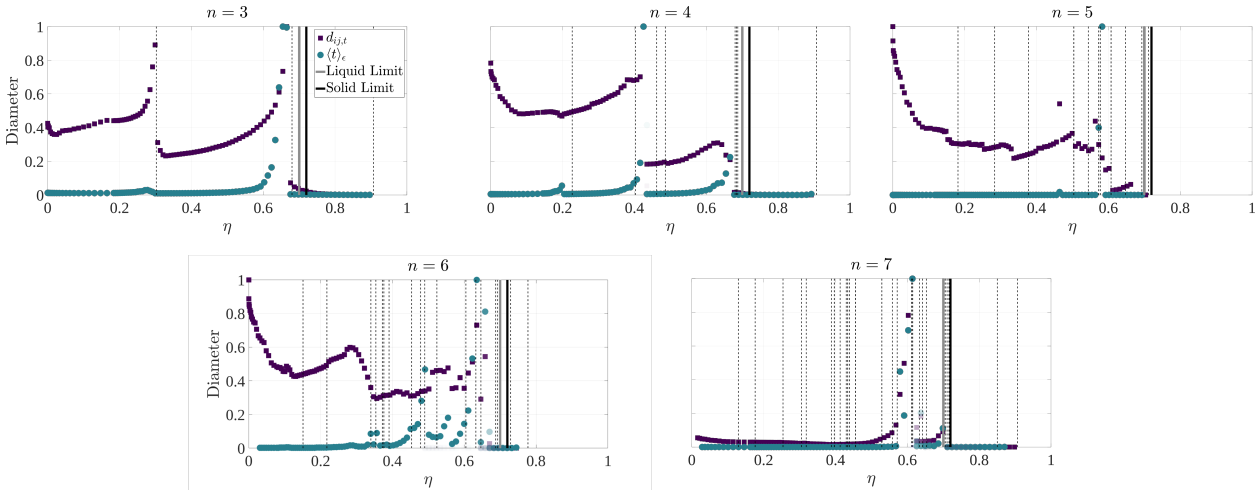

FIGURE 5.6. Diffusion diameters and ϵ-mixing times of the spaces $\Gamma/\{\mathcal{T} \cup \mathcal{P} \cup \mathcal{L}\}$ for $n = 3 \ldots 7$ hard disks as functions of packing fraction. Dashed lines shows the packing fractions of the critical points.

diffusion distance and ϵ-mixing time do not always occur at the same packing fractions, the largest discontinuities generally do.

Figure 5.7 shows the corresponding results to Fig. 5.6 for $n = 3 \ldots 6$ hard spheres. While the number of critical points is greatly increased relative to the hard disk systems in Fig. 5.6, 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. 5.2, the memory requirements increase so rapidly with n that the spaces for $n \geq 7$ could not realistically examined. Even for five and six hard spheres, the volume of the space increases so rapidly with decreasing packing fraction that the diffusion diameters and ϵ-mixing times over the entire domain could not be reported. This does not substantially affect our conclusions though, since the sampled domain of η already extends well below the coexistence interval, and notably includes all of the index-1 critical points.

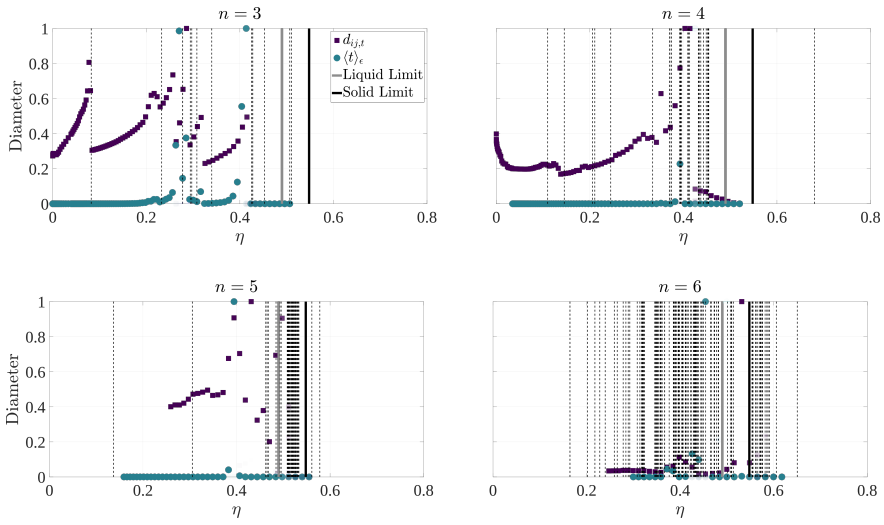

FIGURE 5.7. Diffusion diameters and ϵ-mixing times of the spaces $\Gamma/\{\mathcal{T} \cup \mathcal{P} \cup \mathcal{L}\}$ for $n = 3 \ldots 6$ hard spheres as functions of packing fraction. Dashed lines shows the packing fractions of the critical points.

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 5.8 shows the packing fractions of the largest discontinuities in Figs. 5.6 and 5.7 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 $n \leq 7$ hard disks and $n \leq 6$ hard spheres, populations of critical points for $n \leq 12$ hard disks and hard spheres have been sampled using established techniques that are described elsewhere [**45**, **48**, **113**]. 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. 1.0.1 could be true for hard sphere systems, and perhaps for thermodynamics systems as well.

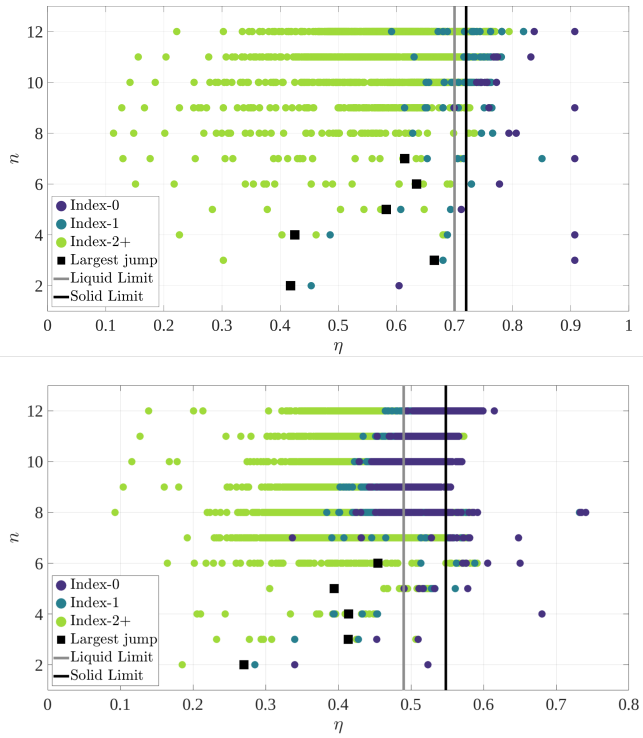

FIGURE 5.8. The largest discontinuous jumps (black squares) observed in Figs. 5.6 (top) and 5.7 (bottom) along with the packing fractions of all known critical points, colored by their indices (index-0 in purple, index-1 in blue, all others in green).

5.5. Heuristic for the distance

Let $\mathcal{S} = \mathcal{T} \cup \mathcal{P} \cup \mathcal{L}$ in Eq. 3.2. Since $\mathcal{T}$ is a continuous group and $\mathcal{P} \cup \mathcal{L}$ is discrete, Eq. 3.2 can be rewritten as

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

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

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

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 $d_{\Lambda}[\mathbf{x}, S(\mathbf{y}) + \mathbf{t}]$ [**131**]. 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.

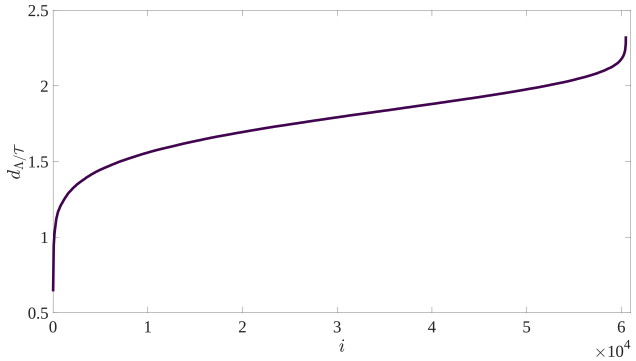

FIGURE 5.9. The sorted true distances $d_{\Lambda/\mathcal{T}}$ calculated for a fixed generic configuration and the $7! \times 12$ symmetric versions of a second generic configuration for 7 hard disks.

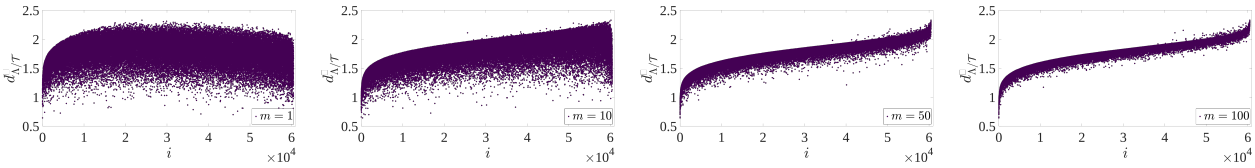

FIGURE 5.10. The same configurations as in Fig. 5.9 are sorted by the approximate values of $d_{\Lambda/\mathcal{T}}[\mathbf{x}, S(\mathbf{y})]$ for $m = 1$, 10, 50, and 100 and are plotted with the true values of $d_{\Lambda/\mathcal{T}}[\mathbf{x}, S(\mathbf{y})]$ on the vertical axis.

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 5.9 shows the sorted true distances $d_{\Lambda/\mathcal{T}}$, and $d_{\Lambda/\mathcal{S}}$ is the minimum of this set. Figure 5.10 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. 5.9, 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}.

5.6. Eigenspectrum analysis

The diffusion coordinates in Eq. 2.5 involve the tth powers of the eigenvalues of $\bar{\bar{P}}$ 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 [**118**]. Figure 5.11 shows the tth powers of the eigenvalues of $\bar{\bar{P}}$ for $\Lambda(4)/\{\mathcal{T} \cup \mathcal{P} \cup \mathcal{L}\}$ 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 precision of less than 10^{-8} and making this source of error negligible compared to other sources.

5.7. Mixing time of the example system

Figure 5.12 shows the ϵ-mixing times $t_\epsilon(\bar{q}_0)$ as a function of $\bar{q}_0$ for the example system in Fig. 5.1. 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

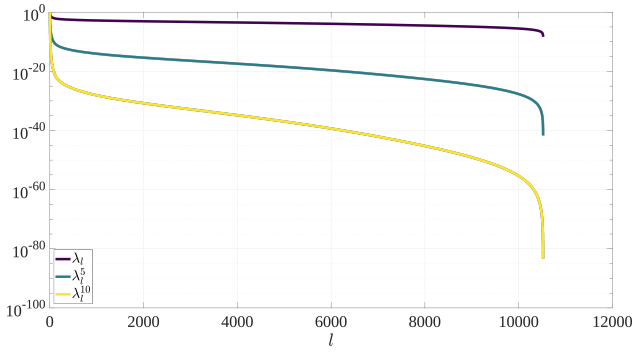

FIGURE 5.11. The tth powers of the eigenvalues of $\bar{\bar{P}}$ for $\Lambda(4)/\{\mathcal{T} \cup \mathcal{P} \cup \mathcal{L}\}$ at times $t = 1$, 5, and 10. The values generally decay rapidly, and are often numerically indistinguishable from zero for sufficiently large t.

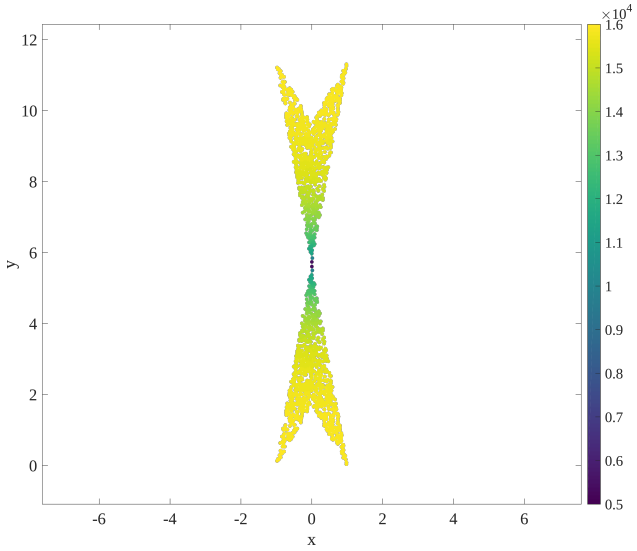

FIGURE 5.12. Individual mixing times $t_\epsilon(\bar{q}_0)$ of the example system shown in Fig. 5.1. The average mixing time $\langle t \rangle_\epsilon$ is around 15 000.

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.

CHAPTER 6

Foundations of a Finite Non-Equilibrium Statistical Thermodynamics: Extrinsic Quantities*

The previous chapters gradually build a notion of how certain symmetries affect the topology and geometry of configuration spaces. In particular, the relationship between the system dynamics and the configuration space topology/geometry is studied. For example, ϵ-mixing time roughly provides a measure for identifying whether an initial probability distribution on the configuration space has reached a steady-state or not. This chapter aims to further extend these notions to finite systems in the context of non-equilibrium statistical thermodynamics.

While thermodynamics is nominally concerned with heat and energy flows, the only states about which anything meaningful can be said about are those infinitesimally close to equilibrium, i.e., those for which heat and energy flows vanish. Moreover, being based on empirical observations, thermodynamics has little to say about the underlying causes for observed phenomena. That is instead the subject of statistical thermodynamics, which has the purpose of establishing connections between the microscopic state of a system (i.e., the particle positions and momenta) and equilibrium thermodynamic quantities. Although statistical thermodynamics can rightfully claim a number of significant successes, the theory says little about non-equilibrium systems by design.

That is instead the concern of non-equilibrium statistical thermodynamics. This subject has received considerable attention (perhaps for the reason that the most physically significant systems are often those not in equilibrium) but is not nearly as well-established as equilibrium statistical thermodynamics. This can be partly attributed to the difficulty of adequately defining and measuring thermodynamic quantities for a non-equilibrium system; without established experimental points of reference, it is difficult to conclude whether any given proposal should be discarded as being discordant with reality. Part of the difficulty is also that there are ontological questions that

*The content of this chapter has previously been published in Ref. [**47**].

have not been satisfactorily resolved. For example, can a thermodynamic entropy be defined for a finite isolated system, one of precisely the type that is frequently simulated by molecular dynamics?

There are at least three practical reasons to search for answers to such a question. First, students and practitioners who are asked to accept the conclusions of statistical thermodynamics without finding satisfactory answers to such questions could easily lose confidence in a subject that is, and should be regarded as, one of the cornerstones of modern physics. Second, the inconsistencies that emerge when pondering such questions could erroneously suggest that our instinctive sense of reality is not relevant to this subject; without this, our ability to reason about and develop non-equilibrium statistical thermodynamics is severely limited. Third, the possible applications of molecular dynamics to the physical sciences would be considerably expanded if thermodynamic quantities like the free energy could be reliably extracted from molecular dynamics simulations.

The purpose of this chapter is to establish part of the foundations for a non-equilibrium statistical thermodynamics that is consistent with existing results in the equilibrium thermodynamic limit, and that introduces a minimum of arbitrary constants and additional physics. It must be mentioned that although nature itself is quantum mechanical at small length scales and quantum mechanics is essential to explain certain phenomena, this chapter approaches the problem in a classical setting. The intention is not to ignore the implications of quantum mechanics, but to develop notions of certain concepts like distributions in a classical context where questions like the nature of the wave function collapse do not need to be simultaneously answered. Section 6.1 motivates this undertaking by considering several limitations of the existing foundations, including situations where the ergodic hypothesis does not hold, the dependence of thermodynamic ensembles on initial conditions, and the increase of entropy during irreversible processes. Section 6.2 proposes definitions applicable to finite non-equilibrium systems of all of the extrinsic quantities in the fundamental thermodynamic relation. These definitions satisfy the same properties as the classical quantities in the equilibrium thermodynamic limit. Section 6.3 evaluates these quantities as functions of time and space for the expansion of an ideal gas in an isolated box, and verifies that our definitions converge to the expected values when interpreted in a classical context.

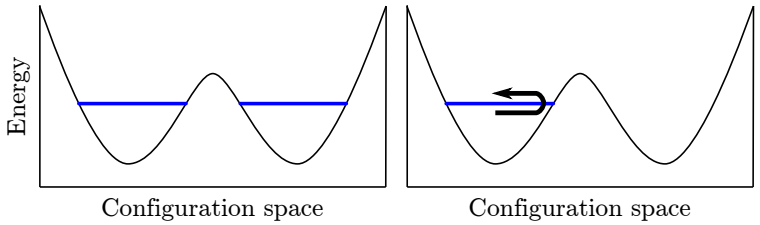

FIGURE 6.1. A schematic showing the confinement of a microcanonical system to a small part of the configuration space. The microcanonical ensemble (left) stipulates that all microstates of a given total energy (blue, with the potential energy in black) be equally probable. For a physical system (right), the trajectory is confined to the path-connected region containing the initial condition.

6.1. Limitations of Statistical Thermodynamics

The following limitations in the foundations of statistical thermodynamics are not intended to be exhaustive, but rather to provide motivation for the developments in subsequent sections.

6.1.1. Thermodynamic Ensembles. The utility of a thermodynamic ensemble derives from the ergodic hypothesis [**16**,**41**], or the effective assumption that the time average of a quantity over a generic trajectory through the phase space can be replaced by the instantaneous average of the quantity over an ensemble of systems prepared in the same macrostate. Practically speaking, the ability to replace a complicated calculation of the evolution of a microstate with an average over the phase space, weighted by an appropriately constructed probability distribution, is the only reason that statistical thermodynamic calculations are at all possible. That said, a variety of systems do not behave ergodically [**35**, **55**, **62**, **80**], and for these the above procedure leads to obvious inconsistencies.

As an initial example, consider a microcanonical system containing two potential energy wells, as shown in Fig. 6.1. The microcanonical ensemble stipulates that, following the principle of equal *a priori* probabilities [**96**, **122**], the time average of an observable for a microcanonical system with total energy in blue is equivalent to the average over the allowed states in both potential energy wells. That said, the allowed states are separated by a disallowed region, any trajectory is necessarily confined to the potential energy well specified by the initial conditions (i.e., the system is not metrically transitive [**13**, **105**]), and the ergodicity hypothesis does not hold. The two-dimensional square lattice Ising model with zero external field below the critical temperature

is a nontrivial system that exhibits such behavior [**106**, **132**], and shows that this situation is not purely hypothetical. This leads to the natural conclusion that the time-averaged behavior of a generic system can depend on the initial conditions, even in the long-time limit.

One could object that it is not really possible to isolate a physical system from its surroundings, and that ergodicity is restored by the thermal fluctuations in a canonical system eventually allowing the trajectory to pass over the potential energy barrier. The characteristic time required for such a fluctuation depends exponentially on the height of the potential energy barrier though, meaning that the mean residence time in a potential energy well could easily be longer than the lifetime of the universe. More generally, whether the time average of an observable should be performed over one potential energy well or two depends on whether the observation time is shorter or longer than the mean residence time. That is, the time-averaged behavior of a generic system apparently can depend not only on the initial conditions, but also on the period of observation.

Further evidence that an ensemble average is not generally equivalent to the finite time-averaged behavior of a system is offered by the phenomenon of supercooling. Consider a system of supercooled water prepared without any heterogenous crystal nucleation sites. For any specified observation time, let the temperature be such that the system has equal probabilities of homogeneously nucleating a crystal and of remaining a supercooled liquid. The ensemble average would include equal numbers of liquid and crystal states, whereas the time-average behavior of a single system would depend stochastically on whether crystallization occurs in that system. The natural objection to this thought experiment is that thermodynamic ensembles are only relevant to systems in thermodynamic equilibrium, and for that reason are not applicable to a supercooled liquid. That is precisely the point; thermodynamic ensembles have well-established utility but cannot be used in every situation, and care should be taken that the underlying assumptions are not violated.

For the final example, suppose that a phase is defined by a probability distribution on the phase space that is a smooth function of the thermodynamic variables, and that phase transitions occur when this probability distribution changes discontinuously. Consider the proposed existence of an ideal glass phase as a resolution to Kauzmann's paradox [**56**, **79**]. Cooling an ensemble of glass forming systems through the glass transition temperature implies that the probability distribution associated with a proposed ideal glass phase should be distributed over all metabasins

in the potential energy landscape. This is inconsistent with the observation that the instantaneous configuration of a single glassy system identifies one or a few metabasins to which the system is confined for reasonable observation times. That is, the actual probability distribution of microstates for a single glassy system is inconsistent with the one that would be characteristic of an ideal glass phase, though confusion about this point could occur as a result of inappropriately applying thermodynamic ensembles to a non-equilibrium system.

6.1.2. Probability Distributions. This discussion raises questions about the interpretation of a probability distribution of microstates. Given a system in thermodynamic equilibrium, the usual interpretation is that if the practitioner were to repeatedly measure the microstate at intervals longer than the decorrelation time of the system, then the microstates would be independently and identically distributed according to this probability distribution. This requires that the probability distribution be stationary, i.e., the partial derivative with respect to time vanishes everywhere. Along with the constraints imposed by the Liouville equation and the requirement that the weighted average of thermodynamic variables over the phase space be equal (to within a small uncertainty) to the experimentally-measured values, this is conjectured to make the probability distribution well-defined [**30**].

Actually calculating a probability distribution in this way would be exceedingly difficult though. Instead, the usual approach is to construct an ensemble of systems with different initial conditions subject to the requirement that, perhaps after an initial transient, every system in the ensemble realizes a probability distribution of microstates the same as that of the single system described above [**16**, **94**]. The probability distribution can then be empirically constructed by examining the microstates of the ensemble at any single instant in time after the initial transient. This has the advantage of removing the dependence on the decorrelation time, but still requires that the probability distribution be time invariant and consistent with the macrostate.

The interpretation of a probability distribution of microstates for a non-equilibrium system is less clear. Since the properties of a non-equilibrium system (and therefore the probability distribution of microstates) are not generally constant in time, a practitioner cannot reliably construct an empirical probability distribution from a time series of microstates. As for the construction of a thermodynamic ensemble, requiring the observation time to be longer than the decorrelation time

is no longer sufficient for ensemble averages to be independent of the choice of initial microstates (as shown by the thought experiment below), and a thermodynamic ensemble cannot be constructed without making this choice explicit.

Consider a thermodynamic ensemble corresponding to any equilibrium state. The time evolution of the systems in the ensemble maps to an ensemble of trajectories in the phase space. If the practitioner makes an instantaneous change to a thermodynamic variable defining the macrostate, the systems in the ensemble would no longer be in a state of equilibrium. A time-dependent empirical probability distribution for the non-equilibrium state could then be constructed from the ensemble of trajectories, and would be equivalent to the result of applying the Liouville equation to the initial probability distribution. The difficulty with this construction is that, as pointed out in Sec. 6.1.1, thermodynamic ensembles are not generally applicable to non-equilibrium systems. If the initial equilibrium system was a glass former in the liquid phase, instantaneously quenching the system through the glass transition temperature would cause the systems in the corresponding ensemble to break into distinct populations confined to separate metabasins in the potential energy landscape. The actual system would follow a trajectory into one of these metabasins, with the result that the probability distribution of microstates for the actual system would be inconsistent with the prediction of the ensemble.

This conflict is resolved by recognizing that the probability distribution of microstates actually represents the practitioner's subjective uncertainty about the system's microstate [**72**, **73**]. The practitioner does not precisely know the initial microstates of the system nor of the surroundings, and therefore does not know the system's subsequent trajectory. Without this information, it is quite reasonable to represent the initial uncertainty by means of a probability distribution and to propagate that uncertainty to subsequent times with the Liouville equation. That the uncertainty is not an objective property of a thermodynamic system should not affect the apparent properties of generic equilibrium states, since for these there is assumed to be a limiting probability distribution that is independent of the initial conditions. This certainly does not apply to all thermodynamic states though, the clearest examples being ones close to a phase transition where the time averages of distinct trajectories can diverge.

A possible objection to the conclusion of the preceding paragraph is that this makes the thermodynamic entropy a subjective quantity, though the consequences of this are less severe than could be supposed. Consider that the part of the internal energy that cannot feasibly be extracted as work should be expected to increase with the practitioner's uncertainty about the microstate. This allows irreversible processes, e.g., the expansion of an ideal gas into an empty volume, to be interpreted as those where the practitioner's uncertainty about the system's microstate spontaneously increases. Conversely, an entity with perfect knowledge of a system's microstate would conclude that the subjective entropy was zero and be able to extract all of the internal energy as work, allowing Maxwell's demon to avoid violating any thermodynamic laws.

Jaynes' discussion of the mixing paradox offers further evidence that the entropy should be considered as a subjective quantity [**74**]. Consider a thermodynamic system consisting of two chambers separated by a removable barrier, the first containing gas A and the second containing gas B. If the practitioner is able to distinguish gases A and B, then the entropy of the system should increase with the removal of the barrier and the mixing of the gases. Conversely, if the practitioner cannot distinguish gases A and B, then the removal of the barrier is reversible and should not change the entropy of the system. The mixing paradox arises even if the gases are arbitrarily similar to one another suggesting that the additional entropy associated with the mixing depends only on the fact that the gases are distinguishable to the practitioner. While the physical properties of the gases are arbitrarily close to each other, the entropy change on removing the barrier is apparently discontinuous at the point where the gases first become indistinguishable. This would be concerning if the entropy were an objective property of the system, but if the entropy is subjective this merely reflects that the practitioner's knowledge of the possible microstates changes discontinuously.

The subjectivity of the probability distribution allows a significant reinterpretation of the discussion of supercooling in Sec. 6.1.1. After placing a liquid in a freezer, the practitioner does not know whether the system has crystallized or not until the freezer door is opened. The probability distribution describing the practitioner's knowledge of the system state at this point includes equal numbers of liquid and crystal states; this accurately reflects that it is not possible to reliably predict the work that could be extracted from the system, or more generally the system's responses to external stimuli, until more information is acquired. Opening the door provides this information

and collapses the probability distribution into one containing only liquid or only crystal states, once again allowing useful predictions to be made. This only changes the practitioner's knowledge of and ability to harness the system. Thermodynamics is then not so much about describing physical systems as it is about describing the ways in which physical systems can be manipulated.

6.1.3. Liouville's Theorem. Given an initial probability distribution on the phase space, the Liouville equation describes the time evolution of this probability distribution on the basis of Hamiltonian mechanics [**57**, **89**]; conceptually, the flow of probability density resembles that of an incompressible fluid. Liouville's theorem states that the convective derivative of the probability distribution is zero, or that any set of points beginning within a specified volume can always be enclosed in a distorted region of identical volume at any subsequent time. Suppose that the thermodynamic entropy is given by Gibbs's entropy formula for a continuous probability distribution

$$S_{\mathrm{G}} = -k_{\mathrm{B}} \int_{\Gamma} \rho \ln \frac{\rho}{\mu} \mathrm{d}\Omega \tag{6.1}$$

where Γ is the phase space, $\mathrm{d}\Omega$ is the infinitesimal volume element, ρ is the probability distribution, and μ is a constant that makes the argument of the logarithm unitless. Liouville's theorem then implies that Gibbs's entropy is constant for any time evolution given by the Liouville equation, even for initial conditions that describe non-equilibrium systems. For this reason, it is unclear how the second law of thermodynamics can be derived from statistical thermodynamics without the introduction of additional physics.

Only two of the resolutions proposed in the literature are considered here. The first is known as coarse-graining, and involves partitioning the phase space into regions of a characteristic volume V_{CG} [**57**, **108**, **122**]. This allows the continuous probability distribution ρ to be converted into a discrete probability distribution q_i indexed by the partition elements. Gibbs's entropy formula for a continuous probability distribution is then converted into with one for a discrete probability distribution (up to an additive constant):

$$S_{\mathrm{CG}} = -k_{\mathrm{B}} \sum_i q_i \ln q_i.$$

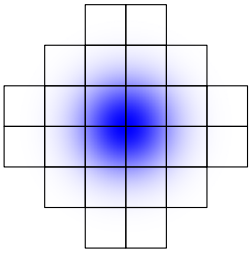

FIGURE 6.2. Representation of a typical partition of the phase space containing a fixed number of elements, with the probability distribution in blue.

This approach has several advantages. With regard to the second law, the increasing distortion of a volume element in the phase space with time results in a continuous probability distribution being distributed over many elements of the partition, allowing the coarse-grained entropy to increase despite Liouville's theorem. More precisely, the introduction of V_{CG} defines a length scale below which the contraction of a continuous probability distribution cannot be resolved, and after an initial transient the rate of expansion of the continuous probability distribution effectively depends only on the positive Lyapunov exponents [**38**,**109**]. With regard to the third law, the concentration of a continuous probability distribution at a single point as the temperature approaches absolute zero would cause the entropy in Eq. 6.1 to diverge to negative infinity, but the coarse-grained entropy approaches the conventional value of zero for any nonzero V_{CG}.

Everything comes with a price though, and for the coarse-graining approach that is introducing V_{CG} as a fundamental physical constant that behaves as an offset to the entropy. Consider that the coarse-grained entropy for a uniform continuous probability distribution on some volume V_Γ of the phase space is

$$\begin{aligned} S_{\mathrm{CG}} &= -k_{\mathrm{B}} \sum_i \frac{V_{\mathrm{CG}}}{V_\Gamma} \ln \frac{V_{\mathrm{CG}}}{V_\Gamma} \\ &= k_{\mathrm{B}} \ln \frac{V_\Gamma}{V_{\mathrm{CG}}}. \end{aligned}$$

Since taking the $V_{\mathrm{CG}} \to 0$ limit causes the coarse-grained entropy to diverge, every thermodynamic system should have a nonzero V_{CG}. Moreover, for thermodynamic comparisons to be possible, this value should be the same for all thermodynamic systems. The definition of a fundamental physical constant is not something to be undertaken lightly though, and the correct identification of V_{CG}

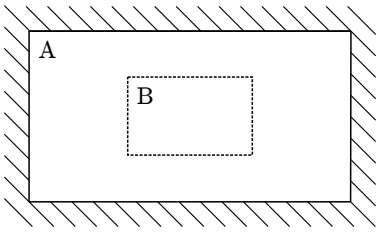

FIGURE 6.3. The thermodynamic system A is assumed to be isolated, and the closed subsystem B is separated from the surroundings $A - B$ by a heat-conductive barrier.

from a physical standpoint is not so obvious. Finally, there remains the problem of the construction of the partition of phase space. Consider two different translations of the partition used to calculate the coarse-grained entropy of the probability distribution in Fig. 6.2. The calculated coarse-grained entropies would be different for each translation in general, making the apparent thermodynamic properties of the system effectively dependent on an arbitrary choice of the origin.

The second proposed resolution instead modifies the Liouville equation by introducing a diffusive term, with the result that an initial probability distribution spreads over time and Gibbs's entropy in Eq. 6.1 increases [**14**, **88**, **112**]. The usual justification offered is that it is not actually possible to construct an isolated system, and the diffusive term reflects the effectively stochastic influence of the surroundings. The consequences of this argument can be seen by considering the isolated system A in Fig. 6.3; it should always be possible to construct such a system since, e.g., the universe is isolated by definition. Let the initial probability distribution of microstates of A be such that the entropy $S_{\mathrm{G}}(A)$ is equal to the minimum possible for any thermodynamic system. Since A does not interact with any surroundings, the probability distribution of microstates of A obeys the standard Liouville equation and $S_{\mathrm{G}}(A)$ remains constant by Liouville's theorem. If a subsystem $B \subset A$ is closed but not isolated, the probability distribution of microstates of B obeys the modified Liouville equation, and the entropy $S_{\mathrm{G}}(B)$ gradually increases to a limiting value. Experimental thermodynamics requires that the entropy be additive (up to the leading terms) in the sense that $S_{\mathrm{G}}(A) = S_{\mathrm{G}}(B) + S_{\mathrm{G}}(A - B)$ though, and if $S_{\mathrm{G}}(B) > S_{\mathrm{G}}(A)$, then $S_{\mathrm{G}}(A - B) < S_{\mathrm{G}}(A)$ must be less than the absolute minimum of entropy. That is, the proposed modification of the Liouville equation apparently violates the third law of thermodynamics.

Proponents of coarse-graining could suggest that a regular tessellation of the phase space by hypercubes be used as a canonical partition, and perhaps average over all possible translations of the tessellation. Proponents of the modified Liouville equation could object that the universe is not isolated, or that the situation in the paragraph above could not occur in practice. There are many other proposed answers to the question of entropy generation in the literature as well. The purpose of this section is not to rigorously evaluate these proposals, but merely to point out that they generally seem to require additional development before they could be entirely satisfactory.

6.2. Fundamental Thermodynamic Relation

The development of classical thermodynamics begins with the fundamental thermodynamic relation for open systems

$$\mathrm{d}U = T\mathrm{d}S + V\sum_{ij}\sigma_{ij}\mathrm{d}\epsilon_{ij} + \sum_{i}\mu_i\mathrm{d}N_i. \tag{6.2}$$

The purpose of this section is to propose suitable definitions of the extrinsic quantities in Eq. 6.2 that can be applied to finite non-equilibrium systems and that reduce to the classical results in the equilibrium limit. This necessarily includes a definition of the entropy that is consistent with the second and third laws of thermodynamics. Our intention is to handle the intrinsic quantities in a separate publication.

It is useful to define precisely what properties are expected of an extensive quantity, particularly since the literature is surprisingly inconsistent on this point [**92**, **123**]. Extensive quantities have been variously defined as those that are equal to the sum of the corresponding values for the constituent subsystems (additive), as those that are proportional to a system's mass (proportional), or as those that are are both additive and proportional. These definitions are not all equivalent, and are not necessarily sensible for non-equilibrium systems. Consider that the volume of a system composed of two subsystems with different densities but the same number of atoms (e.g., a liquid and a solid) is additive but not proportional. The gravitational potential energy of a system increases superlinearly in the number of atoms, implying that the internal energy is not proportional either.

Classical definitions of extensivity apparently make several assumptions that are usually left implicit. First, the extent of the interactions between any two subsystems should be at most proportional to the shared surface area. Second, passing to the thermodynamic limit causes each subsystem's volume to increase faster than its surface area, making interactions between subsystems negligible in the limit. A classically extensive quantity is here defined to be one that is additive in such conditions. Notice that the second assumption is not applicable to the finite systems considered here.

Two types of additivity are defined for finite systems that are equivalent to the classical definition in the conditions described above. A quantity is said to be *strongly additive* if its value for a union of arbitrary subsystems is equal to the sum of its values for those subsystems. Strong additivity is stronger than classical additivity since it does not require either of the assumptions above, and is often a property of quantities that are defined on a per-atom basis. A quantity is said to be *weakly additive* if its value for a union of isolated subsystems (defined in Sec. 6.2.1) is equal to the sum of its values for those subsystems. Notice that the assumptions above effectively isolate each subsystem, making strong and weak additivity equivalent in classical conditions.

The quantities defined in the sections below are said to be strongly or weakly extensive depending on whether they are strongly or weakly additive; this will be found to depend sensitively on the way the subsystems' interactions are handled. Finally, it is worthwhile to briefly discuss whether one is preferable to the other. Certainly they are both consistent with macroscale experimental thermodynamics, and so one does not seem to be preferable to the other on that account. It is true that strong extensivity is mathematically convenient, but it is not at all clear that it should be possible to objectively distribute, e.g., the energy of an interaction over the various subsystems participating in that interaction. Observe that since the energy in question arises from interacting subsystems, the part of the energy assigned to a particular subsystem cannot be a property of that subsystem alone. Indeed, the appealing idea that a subsystem's internal energy should in some way reside within that subsystem was an essential part of the intuitive but long-discredited caloric theory [**101**].

6.2.1. Boundary Conditions. Given that one of our motivations is to evaluate thermodynamic quantities during molecular dynamics simulations, it is useful to consider the ways in which

such a system could be defined. Since molecular dynamics simulations are constructed to conserve internal energy, volume, and particle number unless specifically specified otherwise, the simplest simulation would be of an *isolated* system for which all of these quantities are constant and Eq. 6.2 is the natural governing equation.

If instead the system is *closed* and can exchange only thermal energy with a reservoir, then that is precisely what should be simulated. This would entail a molecular dynamics simulation of a large isolated system A partitioned into a system of interest B and a surrounding reservoir $A - B$. B would need to be separated from $A - B$ by a solid heat-conductive barrier composed of atoms with a high binding energy. Notice that the usual practice of neglecting boundary effects in the thermodynamic limit is not valid for finite systems, and that thermodynamic quantities should include the effects of the system's particles interacting with those of the barrier. The use of a thermostat is inadvisable since it is not yet clear that the definition of temperature for an equilibrium system is relevant to a non-equilibrium one.

For an *open* system that can exchange thermal energy and particles with a reservoir, the procedure would be similar but with the barrier removed. This raises the question of what precisely constitutes the system of interest B and the surrounding reservoir $A-B$ though. Since the particles in B are not fixed, B would presumably be defined so as to include the particles residing within a particular region of space at any instant in time. Notice that defining the thermodynamic properties of B using the probability distribution of microstates of B leads to difficulty since the dimension of the phase space would change discontinuously with the number of particles in B. For this reason, the thermodynamic properties of B should always be defined using the probability distribution of microstates of A.

The only change necessary to allow a closed system to exchange volume with a reservoir would be to make the heat-conductive barrier deformable. It is unclear whether a barrier could be made deformable while disallowing the transfer of thermal energy for a system that can exchange only volume with a reservoir though. A simulation could be performed by subjecting the system to a barostat, but it is not yet clear that the definition of pressure for an equilibrium system is relevant to a non-equilibrium one.

There is not an obvious way to define what constitutes an open system that is able to exchange volume with a reservoir, since neither the particles in B nor the region assigned to B are fixed. Other possible boundary conditions that involve the exchange of particles but not thermal energy with a reservoir are regarded as unphysical.

6.2.2. Internal Energy. The internal energy is the first of the extrinsic quantities appearing in Eq. 6.2, and is defined as the sum of the potential and kinetic energies

$$U(\bar{x}, \bar{p}) = U_x(\bar{x}) + U_p(\bar{p})$$

where $\bar{x}$ and $\bar{p}$ are vectors of all the position and momentum variables describing the microstate of the system. Since energy is not an absolute quantity, the zeros of the potential and kinetic energies need to be defined. For atomic systems with short-range interactions, the potential energy is usually measured with respect to a microstate where the atoms are infinitely separated and in the absence of external fields, and the kinetic energy is measured with respect to a center of momentum frame of the system. While $U_x(\bar{x})$ is usually a complicated function of the position variables, $U_p(\bar{p})$ is simply the sum of the kinetic energies of the constituent atoms.

The internal energy above is a function on the phase space that assigns a value to a particular microstate, but the expectation value $\langle U \rangle = \langle U_x \rangle + \langle U_p \rangle$ of the practitioner depends on the probability distribution of microstates $\rho(\bar{x}, \bar{p})$. Since the potential and kinetic energies respectively depend only on position and momentum variables, their expectation values can be calculated as

$$\langle U_x \rangle = \int_{\Gamma_x} U_x(\bar{x}) \rho_x(\bar{x}) \mathrm{d}\bar{x}$$
$$\langle U_p \rangle = \int_{\Gamma_p} U_p(\bar{p}) \rho_p(\bar{p}) \mathrm{d}\bar{p}$$

where $\rho_x = \int_{\Gamma_p} \rho \, \mathrm{d}\bar{p}$ and $\rho_p = \int_{\Gamma_x} \rho \, \mathrm{d}\bar{x}$ are the marginal probability distributions on the configuration and momentum spaces Γ_x and Γ_p. The change in internal energy is the more physically relevant quantity though. Suppose that the probability distribution ρ is changed into the probability distribution $\rho + \delta\rho$, where $\delta\rho(\bar{x}, \bar{p})$ is arbitrarily small with respect to the L^1 norm and is required to satisfy $\int_\Gamma \delta\rho \, \mathrm{d}\Omega = 0$ for the probability distribution to remain normalized. The resulting change $\mathrm{d}\langle U \rangle = \mathrm{d}\langle U_x \rangle + \mathrm{d}\langle U_p \rangle$ in the expectation value of the internal energy is the sum of the changes in

the expectation values of the potential and kinetic energies, or

$$\mathrm{d}\langle U_x\rangle = \int_{\Gamma_x} U_x(\bar{x})\delta\rho_x(\bar{x})\mathrm{d}\bar{x}$$
$$\mathrm{d}\langle U_p\rangle = \int_{\Gamma_p} U_p(\bar{p})\delta\rho_p(\bar{p})\mathrm{d}\bar{p}$$

where $\delta\rho_x = \int_{\Gamma_p} \delta\rho\,\mathrm{d}\bar{p}$ and $\delta\rho_p = \int_{\Gamma_x} \delta\rho\,\mathrm{d}\bar{x}$ are defined analogously to ρ_x and ρ_p. This is in principle enough to define the change in internal energy of an isolated system.

Physical systems are rarely isolated though. Even noble gas atoms in a rigid container interact by dispersion forces with the atoms in the container walls, and there should be a procedure to partition the potential energy of such interactions into two parts, one pertaining to the system and the other to the surroundings. More generally, let a system A be partitioned into n subsystems A^i where $i \in [1, n]$. A point in the phase space Γ defines the microstate of A and therefore the microstates of all A^i. This implies that there should exist functions $U_x^i(\bar{x})$ and $U_p^i(\bar{p})$ defined on Γ that indicate the potential and kinetic energies of A^i and that have expectation values defined analogously to $\langle U_x\rangle$ and $\langle U_p\rangle$. Moreover, the extensivity of the internal energy requires that these functions satisfy $U_x(\bar{x}) = \sum_i U_x^i(\bar{x})$ and $U_p(\bar{p}) = \sum_i U_p^i(\bar{p})$.

A strongly extensive $U_p^i(\bar{p})$ can be defined by assigning to A^i the kinetic energies of all the atoms in A^i, or

$$U_p^i(\bar{p}) = \sum_{j\in\mathcal{L}^i} \|\bar{p}_j\|^2/(2m_j) \tag{6.3}$$

where $\mathcal{L}^i$ is the set of labels of the atoms in A^i and $\bar{p}_j$ is the momentum of the jth atom. Defining $U_x^i(\bar{x})$ is more difficult though. Let N be the number of atoms in A, and $\mathcal{S}$ be the set of all subsets of labels of atoms in A except for the empty set. Formally, the potential energy of A can be represented as a sum of many-body interactions $U_x(\bar{x}) = \sum_{s\in\mathcal{S}} \phi_s(\bar{x})$ where $\phi_s(\bar{x})$ is the interaction among the atoms with labels in s [**64**]. A reasonable definition for $U_x^i(\bar{x})$ is then

$$U_x^i(\bar{x}) = \sum_{s\in\mathcal{S}} \frac{|\mathcal{L}^i \cap s|}{|s|}\phi_s(\bar{x}) \tag{6.4}$$

where the contribution of an interaction to the potential energy of a subsystem is proportional to the number of atoms of the subsystem that participate in the interaction. It is straightforward to show that this definition is strongly extensive.

The difficulty with this approach is that it is not clear whether the potential energy can always be decomposed into a sum of many-body interactions in practice; electronic structure calculations indicate that the presence of one atom can affect the electron distributions of surrounding atoms in complicated ways. That said, it is always possible to measure the potential energy change of removing an atom (i.e., displacing the atom to infinity). The question is therefore whether such information is sufficient to construct $U_x^i(\bar{x})$ as in the definition above.

Observe that $\mathcal{S}$ not only indexes the many-body interactions, but also the sets of atoms that could be removed from the system. Let $\chi_r(\bar{x})$ be the potential energy change of A when removing the atoms with labels in $r \in \mathcal{S}$. By construction, $\chi_r(\bar{x})$ is the sum of all many-body interactions that involve any atoms with labels in r, and is written as $\chi_r(\bar{x}) = \sum_{s\in\mathcal{S}} a_{rs}\phi_s(\bar{x})$ where $a_{rs} \in \{-1, 0\}$. The matrix with elements a_{rs} is invertible as shown by the following argument. The atom removal event that includes all of the atoms of A gives the potential energy of A. Adding the potential energy of A to that of events that remove all but one atom gives the one-body interactions. Adding the potential energy of A to that of events that remove all but two atoms and subtracting the appropriate one-body interactions gives the two-body interactions. Repeating this process with increasing numbers of remaining atoms shows that each atom removal event introduces precisely one multi-body interaction that is not already known. The rows of the matrix with elements a_{rs} are therefore linearly independent, and since the matrix is square, it is invertible. If b_{sr} are the elements of the inverse matrix, the multi-body interactions can be found by $\phi_s(\bar{x}) = \sum_{r\in\mathcal{S}} b_{sr}\chi_r(\bar{x})$. Substituting this into Eq. 6.4 gives

$$U_x^i(\bar{x}) = \sum_{s\in\mathcal{S}} \frac{|\mathcal{L}^i \cap s|}{|s|} \sum_{r\in\mathcal{S}} b_{sr}\chi_r(\bar{x}) \tag{6.5}$$

An example of this procedure is given in Sec. 6.4.

Since the number of multi-body interactions increases as 2^N, the definition in Eq. 6.5 could be difficult to use in practice when the number of atoms is not sufficiently small. As an alternative, the potential energy of a multi-body interaction could be distributed equally over the subsystems

that participate in the interaction. Let S_n be the set of all permutations of n elements, $\sigma \in S_n$ be a permutation that indicates an ordering of the subsystems, and consider the process that begins with the system A and successively removes all atoms in $A^{\sigma(j)}$ for $j \in [1, n]$. Let $\psi^i_\sigma(\bar{x})$ be the potential energy change when the atoms in A^i are removed as part of this process. Then

$$U^i_x(\bar{x}) = \frac{1}{n!} \sum_{\sigma \in S_n} \psi^i_\sigma(\bar{x}) \tag{6.6}$$

is a weakly extensive definition for the potential energy of A^i. A proof that this effectively distributes many-body interactions equally over the participating subsystems is given in Sec. 6.5. The advantage of this definition is that if an isolated system A is partitioned into a subsystem A^1 of interest and the surroundings A^2, the equation for $U^1_x(\bar{x})$ has only two terms and Eq. 6.6 could be evaluated in practice.

One property of these definitions is that if every atom of A is assigned to a separate subsystem, Eqs. 6.5 and 6.6 give identical and well-defined potential energies for individual atoms. The alternatives in the literature instead require the careful definition and justification of volumes associated with individual atoms over which an energy density is integrated [**4**, **134**].

6.2.3. Entropy. If the definitions for the extrinsic quantities in the fundamental thermodynamic relation are to reduce to the classical ones in the equilibrium limit, then the entropy is likely to be defined by an equation resembling Gibbs's formula in Eq. 6.1. Section 6.1.3 describes why this appears to not allow entropy generation during irreversible processes as required by the second law of thermodynamics though. This suggests that the entropy formula be modified, though it is useful to initially consider the symmetries that the entropy is expected to satisfy. Certainly the entropy, like all other thermodynamic quantities, should be invariant to the choice of coordinate system. The entropy in Eq. 6.1 not only has this property, but is actually invariant to all canonical transformations (ones that preserve the form of Hamilton's equations of motion) [**59**, **83**]. These include transformations that, e.g., exchange coordinates with momenta, though since other thermodynamic variables are not expected to be invariant to such transformations, it is possible that the entropy in Eq. 6.1 is more symmetric than is necessary.

Following this line of thought, it is curious that position and momentum variables enter equivalently into Eq. 6.1, simply as variables over which to integrate. They are certainly not equivalent in the context of Lagrangian mechanics, nor in that of quantum mechanics [**84**,**116**]. Indeed, quantum mechanics even seems to suggest that the marginal probability distributions ρ_x and ρ_p are more fundamental objects than the joint probability distribution ρ. Suppose then that the entropy is defined as a sum of configurational and vibrational entropies

$$\langle S \rangle = \langle S_x \rangle + \langle S_p \rangle. \tag{6.7}$$

Applying Gibbs's formula to the marginal probability distributions ρ_x and ρ_p allows these to be explicitly defined as

$$\begin{aligned} \langle S_x \rangle &= -k_{\mathrm{B}} \int_{\Gamma_x} \rho_x(\bar{x}) \ln \frac{\rho_x(\bar{x})}{\mu_x} \mathrm{d}\bar{x} \\ \langle S_p \rangle &= -k_{\mathrm{B}} \int_{\Gamma_p} \rho_p(\bar{p}) \ln \frac{\rho_p(\bar{p})}{\mu_p} \mathrm{d}\bar{p}. \end{aligned}$$

The quantum entropic uncertainty principle [**8**, **12**] states that, for suitable values of μ_x and μ_p, the entropy in Eq. 6.7 can be made nonnegative for any ρ_x and ρ_p that are admissible on the basis of quantum mechanics; while the intention is to develop a classical theory, it is satisfying that the entropy remains sensible even in the low-temperature limit where quantum effects are likely to be significant.

The entropies in Eqs. 6.1 and 6.7 are equivalent whenever $\bar{x}$ and $\bar{p}$ are independent random variables, or $\rho = \rho_x \rho_p$, since

$$\begin{aligned} -\frac{\langle S \rangle}{k_{\mathrm{B}}} &= \int_{\Gamma_x} \rho_x(\bar{x}) \ln \frac{\rho_x(\bar{x})}{\mu_x} \mathrm{d}\bar{x} + \int_{\Gamma_p} \rho_p(\bar{p}) \ln \frac{\rho_p(\bar{p})}{\mu_p} \mathrm{d}\bar{p} \\ &= \int_{\Gamma} \rho(\bar{x}, \bar{p}) \ln \frac{\rho_x(\bar{x})}{\mu_x} \mathrm{d}\Omega + \int_{\Gamma} \rho(\bar{x}, \bar{p}) \ln \frac{\rho_p(\bar{p})}{\mu_p} \mathrm{d}\Omega \\ &= \int_{\Gamma} \rho(\bar{x}, \bar{p}) \ln \frac{\rho_x(\bar{x}) \rho_p(\bar{p})}{\mu_x \mu_p} \mathrm{d}\Omega. \end{aligned} \tag{6.8}$$

Significantly, the Gibbs measure for the canonical distribution has this property, and can be written as the product of marginal measure on the configuration space and a Maxwell–Boltzmann distribution on the momentum space. The definition in Eq. 6.7 is therefore equivalent to that in Eq.

6.1 for the canonical ensemble, and by the principle of the equivalence of ensembles [**43**, **124**], is consistent with effectively all prior thermodynamic results in the equilibrium thermodynamic limit. This includes the Sackur–Tetrode equation [**114**, **120**], one of the few explicit predictions for the entropy of a thermodynamic system. That said, and despite the entropy often being separated into configurational and vibrational parts in the literature, the two definitions of entropy are not equivalent in general; their difference is proportional to the mutual information of $\bar{x}$ and $\bar{p}$ (here the mutual information of two disjoint sets of random variables is defined analogously to that of two random variables [**31**]).

With regard to the second law, the entropy in Eq. 6.7 generically increases with time for a non-equilibrium isolated system essentially as a result of projective geometry. Since the Lyapunov exponents of a Hamiltonian system occur in pairs $(\lambda, -\lambda)$ [**1**, **33**], a generic initial probability distribution expands exponentially in some directions and contracts exponentially in others. Though Liouville's theorem indicates that the overall volume of this region is constant in time, the projected volumes on the configuration and momentum spaces are not, as can be seen by considering the projected shadow of an immiscible fluid stirred into water. More precisely, the evolution of ρ as prescribed by the Liouville equation generically causes the mutual information of $\bar{x}$ and $\bar{p}$ and therefore the entropy in Eq. 6.7 to increase with time up to some limiting value when the system is said to have reached equilibrium.

The notion of a generic initial probability distribution should be examined further. If the entropy increase required by the second law is provided by an increase in mutual information, then a decrease of entropy with time as in Loschmidt's paradox is not strictly forbidden [**18**,**90**]. Consider any initial probability distribution ρ_0 for which the entropy increases in time. After some period of time τ has passed and ρ_0 has evolved to ρ_1, transform ρ_1 to ρ_1' by mapping each point in the phase space to the corresponding point with the particle velocities reversed. The evolution of ρ_1' would then proceed back to ρ_0' with decreasing entropy. That is, for each initial probability distribution ρ_0 for which entropy increases over τ, there is a corresponding initial probability distribution ρ_1' for which entropy decreases by the same amount over τ. Apparently, a generic initial probability distribution does not refer to most probability measures on the phase space in the usual sense.

Recall the conclusion of Sec. 6.1.2 that the probability distribution reflects the practitioner's uncertainty about the system's microstate; perhaps a generic probability distribution instead refers to those that a practitioner could feasibly specify as an initial condition. Certainly probability distributions with little mutual information (e.g., where $\bar{x}$ and $\bar{p}$ are independent random variables) are simpler to write down and to reason about, but the more fundamental difficulty relates to the possible sources of information about correlations of position and momentum variables. It seems unlikely that the practitioner would be able to either prepare an initial condition with highly correlated position and momentum variables or measure such correlations in an evolving system. Again consider stirring an immiscible fluid into water. Certainly the most likely initial condition from a practical standpoint is one where the immiscible fluid is a roughly spherical droplet, i.e., where the marginal mass distributions along orthogonal directions are nearly uncorrelated. Moreover, once stirring begins the distribution of mass quickly becomes so complicated that the details are practically unavailable, and the only quantity that can reasonably be measured is the mass density. This suggests that the resolution to Loschmit's paradox is that the available initial conditions do not include all possible measures on the phase space due to the limitations of human capabilities and human knowledge.

With regard to symmetries, the entropy defined in Eq. 6.7 is invariant to coordinate transformations but not to all canonical transformations. This would not be true if the reference quantities μ_x and μ_p were not included with the appropriate units of inverse volume though, since then the absolute value of the Jacobian determinant would appear inside the logarithms during a coordinate transformation. The presence of μ_x and μ_p is therefore essential, and their values can be determined by requiring that the entropy vanish at the absolute zero of temperature. For simplicity, consider the entropy of a perfect crystal in the Einstein model [**42**, **91**]. Let the crystal contain N atoms of mass m and have a potential energy of $U_x(\bar{x}) = \frac{1}{2}m\omega^2 \sum_i |\bar{x}_i - \bar{y}_i|^2$, where $\bar{x}_i$ and $\bar{y}_i$ are the absolute position and minimum potential energy position of the ith atom. The marginal

probability distributions in the ground state can then be shown to be [**84**, **116**]

$$\rho_x(\bar{x}) = \left(\frac{m\omega}{\pi\hbar}\right)^{3N/2} \exp\left(-\frac{m\omega}{\hbar}\sum_j |\bar{x}_j - \bar{y}_j|^2\right)$$
$$\rho_p(\bar{p}) = \left(\frac{1}{\pi\hbar m\omega}\right)^{3N/2} \exp\left(-\frac{1}{\hbar m\omega}\sum_j |\bar{p}_j|^2\right).$$

The third law of thermodynamics is interpreted as requiring that the corresponding configurational and vibrational entropies vanish independently. Further assuming that μ_x and μ_p are constant functions on their respective spaces gives

$$\mu_x = \left(\frac{2m\omega}{he}\right)^{3N/2}$$
$$\mu_p = \left(\frac{2}{hm\omega e}\right)^{3N/2}$$

where e is Euler's number. While these depend on the choice of reference crystal through m and ω, there is not an obvious physical reason why the entropy of all reference crystals should be the same at the absolute zero of temperature. Certainly one does not expect, e.g., the heat capacities of all reference crystals to be the same in the low-temperature limit. The appearance of Plank's constant could be argued as being inconsistent with a classical theory, but since μ_x and μ_p are defined by systems at the absolute zero of temperature where quantum effects are significant, this is not regarded as a serious objection. Finally, the dependence on N is necessary for μ_x and μ_p to have the appropriate units of inverse volume in the configuration and momentum spaces.

The derivation of the equation for the entropy change $\mathrm{d}\langle S\rangle = \mathrm{d}\langle S_x\rangle + \mathrm{d}\langle S_p\rangle$ is slightly more involved than the one for the internal energy since the entropy depends nonlinearly on the probability distribution ρ. Consider the configurational entropy of a system with probability distribution $\rho + \delta\rho$:

$$\begin{aligned}
-\frac{\langle S_x\rangle'}{k_\mathrm{B}} &= \int_{\Gamma_x} (\rho_x + \delta\rho_x) \ln\frac{\rho_x + \delta\rho_x}{\mu_x}\mathrm{d}\bar{x} \\
&= \int_{\Gamma_x} \left[\rho_x \ln\frac{\rho_x}{\mu_x} + \delta\rho_x \ln\frac{\rho_x}{\mu_x} + \delta\rho_x + \frac{(\delta\rho_x)^2}{\rho_x}\right]\mathrm{d}\bar{x} \\
&= -\frac{\langle S_x\rangle}{k_\mathrm{B}} + \int_{\Gamma_x} \delta\rho_x \ln\frac{\rho_x}{\mu_x}\mathrm{d}\bar{x}.
\end{aligned}$$

The second line uses a first-order Taylor series expansion of the logarithm about ρ_x/μ_x, and the third line only retains terms to first order in $\delta\rho_x$. This allows the changes in the configurational and vibrational entropies to be written as

$$\mathrm{d}\langle S_x\rangle = -k_\mathrm{B}\int_{\Gamma_x}\delta\rho_x(\bar{x})\ln\frac{\rho_x(\bar{x})}{\mu_x}\mathrm{d}\bar{x}$$
$$\mathrm{d}\langle S_p\rangle = -k_\mathrm{B}\int_{\Gamma_p}\delta\rho_p(\bar{p})\ln\frac{\rho_p(\bar{p})}{\mu_p}\mathrm{d}\bar{p}.$$

Curiously, $\mathrm{d}\langle S_x\rangle$ and $\mathrm{d}\langle S_p\rangle$ are independent of μ_x and μ_p despite their appearance in the equations above. Observe that, e.g., multiplying μ_x by an arbitrary constant a is equivalent to subtracting $\ln a\int_{\Gamma_x}\delta\rho_x\mathrm{d}\bar{x}$ from the first integral, but this vanishes since $\int_{\Gamma_x}\delta\rho_x\mathrm{d}\bar{x}=0$. This indicates that while μ_x, μ_p, and the absolute entropy might depend on the choice of reference crystal, the validity of the fundamental thermodynamic relation does not.

There is still the issue of the extensivity of the entropy and the frequent division of the partition function by a factor of $N!$ to account for the indistinguishability of identical particles. This practice is apparently motivated by a thought experiment proposed by Gibbs [**57**]. Consider two chambers of the same volume V separated by a removable barrier, each containing N atoms of an ideal monatomic gas at the same temperature. The ρ_x of a single chamber is a constant function over a volume V^N, with the corresponding configurational entropy

$$\langle S_x\rangle = k_\mathrm{B}\ln(V^N\mu_x^{(N)})$$

where the superscript on μ_x indicates the number of atoms for which the constant is defined. Removing the barrier results in a single system of volume $2V$ with $2N$ atoms for which ρ_x is a constant function over a volume $(2V)^{2N}$. The configurational entropy after removing the barrier is therefore

$$\langle S_x\rangle' = 2k_\mathrm{B}\ln(V^N\mu_x^{(N)}) + 2k_\mathrm{B}N\ln 2$$

where μ_x was squared to account for the change in the dimension of the configuration space. The change in configurational entropy $\Delta\langle S_x\rangle = \langle S_x\rangle' - 2\langle S_x\rangle$ is then

$$\Delta\langle S_x\rangle = 2k_\mathrm{B}N\ln 2$$

which is inconsistent with the second law; since the chambers were initially in equilibrium, the removal of the barrier should be reversible and the entropy change should be zero.

This paradox is usually resolved by postulating that the N gas atoms in a chamber are identical and indistinguishable, reducing the volume of distinct states in the configuration space to $V^N/N!$ instead of V^N [**68**,**87**]. Repeating the steps above with this modification gives the expected $\Delta\langle S_x\rangle \approx 0$ (to a precision limited by Stirling's approximation). This treatment is unsatisfactory for at least three reasons. First, the Liouville equation does not indicate what happens when the dimension of the phase space changes. More precisely, removing the barrier transforms the phase space from a disjoint union of those of the two chambers into a product space of the same, a troublingly discontinuous change given that the physical system evolves continuously. Second, if the physical properties of the gas atoms in one chamber could be varied continuously, the equation for the entropy would change discontinuously at the point when the gases are first distinguishable. This seems less preferable to assigning the discontinuity to the practitioner's uncertainty about the microstate, as was done with the mixing paradox in Sec. 6.1.2. Third, the modern justification for the indistinguishability of identical atoms relies on quantum mechanical principles, and a classical statistical thermodynamics would ideally depend on quantum mechanics as little as possible.

The change in the dimension of the phase space can be addressed by instead considering the two chambers as parts of a single thermodynamic system at all points in time. If, e.g., the first chamber is known to hold the first N atoms and the second chamber the second N atoms, then ρ_x is the product of two independent probability distributions for the atoms in each of the two chambers. This makes ρ_x a constant function over a volume V^{2N}, and the initial configurational entropy

$$\langle S_x\rangle = k_\mathrm{B}\ln(V^{2N}\mu_x^{(2N)}).$$

Removing the barrier gives the same ρ_x and $\langle S_x\rangle'$ as above (up to the change in the definition of μ_x), with the result that the change in configurational entropy $\Delta\langle S_x\rangle = \langle S_x\rangle' - \langle S_x\rangle$ is

$$\Delta\langle S_x\rangle = 2k_\mathrm{B}N\ln 2.$$

An increase in the configurational entropy is expected though, since removing the barrier irreversibly increases the practitioner's uncertainty about the microstate; before removing the barrier the first

N atoms were known to be in the first chamber and the second N atoms in the second chamber, but afterwards each atoms has equal probability of being in either chamber.

While justifiable, this is inconsistent with the expectation that a practitioner would not observe any measurable change in the macrostate on removing and reinserting the barrier. The source of the inconsistency is not the equation for the entropy though, but rather that the initial ρ_x does not accurately reflect the practitioner's uncertainty about the microstate. By hypothesis the practitioner knows that the system begins with N atoms in the first chamber and N atoms in the second, but any permutation of the $2N$ atoms would result in an experimentally indistinguishable system in all respects. The initial ρ_x should therefore be a constant function over $(2N)!/(N!)^2$ copies of a volume V^{2N}, one for each binning of the atoms. The corresponding initial configurational entropy is

$$\begin{aligned}\langle S_x \rangle &= k_{\mathrm{B}} \ln(V^{2N} \mu_x^{(2N)}) + k_{\mathrm{B}} \ln \frac{(2N)!}{(N!)^2} \\ &\approx k_{\mathrm{B}} \ln(V^{2N} \mu_x^{(2N)}) + 2k_{\mathrm{B}} N \ln 2\end{aligned}$$

where the second line uses Stirling's approximation, with the consequence that $\Delta\langle S_x \rangle \approx 0$. That is, the extensivity of the entropy does not require that identical atoms actually be indistinguishable (a postulate not obviously compatible with a classical theory), but merely that they be effectively indistinguishable as a consequence of the limited knowledge of the practitioner.

Following the discussion in Sec. 6.2.1, it would be useful to be able to distribute the entropy over several subsystems while maintaining extensivity. As with the internal energy in Sec. 6.2.2, suppose that a system A is partitioned into n subsystems A^i where $i \in [1, n]$. Let $\mathcal{T}$ be the set of all assignments of the N atoms in A to the n subsystems, and associate with the assignment $t \in \mathcal{T}$ the part of the configuration space in which all the atoms are in the indicated subsystems. Given assignment t, let N_t^i be the number of atoms in A^i, and $\bar{x}_t^i$ and $\bar{p}_t^i$ be the sets of position and momentum variables for the atoms in A^i. Starting with Eq. 6.8, the contribution of the region

associated with assignment t to the overall entropy is

$$
\begin{aligned}
-\frac{\langle S\rangle_t}{k_{\mathrm{B}}} &= \int_{\Gamma_p}\int_{A^n}\cdots\int_{A^1}\rho(\bar{x},\bar{p})\ln\frac{\rho_x(\bar{x})\rho_p(\bar{p})}{\mu_x\mu_p}\mathrm{d}\bar{x}_t^1\ldots\mathrm{d}\bar{x}_t^n\mathrm{d}\bar{p}\\
&= \int_{A^n}\cdots\int_{A^1}\rho_x(\bar{x})\ln\frac{\rho_x(\bar{x})}{\mu_x}\mathrm{d}\bar{x}_t^1\ldots\mathrm{d}\bar{x}_t^n\\
&\quad + \int_{\Gamma_p}\left[\int_{A^n}\cdots\int_{A^1}\rho(\bar{x},\bar{p})\mathrm{d}\bar{x}_t^1\ldots\mathrm{d}\bar{x}_t^n\right]\ln\frac{\rho_p(\bar{p})}{\mu_p}\mathrm{d}\bar{p}
\end{aligned}
$$

where the domain of integration for each of the N_t^i atoms in A^i is A^i. The additive decomposition on the second line suggests the definitions

$$
\begin{aligned}
\langle S_x\rangle_t &= -k_{\mathrm{B}}\int_{A^n}\cdots\int_{A^1}\rho_x(\bar{x})\ln\frac{\rho_x(\bar{x})}{\mu_x}\mathrm{d}\bar{x}_t^1\ldots\mathrm{d}\bar{x}_t^n\\
\langle S_p\rangle_t &= -k_{\mathrm{B}}\int_{\Gamma_p}\left[\int_{A^n}\cdots\int_{A^1}\rho(\bar{x},\bar{p})\mathrm{d}\bar{x}_t^1\ldots\mathrm{d}\bar{x}_t^n\right]\ln\frac{\rho_p(\bar{p})}{\mu_p}\mathrm{d}\bar{p}
\end{aligned}
\tag{6.9}
$$

for the contributions of the assignment t to the configurational and vibrational entropies of A.

The procedure to distribute the quantities $\langle S_x\rangle_t$ and $\langle S_p\rangle_t$ over the subsystems A^i is motivated by the chain rule for conditional entropies [**31**]. Let S_n be the set of all permutations of n elements, $\sigma\in S_n$ be a permutation that indicates an ordering of the subsystems, and q be the number of elements before i in σ. Let

$$
\begin{aligned}
\alpha_t^i(\bar{x}|\sigma) &= \int_{\Gamma_x}\cdots\int_{\Gamma_x}\rho_x(\bar{x})\mathrm{d}\bar{x}_t^{\sigma(1)}\ldots\mathrm{d}\bar{x}_t^{\sigma(q)}\\
\beta_t^i(\bar{p}|\sigma) &= \int_{\Gamma_p}\cdots\int_{\Gamma_p}\rho_p(\bar{p})\mathrm{d}\bar{p}_t^{\sigma(1)}\ldots\mathrm{d}\bar{p}_t^{\sigma(q)}
\end{aligned}
$$

be the marginal probability distributions for atoms in subsystems with labels in the final $n-q$ elements of σ. The quantities

$$
\begin{aligned}
\zeta_t^i(\bar{x}) &= \frac{1}{n!}\sum_{\sigma\in S_n}\ln\left[\frac{\alpha_t^i(\bar{x}|\sigma)}{\mu_x^{(N_t^i)}\int_{\Gamma_x}\alpha_t^i(\bar{x}|\sigma)\mathrm{d}\bar{x}_t^i}\right]\\
\eta_t^i(\bar{p}) &= \frac{1}{n!}\sum_{\sigma\in S_n}\ln\left[\frac{\beta_t^i(\bar{p}|\sigma)}{\mu_p^{(N_t^i)}\int_{\Gamma_p}\beta_t^i(\bar{p}|\sigma)\mathrm{d}\bar{p}_t^i}\right]
\end{aligned}
$$

effectively average the conditional entropies of the atoms in A^i over all possible orderings of the subsystems. This gives

$$\langle S_x^i \rangle_t = -k_\mathrm{B} \int_{A^n} \dots \int_{A^1} \rho_x(\bar{x}) \zeta_t^i(\bar{x}) \mathrm{d}\bar{x}_t^1 \dots \mathrm{d}\bar{x}_t^n$$

$$\langle S_p^i \rangle_t = -k_\mathrm{B} \int_{\Gamma_p} \left[\int_{A^n} \dots \int_{A^1} \rho(\bar{x}, \bar{p}) \mathrm{d}\bar{x}_t^1 \dots \mathrm{d}\bar{x}_t^n \right] \eta_t^i(\bar{p}) \mathrm{d}\bar{p} \tag{6.10}$$

for the configurational and vibrational entropies of A^i for assignment t. Observe that summing $\langle S_x^i \rangle_t$ and $\langle S_p^i \rangle_t$ over all subsystems gives $\langle S_x \rangle_t$ and $\langle S_p \rangle_t$ as defined in Eq. 6.9. Finally,

$$\langle S^i \rangle = \sum_{t \in \mathcal{T}} (\langle S_x^i \rangle_t + \langle S_p^i \rangle_t). \tag{6.11}$$

is defined as the entropy of subsystem A^i. This definition is weakly extensive; since the microstates of isolated subsystems are necessarily independently distributed, $\zeta_t^i(\bar{x})$ and $\eta_t^i(\bar{p})$ reduce to functions of $\bar{x}_t^i$ and $\bar{p}_t^i$ alone, and $\langle S_x^i \rangle_t$ and $\langle S_p^i \rangle_t$ are weakly additive.

The procedure above is for subsystems that are open and defined by particular regions of space. If the subsystems are instead defined by sets of atoms, the only change is that there is a single assignment $t \in \mathcal{T}$. If the subsystems are able to exchange volume, the domain of each subsystem is extended to that of A. These changes significantly simplify the calculations for subsystems that are isolated or closed.

6.2.4. Volume. If the system is not specified by a region of space but rather by a set of atoms, then the volume likely should be calculated as the expectation value of the sum of the atomic volumes. Only the overall system volume is considered here, the intention being to discuss the strains along with the stresses in a separate publication.

Let $\mathcal{L}$ the set of labels of atoms in the system, $R_i(\bar{x})$ be the Voronoi cell of atom i, and $|R_i(\bar{x})|$ be the volume of this cell. The proposed definition for the volume is then

$$\langle V \rangle = \int_{\Gamma_x} \rho_x(\bar{x}) \sum_{i \in \mathcal{L}} |R_i(\bar{x})| \mathrm{d}\bar{x},$$

though this is likely sensible only when the distribution of the atoms is relatively homogeneous. It is straightforward to show that this definition is strongly extensive.

6.2.5. Particle Number. The final extensive variables in Eq. 6.2 are the particle numbers N_i. As discussed in Sec. 6.2.1, the system should be open and specified by a region of space for the calculation of this quantity to be necessary. Let $N_i(\bar{x})$ be the number of atoms of species i within the specified region for a given microstate. Then

$$\langle N_i \rangle = \int_{\Gamma_x} \rho_x(\bar{x}) N_i(\bar{x}) \mathrm{d}\bar{x} \tag{6.12}$$

is the proposed definition for the number of atoms of species i in the system, and significantly is a continuous function of the probability distribution on the configuration space. As with the other definitions that can be written as sums of per-atom quantities, this definition is strongly extensive.

6.3. Application to an Ideal Gas

This section applies the definitions developed in Sec. 6.2 to the expansion of an ideal gas in an isolated box. This system that has been studied in the context of non-equilibrium statistical mechanics before [**49**, **50**, **67**, **77**], though the present approach differs in several respects. For example, Hobson and Loomis [**67**] considered a system of N ideal gas particles in an isolated box, solved Liouville's equation for the evolution of the probability distribution, and used this to calculate various macroscopic observables like the center of mass, total energy, and total momentum. Then they discarded the previously-derived probability distribution, constructed a different one using the macroscopic observables and Jaynes' principle of maximum uncertainty [**71**], and used that to calculate various thermodynamic quantities including the entropy. While they observed that the entropy increased to the expected equilibrium value, this should not be confused for a signature of irreversibility. Instead, this was a result of willfully ignoring previously-known information about the probability distribution. This approach, vocally advocated by Jaynes [**72**], has been widely criticized in the literature [**37**, **86**, **125**].

That said, their solution of Liouville's equation for a system of N ideal gas particles initially confined by a removable barrier to one side of a d-dimensional isolated box is used here. The system is initially assumed to be in thermodynamic equilibrium with the joint probability distribution

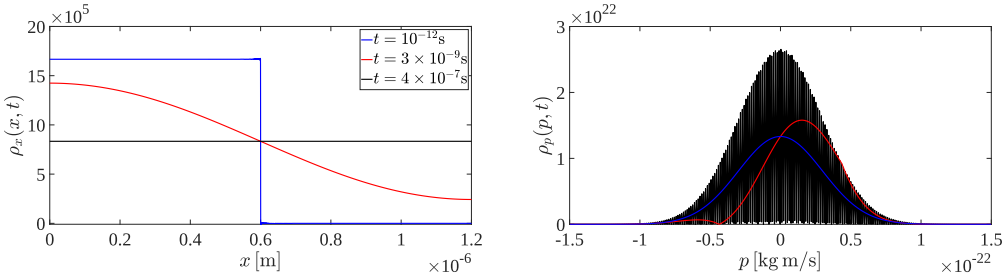

FIGURE 6.4. The evolution of the marginalized probability distributions $\rho_x(x,t)$ (left) and $\rho_p(p,t)$ (right) for a single xenon atom that is initially confined to the left half of a one-dimensional box of length $l = 1.2 \times 10^{-6}\,\mathrm{m}$ at $T = 298\,\mathrm{K}$.

$\rho(\bar{x},\bar{p},0) = \rho_x(\bar{x})\rho_p(\bar{p})$ where $\rho_x(\bar{x})$ is a uniform distribution and $\rho_p(\bar{p})$ is a Maxwell–Boltzmann distribution. Removing the barrier at time $t = 0$ allows the particles to diffuse into the rest of the box, with the evolution of the time-dependent probability distribution $\rho(\bar{x},\bar{p},t)$ found by the procedure in Sec. 6.6. Since the nature of an ideal gas allows the solution to be written as a product of the solutions for a single particle in a one-dimensional box, it is the single-particle solution that is analyzed in detail below.

Figure 6.4 shows the evolution of the marginalized probability distributions given in Eq. 6.16 for a single xenon atom that is initially confined to the left half of a one-dimensional box of length $l = 1.2 \times 10^{-6}\,\mathrm{m}$ at $T = 298\,\mathrm{K}$. $\rho_x(x,t)$ is initially a uniform probability distribution $2/l$ on the left half of the box and gradually evolves to a uniform distribution $1/l$ on the entire box. $\rho_p(p,t)$ starts as a Maxwell–Boltzmann distribution, but skews to the right shortly after the barrier is removed due to the reflection of particles at $x = 0$. With the further passage of time $\rho_p(p,t)$ actually fails to converge as a function at all, but does weakly converge to a Maxwell–Boltzmann distribution as a measure. This is particularly significant since the weak convergence to a Maxwell–Boltzmann distribution occurs without any recourse to the assumption of molecular chaos (the *Stozahlansatz*) that Boltzmann believed to be essential in his derivation of the H-theorem [**17**, **93**].

One of the contributions of Sec. 6.2 are equations to distribute the extensive thermodynamic quantities in Eq. 6.2 over arbitrary spatial partitions. This allows, e.g., densities of extrinsic quantities to be defined without resorting to the standard practice of spatial coarse-graining [**5**]. Densities are found here by partitioning the one-particle system into 1000 equal-sized subsystems and plotting the ratio of an extensive quantity and the subsystem width Δl. For example, the

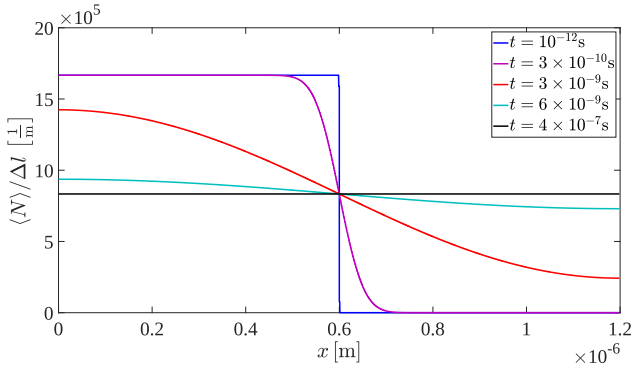

FIGURE 6.5. The evolution of the particle density of the single-particle solution as a function of position. The total number of particles is conserved at all times with a relative error on the order of 10^{-15}.

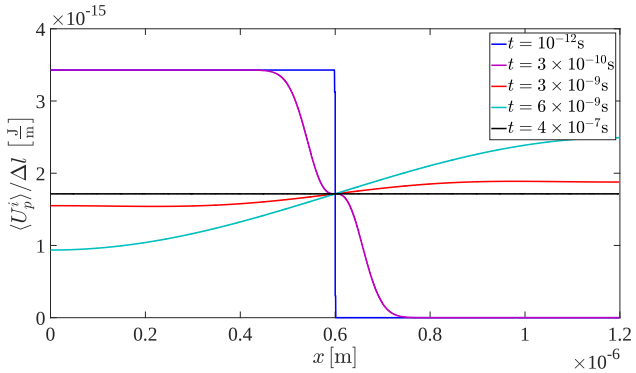

FIGURE 6.6. The evolution of the kinetic energy density of the single-particle solution as a function of position. The total kinetic energy is conserved at all times with a relative error on the order of 10^{-15}.

particle density is calculated using Eq. 6.12 and plotted as a function of position in Fig. 6.5 at various times. The evolution of the particle density closely follows that of $\rho_x(x,t)$ in Fig. 6.4, initially being uniformly distributed over the left half of the box and converging to a uniform distribution over the entire box.

Since ideal gas particles do not interact with each other and have zero potential energy by definition, only the kinetic energy contribution to the internal energy of a single particle is considered. The kinetic energy density is calculated using Eq. 6.3 and plotted as a function of position in Fig. 6.6 at various times. Initially the kinetic energy is uniformly distributed over the left half of the box, though as the expected position of the particle moves to the right the expected kinetic energy of the right half increases. For a brief period the expected kinetic energy on the right actually

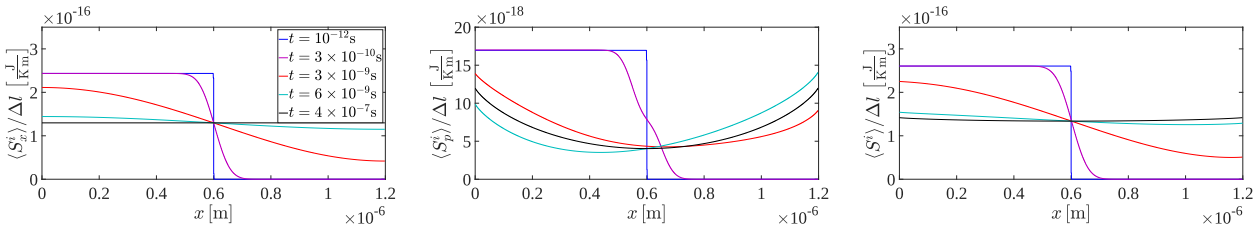

FIGURE 6.7. The evolution of the configurational (left), vibrational (middle), and total (right) entropy densities of the single-particle solution as functions of position.

exceeds that on the left (light blue curve) since a faster (slower) particle has a higher probability of being on the right (left). Predictably, the kinetic energy density eventually converges to a uniform distribution over the entire box.

The densities of the configurational, vibrational, and total entropies are calculated using Eqs. 6.10 and 6.11 and are plotted as functions of position in Fig. 6.7 at various times. While the evolution of the configurational entropy density closely follows that of the particle density in Fig. 6.5, the behavior of the vibrational entropy is quite different. It is initially distributed uniformly over the left half of the box, but converges to a distribution with higher vibrational entropy density at the ends of the box than in the middle. The precise reason for this is not known, but is apparently related to $\rho(x, p, t)$ not being a constant function of x at any time. For example, $\rho(0, p, t)$ and $\rho(l, p, t)$ are seen in Fig. 6.8 to be even functions of p with complementary supports (the sets of possible momenta of a particle at $x = 0$ and $x = l$ are disjoint at all times), whereas the support of $\rho(l/2, p, t)$ is antisymmetric about the origin. Moreover, while all three of the distributions in Fig. 6.8 seem to weakly converge to a Maxwell–Boltzmann distribution, the vibrational entropy density for a Maxwell–Boltzmann distribution would be $8.5 \times 10^{-18}\,\mathrm{J/(K\,m)}$ whereas the observed values are $1.2 \times 10^{-17}\,\mathrm{J/(K\,m)}$ at the ends of the box and $4 \times 10^{-18}\,\mathrm{J/(K\,m)}$ in the middle. This results in the nonuniformity of the converged total entropy density on the right of Fig. 6.7.

The integrated configurational, vibrational, and total entropies are shown in Fig. 6.9 as functions of time. The Sackur–Tetrode equation predicts that the total entropy should change by $9.5 \times 10^{-24}\,\mathrm{J/K}$ during the gas expansion, precisely equal to the observed increase in the configurational entropy. The Sackur–Tetrode equation does not predict a decrease in the vibrational entropy though, the effect of which is to reduce the total entropy change to $6.9 \times 10^{-24}\,\mathrm{J/K}$; this is a consequence of the Sackur–Tetrode equation assuming that the momentum distribution is unaffected

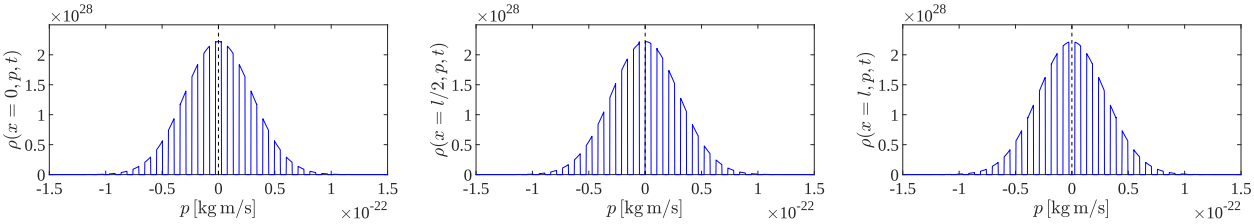

FIGURE 6.8. The probability distribution $\rho(x, p, t)$ of the single-particle solution at $x = 0$ (left), $x = l/2$ (middle) and $x = l$ (right) at $t = 5 \times 10^{-8}$ s.

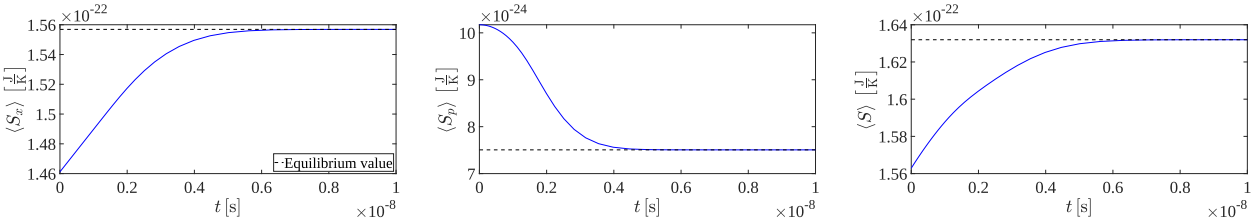

FIGURE 6.9. The configurational (left), vibrational (middle) and the total entropy (right) of the single-particle solution as a function of time.

by the expansion of the gas. While the decrease of vibrational entropy is not inconsistent with the second law of thermodynamics (which is conjectured in Sec. 6.2.3 to apply only to the total entropy), there are several conclusions that could be drawn from this phenomenon.

A first possibility is that the Sackur–Tetrode equation should hold, and that the inconsistency is in the use of Eq. 6.7 for the entropy. Indeed, the weak convergence of the probability distribution of momentum to a Maxwell–Boltzmann distribution seems to suggest that the equation for the entropy could be modified to make the vibrational entropy of the two measures the same. Considering Fig. 6.8, this would likely involve a variation on the coarse-graining procedure discussed in Sec. 6.1.3. Since the concerns raised there have not yet been resolved, this line of thought is not considered further.

A second possibility is that the Sackur–Tetrode equation should hold, and the inconsistency be resolved by a variation on Jaynes' forgetting of information. Observe that a practitioner who makes repeated and independent measurements of the position and momentum of the particle would likely conclude that $\rho(x, p, t)$ is a product of a uniform distribution $\rho_x(x, t)$ and a Maxwell–Boltzmann distribution $\rho_p(p, t)$, and hence that the Sackur–Tetrode equation holds. The argument is effectively that some information about the probability distribution of microstates is lost as

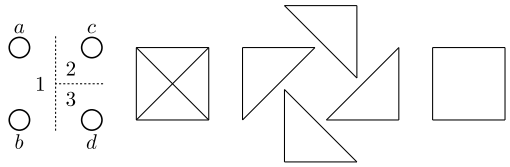

FIGURE 6.10. A partition of a four-atom system into three subsystems with the subsystem and atom labels on the left. The potential energy includes contributions from four 1-body, six 2-body, four 3-body, and one 4-body interactions, represented graphically on the right.

a consequence of experimental measurements not being able to resolve the fine structure of the probability distribution of momentum. This seems more reasonable, but raises additional questions about precisely what information should be forgotten and how to do so.

A third possibility (and the one advocated in this chapter) is that the Sackur–Tetrode equation does not hold unconditionally, but only when manipulating systems in a classical context. Knowledge of the fine structure of the probability distribution of momentum could, in principle, be used by a microscopic entity resembling Maxwell's demon to extract more work from the system than would be available classically. This would presumably be reflected by a lower entropy than that predicted by the Sackur–Tetrode equation, and could frequently occur when finite systems are concerned. If the practitioner is only able to view the system in a classical context though, then the probability distribution of momentum would likely be mistaken for a Maxwell–Boltzmann distribution as in the previous paragraph and the classical result recovered. Thus the discrepancy with the Sackur–Tetrode equation is not viewed as an inconsistency, but as an indication of the additional opportunities available with a finite non-equilibrium statistical thermodynamics.

6.4. Partitioning Potential Energy

Consider the example system in Fig. 6.10 where there are four atoms and three partitions. Table 6.1 shows all of the multiplicative factors $|\mathcal{L}^i \cap s|/|s|$ that appear in Eq. 6.4. Similarly, Table 6.2 shows the coefficients a_{rs} for events that remove atoms with labels in r. The coefficients form a 15×15 invertible matrix for four atoms, allowing potential energies to be assigned to the three partitions by means of Eq. 6.5.

TABLE 6.1. The weights $|\mathcal{L}^i \cap s|/|s|$ associated with each ϕ_s for the example system in Fig. 6.10.

i	ϕ_a	ϕ_b	ϕ_c	ϕ_d	ϕ_{ab}	ϕ_{ac}	ϕ_{ad}	ϕ_{bc}	ϕ_{bd}	ϕ_{cd}	ϕ_{abc}	ϕ_{abd}	ϕ_{acd}	ϕ_{bcd}	ϕ_{abcd}
1	1	1	0	0	1	1/2	1/2	1/2	1/2	0	2/3	2/3	1/3	1/3	1/2
2	0	0	1	0	0	1/2	0	1/2	0	1/2	1/3	0	1/3	1/3	1/4
3	0	0	0	1	0	0	1/2	0	1/2	1/2	0	1/3	1/3	1/3	1/4

TABLE 6.2. The coefficients a_{rs} for the example system in Fig. 6.10.

	ϕ_a	ϕ_b	ϕ_c	ϕ_d	ϕ_{ab}	ϕ_{ac}	ϕ_{ad}	ϕ_{bc}	ϕ_{bd}	ϕ_{cd}	ϕ_{abc}	ϕ_{abd}	ϕ_{acd}	ϕ_{bcd}	ϕ_{abcd}
χ_a	−1	0	0	0	−1	−1	−1	0	0	0	−1	−1	−1	0	−1
χ_b	0	−1	0	0	−1	0	0	−1	−1	0	−1	−1	0	−1	−1
χ_c	0	0	−1	0	0	−1	0	−1	0	−1	−1	0	−1	−1	−1
χ_d	0	0	0	−1	0	0	−1	0	−1	−1	0	−1	−1	−1	−1
χ_{ab}	−1	−1	0	0	−1	−1	−1	−1	−1	0	−1	−1	−1	−1	−1
χ_{ac}	−1	0	−1	0	−1	−1	−1	−1	0	−1	−1	−1	−1	−1	−1
χ_{ad}	−1	0	0	−1	−1	−1	−1	0	−1	−1	−1	−1	−1	−1	−1
χ_{bc}	0	−1	−1	0	−1	−1	0	−1	−1	−1	−1	−1	−1	−1	−1
χ_{bd}	0	−1	0	−1	−1	0	−1	−1	−1	−1	−1	−1	−1	−1	−1
χ_{cd}	0	0	−1	−1	0	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1
χ_{abc}	−1	−1	−1	0	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1
χ_{abd}	−1	−1	0	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1
χ_{acd}	−1	0	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1
χ_{bcd}	0	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1
χ_{abcd}	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1

6.5. Formula for Weakly Extrinsic Potential Energy

The proof that Eq. 6.6 distributes the potential energy of a system A in the desired way over the n subsystems A^i begins with the assumption that the potential energy can be written as a sum of many-body interactions. Consider a single m-body interaction with atoms contained in p subsystems. The hypothesis is that the number of terms on the right hand side of Eq. 6.6 containing this interaction is precisely $n!/p$, or that the first time an atom participating in the interaction is removed coincides with removing all the atoms of A^i for $n!/p$ of the $n!$ possible permutations. This would result in the overall contribution of the interaction to $U^i_x(\bar{x})$ being $1/p$ times the energy of the interaction, as desired.

Consider the number of permutations such that the q subsystems selected before A^i do not participate in the interaction. There are $(n-p)!/(n-p-q)!$ ways to select such initial subsystems, and $(n-1-q)!$ ways to select the subsystems afterwards. This implies that the hypothesis is true

if the equation

$$\frac{n!}{p} = \sum_{q=0}^{n-p} \frac{(n-p)!}{(n-p-q)!}(n-1-q)!$$

holds. This can be shown to be so by making the substitution $r = n - p - q$ and replacing the summand with a binomial coefficient to find

$$\frac{n!}{p} = (n-p)!(p-1)! \sum_{r=0}^{n-p} \binom{p-1+r}{r}.$$

The summation can be replaced by a single binomial coefficient using a variant of the Chu–Vandermonde identity with the result

$$\frac{n!}{p} = (n-p)!(p-1)! \binom{n}{n-p}.$$

Expanding the binomial coefficient and canceling the resulting terms reduces this to an identity.

6.6. Evolution of an Ideal Gas in a Box

Liouville's equation in principle describes the evolution of the probability distribution $\rho(\bar{x}, \bar{p}, t)$ for any isolated system, but is difficult to solve in practice. An ideal gas in an isolated box is one of the few systems for which a general solution is known [**67**], and relies on the observation that the problem can be reduced to a single particle moving in a one-dimensional box. The general solution can be written as a product of the one-dimensional solutions, or

$$\rho(\bar{x}, \bar{p}, t) = \prod_{j=1}^{N} \prod_{\mu=1}^{d} \rho(x_{\mu j}, p_{\mu j}, t)$$

$$\rho(x_{\mu j}, p_{\mu j}, t) = \mathrm{e}^{-i\hat{L}_{\mu j} t} \rho(x_{\mu j}, p_{\mu j}, 0) \tag{6.13}$$

where j and μ index the particles and spatial dimensions and $\hat{L}_{\mu j}$ is the one-dimensional Liouville operator. A procedure to obtain the one-dimensional solution was developed by Born [**20**] using the method of images. This replaces the actual Hamiltonian H by the free-particle Hamiltonian $\bar{H} = p^2/2m$ (the subscripts μ and j are dropped for simplicity) and handles the boundary conditions by modifying the probability distribution. The initial probability distribution $\rho(x, p, 0)$ is replaced by an extended probability distribution $\bar{\rho}(x, p, 0)$ such that $\bar{\rho}(-x, -p, 0) = \bar{\rho}(x, p, 0)$ and $\bar{\rho}(x +$

$2l, p, 0) = \bar{\rho}(x, p, 0)$ where l is the length of the box. The extended problem is easier to solve, and the solution of the actual problem is $\rho(x, p, t) = \bar{\rho}(x, p, t)$ for $0 < x < l$. Details of this approach can be found in Refs. [**20**, **67**].

Consider a system of N ideal gas particles initially confined to the left-hand side of a one-dimensional isolated box. The gas is in thermodynamic equilibrium with the initial condition

$$\rho(x, p, 0) = \begin{cases} \rho_x(x, 0)\rho_p(p, 0) & \text{for } 0 < x < l', \\ 0 & \text{for } l' < x < l. \end{cases}$$

where $\rho_x(x, 0) = 1/l'$ is a uniform probability distribution and $\rho_p(p, 0)$ is the Maxwell–Boltzmann distribution. The extended probability distribution $\bar{\rho}(x, p, 0)$ is constructed as above, and along with the free-particle Hamiltonian $\bar{H}$ gives [**67**]

$$\bar{\rho}(x, p, t) = \bar{\rho}(x - pt, p, 0) \tag{6.14}$$

for the time evolution. Since $\bar{\rho}(x, p, 0)$ is an even function of x it can be written as the Fourier cosine series

$$\begin{aligned} \bar{\rho}(x, p, 0) &= a_0(p) + \sum_{n=1}^{\infty} a_n(p) \cos\Big(\frac{n\pi x}{l}\Big) \\ a_0(p) &= \frac{1}{l}\rho_p(p, 0) \\ a_n(p) &= \frac{2}{n\pi l'} \sin\Big(\frac{n\pi l'}{l}\Big)\rho_p(p, 0). \end{aligned}$$

Using Eq. 6.14 then yields the solution of a single particle with the free-particle Hamiltonian as

$$\bar{\rho}(x, p, t) = \left\{ \frac{1}{l} + \frac{2}{\pi l'} \sum_{n=1}^{\infty} \frac{1}{n} \sin\Big(\frac{n\pi l'}{l}\Big) \cos\Big[\frac{n\pi}{l}\Big(x - \frac{pt}{m}\Big)\Big] \right\} \rho_p(p, 0).$$

Recall that the solution of the actual problem is

$$\rho(x, p, t) = \bar{\rho}(x, p, t) \quad \text{for} \quad 0 < x < l. \tag{6.15}$$

Along with Eq. 6.13, this can be used to find, e.g., the solution for N particles initially confined to the left-hand side of a three-dimensional box. The single particle marginal probability distributions

$\rho_x(x,t)$ and $\rho_p(p,t)$ are

$$\rho_x(x,t) = \left[\frac{1}{l} + \frac{2}{\pi l'}\sum_{n=1}^{\infty}\frac{1}{n}\sin\Big(\frac{n\pi l'}{l}\Big)\exp\Big(-\frac{n^2t^2}{2\sigma^2}\Big)\cos\Big(\frac{n\pi x}{l}\Big)\right],$$

$$\rho_p(p,t) = \left[1 + \frac{4l}{\pi^2 l'}\sum_{n=1}^{\infty}{}'\frac{1}{n^2}\sin\Big(\frac{n\pi l'}{l}\Big)\sin\Big(\frac{n\pi pt}{lm}\Big)\right]\rho_p(p,0) \tag{6.16}$$

where $\sigma = l/\pi\sqrt{m/(k_BT)}$ and the primed summation is only over odd n. With increasing time the exponential term in $\rho_x(x,t)$ goes to zero and $\rho_x(x,t)$ converges to $1/l$ as expected. The dependence of $\rho_p(p,t)$ on time suggests that the limit is not well-defined as a function, but $\rho_p(p,t)$ does appear to weakly converge to a Maxwell–Boltzmann distribution as a measure [**67**].

CHAPTER 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. 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 [**23**, **52**] suggests that topological changes to the accessible region of the configuration space is a necessary condition for a phase transition. This study instead suggest that a substantial change to the geometry of the accessible region is a necessary condition for a phase transition, and that such a change often (but not always) associated with a topological change. More specifically, Conj. 1.0.1 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.

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, 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 or graph networks, and thereby offer new approaches to studying the hard disk and sphere systems. Specifically, the topological and geometric properties of the spaces can be directly evaluated as functions of disk

*The content of this chapter has previously been published in Refs. [**45**, **47**, **48**].

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.

First, the critical points of the configuration spaces for $n = 1 \dots 12$ hard disks and spheres were sampled, tabulated in an online database, and made freely available as a Colab Notebook. The second finding is that the use of a square or hexagonal torus in the case of hard disks 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 third 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 fourth 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 (index-3) critical point that closes the space for the configuration spaces of hard disks (spheres). The minimum-energy paths between the critical configurations were found using the zero-temperature string method, and a new topological invariant called the minimum-energy graph was proposed.

Considering the geometry of the configuration spaces, it is found that 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. 1.0.1, 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 and spheres 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 [**3**], and further development of a suitable min-type Morse theory would also be helpful [**6**, **99**].

Underlying all of this discussion is the assumption that the probability distribution of microstates, an essential concept in statistical thermodynamics and the definition of a phase, is a sensible quantity. Since precisely defining thermodynamic ensembles for non-equilibrium states seems to be fraught with difficulties, this study instead interprets the probability distribution of microstates without using ensembles. This quickly leads to the position advocated by Jaynes [**74**], that the probability distribution describes the subjective uncertainty of the practitioner about the microstate. The implication is that thermodynamics is not really concerned with the objective state of a thermodynamic system, but rather with the ability of the practitioner to manipulate that state. The entropy is interpreted in the context of the fundamental thermodynamic relation as indicating the amount of internal energy that cannot be extracted as work as due to the practitioner's imperfect knowledge of the microstate. Consistent resolutions to the mixing paradox, Gibbs' paradox, Loschmidt's paradox, and Maxwell's demon thought experiment are further consequences of the subjectivity of the probability distribution.

An equation for the entropy is proposed in Sec. 6.2.3 that coincides with Gibbs' entropy for the canonical ensemble, and by the equivalence of ensembles, is consistent with effectively all prior thermodynamic results in the equilibrium thermodynamic limit. The essential difference with Gibbs' entropy is that the proposed definition increases with the mutual information of the position and momentum variables, allowing the dynamics imposed by Liouville's equation to cause the spontaneous increase of entropy in non-equilibrium systems. That is, an answer is offered to the question of irreversibility. Definitions are proposed that allow all of the extensive quantities in the fundamental thermodynamic relation to be distributed over arbitrary thermodynamic subsystems. These definition are used to analyze the expansion of an ideal gas into an isolated box. The

results are found to coincide with those expected in a classical context, and suggest that there are additional opportunities offered by a finite non-equilibrium statistical thermodynamics.

The finite non-equilibrium statistical thermodynamics begun here is not complete. Notably, the intrinsic quantities appearing in the fundamental thermodynamic relation are not addressed, and there is reason to believe that their definition will follow along different lines than those explored here. Their development is intended for a future study.

Bibliography

[1] R. Abraham and J. E. Marsden, *Foundations of mechanics*, no. 364, American Mathematical Soc., 2008.

[2] B. Alder and T. Wainwright, *Phase transition in elastic disks*, Physical Review, 127 (1962), p. 359.

[3] B. Aronov, M. Dulieu, and F. Hurtado, *Witness gabriel graphs*, Computational Geometry, 46 (2013), pp. 894–908. EuroCG 2009.

[4] R. Bader, *Atoms in molecules: A quantum theory*, Clarendon Press, 1994.

[5] R. W. Balluffi, S. M. Allen, and W. C. Carter, *Kinetics of materials*, John Wiley & Sons, 2005.

[6] Y. Baryshnikov, P. Bubenik, and M. Kahle, *Min-type morse theory for configuration spaces of hard spheres*, International Mathematics Research Notices, 2014 (2014), p. 25772592.

[7] O. M. Becker and M. Karplus, *The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics*, The Journal of chemical physics, 106 (1997), pp. 1495–1517.

[8] W. Beckner, *Inequalities in fourier analysis*, Annals of Mathematics, 102 (1975), pp. 159–182.

[9] G. Bel-Hadj-Aissa, M. Gori, R. Franzosi, and M. Pettini, *Geometrical and topological study of the kosterlitz–thouless phase transition in the xy model in two dimensions*, Journal of Statistical Mechanics: Theory and Experiment, 2021 (2021), p. 023206.

[10] E. P. Bernard and W. Krauth, *Two-step melting in two dimensions: First-order liquid-hexatic transition*, Physical review letters, 107 (2011), p. 155704.

[11] S. Bhattacharya, *Towards optimal path computation in a simplicial complex*, The International Journal of Robotics Research, 38 (2019), pp. 981–1009.

[12] I. Białynicki-Birula and J. Mycielski, *Uncertainty relations for information entropy in wave mechanics*, Communications in Mathematical Physics, 44 (1975), pp. 129–132.

[13] G. D. Birkhoff, *Proof of the ergodic theorem*, Proceedings of the National Academy of Sciences, 17 (1931), pp. 656–660.

[14] J. Blatt, *An alternative approach to the ergodic problem*, Progress of Theoretical Physics, 22 (1959), pp. 745–756.

[15] J.-D. Boissonnat, L. J. Guibas, and S. Y. Oudot, *Manifold reconstruction in arbitrary dimensions using witness complexes*, Discrete & Computational Geometry, 42 (2009), pp. 37–70.

[16] L. Boltzmann, *Einige allgemeine sätze über wärmegleichgewicht*, Wiener Berichte, 63 (1871), pp. 679–711.

[17] ———, *Weitere studien über das wärmegleichgewicht unter gasmolekülen*, Wiener Berichte, 66 (1872), pp. 275–370.

[18] ———, *Bemerkungen über einige probleme der mechanischen wärmetheorie*, Wiener Berichte, 75 (1877), pp. 62–100.

[19] ———, *Vorlesungen über gastheorie*, vol. 2, Barth, 1912.

[20] M. Born, *Continuity, determinism, and reality*, Munksgaard, 1955.

[21] M. Buchet, F. Chazal, S. Y. Oudot, and D. R. Sheehy, *Efficient and robust persistent homology for measures*, Computational Geometry, 58 (2016), pp. 70–96.

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

[23] L. Caiani, L. Casetti, C. Clementi, and M. Pettini, *Geometry of dynamics, lyapunov exponents, and phase transitions*, Physical review letters, 79 (1997), p. 4361.

[24] G. Carlsson, *Topology and data*, Bulletin of the American Mathematical Society, 46 (2009), pp. 255–308.

[25] G. Carlsson, J. Gorham, M. Kahle, and J. Mason, *Computational topology for configuration spaces of hard disks*, Phys. Rev. E, 85 (2012), p. 011303.

[26] F. Chazal and S. Y. Oudot, *Towards persistence-based reconstruction in euclidean spaces*, in Proceedings of the twenty-fourth annual symposium on Computational geometry, 2008, pp. 232–241.

[27] R. Chelouah and P. Siarry, *Tabu search applied to global optimization*, European journal of operational research, 123 (2000), p. 256270.

[28] F. R. Chung and F. C. Graham, *Spectral graph theory*, no. 92, American Mathematical Soc., 1997.

[29] R. R. Coifman and S. Lafon, *Diffusion maps*, Applied and computational harmonic analysis, 21 (2006), p. 530.

[30] I. P. Cornfeld, S. V. Fomin, and Y. G. Sinai, *Ergodic theory*, vol. 245, Springer Science & Business Media, 2012.

[31] T. Cover and J. Thomas, *Elements of information theory*, Wiley, 2012.

[32] H. H. Denman, *Invariance and conservation laws in classical mechanics*, Journal of Mathematical Physics, 6 (1965), p. 16111616.

[33] C. P. Dettmann and G. Morriss, *Proof of lyapunov exponent pairing for systems at constant kinetic energy*, Physical Review E, 53 (1996), p. R5545.

[34] E. W. Dijkstra et al., *A note on two problems in connexion with graphs*, Numerische mathematik, 1 (1959), p. 269271.

[35] H. S. Dumas, *The KAM story: A friendly introduction to the content, history, and significance of classical Kolmogorov-Arnold-Moser theory*, World Scientific Publishing Company, 2014.

[36] J. C. Dyre, *Simple liquids quasiuniversality and the hard-sphere paradigm*, Journal of Physics: Condensed Matter, 28 (2016), p. 323001.

[37] J. Earman, *The problem of irreversibility*, in PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, vol. 2, University of Chicago Press, Springer, Philosophy of Science Association, 1986, pp. 226–233.

[38] J.-P. Eckmann and D. Ruelle, *Ergodic theory of chaos and strange attractors*, The theory of chaotic attractors, (1985), pp. 273–312.

[39] H. Edelsbrunner and J. Harer, *Persistent homology-a survey*, Contemporary mathematics, 453 (2008), pp. 257–282.

[40] H. Edelsbrunner, D. Kirkpatrick, and R. Seidel, *On the shape of a set of points in the plane*, IEEE Transactions on information theory, 29 (1983), p. 551559.

[41] P. Ehrenfest and T. Ehrenfest, *The conceptual foundations of the statistical approach in mechanics*, Courier Corporation, 1990.

[42] A. Einstein, *Die plancksche theorie der strahlung und die theorie der spezifischen wärme*, Annalen der Physik, 327 (1907), pp. 180–190.

[43] R. S. Ellis, K. Haven, and B. Turkington, *Large deviation principles and complete equivalence and nonequivalence results for pure and mixed ensembles*, Journal of Statistical Physics, 101 (2000), pp. 999–1064.

[44] M. Engel, J. A. Anderson, S. C. Glotzer, M. Isobe, E. P. Bernard, and W. Krauth, *Hard-disk equation of state: First-order liquid-hexatic transition in two dimensions with three simulation methods*, Physical Review E, 87 (2013), p. 042134.

[45] O. Ericok and J. Mason, *Quotient maps and configuration spaces of hard disks*, arXiv preprint arXiv:2101.00780, (2021).

[46] ———, *A geometric conjecture about phase transitions*, arXiv preprint arXiv:2203.03154, (2022).

[47] ———, *Foundations of a finite non-equilibrium statistical thermodynamics: Extrinsic quantities*, arXiv preprint arXiv:2201.02777, (2022).

[48] O. B. Eriçok, K. Ganesan, and J. K. Mason, *Configuration spaces of hard spheres*, Phys. Rev. E, 104 (2021), p. 055304.

[49] J. Español, F. J. De La Rubia, and J. M. Rubí, *Initial non-equilibrium ensembles: Application to the ideal gas*, Physica A: Statistical Mechanics and its Applications, 171 (1991), pp. 120–138.

[50] A. K. Evans, *Long-time behavior of correlation functions in the finite ideal gas*, Physical Review E, 51 (1995), p. 5462.

[51] F. Fouss, A. Pirotte, J.-M. Renders, and M. Saerens, *A novel way of computing dissimilarities between nodes of a graph, with application to collaborative filtering and subspace projection of the graph nodes*, Technical Report, Tech. Rep., (2006).

[52] R. Franzosi and M. Pettini, *Theorem on the origin of phase transitions*, Physical review letters, 92 (2004), p. 060601.

[53] ——, *Topology and phase transitions ii. theorem on a necessary relation*, Nuclear Physics B, 782 (2007), p. 219240.

[54] R. Franzosi, M. Pettini, and L. Spinelli, *Topology and phase transitions i. preliminary results*, Nuclear Physics B, 782 (2007), p. 189218.

[55] G. Gallavotti, *The Fermi-Pasta-Ulam problem: A status report*, vol. 728, Springer, 2007.

[56] J. H. Gibbs and E. A. DiMarzio, *Nature of the glass transition and the glassy state*, The Journal of Chemical Physics, 28 (1958), pp. 373–383.

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

[58] F. Glover, *Tabu searchpart i*, ORSA Journal on computing, 1 (1989), p. 190206.

[59] H. Goldstein and C. Poole, *Classical mechanics*, Addison-Wesley Publishing Company, 1980.

[60] M. Gori, R. Franzosi, and M. Pettini, *Topological origin of phase transitions in the absence of critical points of the energy landscape*, Journal of Statistical Mechanics: Theory and Experiment, 2018 (2018), p. 093204.

[61] H. Gould and J. Tobochnik, *Statistical and Thermal Physics*, Princeton University Press, 2010.

[62] E. L. Hahn, *Spin echoes*, Physical Review, 80 (1950), p. 580.

[63] B. Halperin and D. R. Nelson, *Theory of two-dimensional melting*, Physical Review Letters, 41 (1978), p. 121.

[64] D. Hankins, J. Moskowitz, and F. Stillinger, *Water molecule interactions*, The Journal of Chemical Physics, 53 (1970), pp. 4544–4554.

[65] A. Hatcher, *Algebraic topology*, Cambridge University Press, 2002.

[66] G. Henkelman, B. P. Uberuaga, and H. Jónsson, *A climbing image nudged elastic band method for finding saddle points and minimum energy paths*, The Journal of chemical physics, 113 (2000), pp. 9901–9904.

[67] A. Hobson and D. N. Loomis, *Exact classical nonequilibrium statistical-mechanical analysis of the finite ideal gas*, Physical Review, 173 (1968), p. 285.

[68] K. Huang, *Introduction to statistical physics*, Chapman and Hall/CRC, 2009.

[69] M. Isobe and W. Krauth, *Hard-sphere melting and crystallization with event-chain monte carlo*, The Journal of chemical physics, 143 (2015), p. 084509.

[70] G. Jaeger, *The ehrenfest classification of phase transitions: introduction and evolution*, Archive for history of exact sciences, 53 (1998), pp. 51–81.

[71] E. T. Jaynes, *Information theory and statistical mechanics*, Physical Review, 106 (1957), p. 620.

[72] ——, *Gibbs vs boltzmann entropies*, American Journal of Physics, 33 (1965), pp. 391–398.

[73] ——, *Prior probabilities*, IEEE Transactions on Systems Science and Cybernetics, 4 (1968), pp. 227–241.

[74] ———, *The gibbs paradox*, in Maximum entropy and Bayesian methods, Springer, 1992, pp. 1–21.

[75] H. Jónsson, G. Mills, and K. W. Jacobsen, *Nudged elastic band method for finding minimum energy paths of transitions*, (1998).

[76] H. Kahn and T. E. Harris, *Estimation of particle transmission by random sampling*, National Bureau of Standards applied mathematics series, 12 (1951), pp. 27–30.

[77] P. Kasperkovitz and C. Foidl, *Approach to equilibrium in integrable gas models*, Physical Review A, 37 (1988), p. 3172.

[78] M. Kastner and D. Mehta, *Phase transitions detached from stationary points of the energy landscape*, Physical review letters, 107 (2011), p. 160602.

[79] W. Kauzmann, *The nature of the glassy state and the behavior of liquids at low temperatures.*, Chemical Reviews, 43 (1948), pp. 219–256.

[80] A. N. Kolmogorov, *On the conservation of conditionally periodic motions for a small change in hamiltons function [en rus]*, in Dokl. Akad. Nauk SSSR, vol. 98, 1954, pp. 525–530.

[81] S. Krantz and H. Parks, *The Implicit Function Theorem: History, Theory, and Applications*, Modern Birkhuser Classics, Springer New York, 2012.

[82] S. Kullback and R. A. Leibler, *On information and sufficiency*, The annals of mathematical statistics, 22 (1951), pp. 79–86.

[83] L. Landau, E. Lifshitz, J. Sykes, and J. Bell, *Mechanics: Volume 1*, Elsevier Science, 1976.

[84] L. D. Landau and E. M. Lifshitz, *Quantum mechanics: Non-relativistic theory*, vol. 3, Elsevier, 2013.

[85] L. D. Landau and E. M. Lifshitz, *Statistical Physics: Volume 5*, vol. 5, Elsevier, 2013.

[86] D. A. Lavis and P. J. Milligan, *The work of e. t. jaynes on probability, statistics and statistical physics*, The British Journal for the Philosophy of Science, 36 (1985), pp. 193–210.

[87] J. M. Leinaas and J. Myrheim, *On the theory of identical particles*, Il Nuovo Cimento B (1971-1996), 37 (1977), pp. 1–23.

[88] C. Lindblad, *Non-equilibrium entropy and irreversibility*, vol. 5, Springer Science & Business Media, 2001.

[89] J. Liouville, *Note sur la thorie de la variation des constantes arbitraires*, Journal de Mathmatiques Pures et Appliques, 3 (1838), pp. 342–349.

[90] J. Loschmidt, *ber die zustand des wrmegleichgewichtes eines systems von krpern mit rcksicht auf die schwerkraft*, Wiener Berichte, 73 (1876), pp. 128–142.

[91] F. Mandl, *Statistical physics*, Wiley, 2013.

[92] S. H. Mannaerts, *Extensive quantities in thermodynamics*, European Journal of Physics, 35 (2014), p. 035017.

[93] J. C. Maxwell, *Iv. on the dynamical theory of gases*, Philosophical Transactions of the Royal Society of London, (1867), pp. 49–88.

[94] J. C. Maxwell, *On Boltzmann's theorem on the average distribution of energy in a system of material points*, Transactions of the Cambridge Philosophical Society, 12 (1879), pp. 547–70.

[95] B. D. McKay and A. Piperno, *Practical graph isomorphism, ii*, Journal of Symbolic Computation, 60 (2014), pp. 94–112.

[96] D. McQuarrie, *Statistical mechanics*, New York: Harper & Row, 1981.

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

[98] A. Mitus, H. Weber, and D. Marx, *Local structure analysis of the hard-disk fluid near melting*, Physical Review E, 55 (1997), p. 6855.

[99] F. Mori and M. Salvetti, *(discrete) morse theory on configuration spaces*, Mathematical Research Letters, 18 (2011), pp. 39–57.

[100] D. Morozov and G. Weber, *Distributed merge trees*, in Proceedings of the 18th ACM SIGPLAN symposium on Principles and practice of parallel programming, 2013, pp. 93–102.

[101] R. J. Morris, *Lavoisier and the caloric theory*, The British Journal for the History of Science, 6 (1972), pp. 1–38.

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

[103] A. Mulero, C. Galán, M. Parra, and F. Cuadros, *Equations of state for hard spheres and hard disks*, in Theory and Simulation of Hard-Sphere Fluids and Related Systems, Springer, 2008, pp. 37–109.

[104] J. R. Munkres, *Elements of algebraic topology*, CRC press, 2018.

[105] J. v. Neumann, *Proof of the quasi-ergodic hypothesis*, Proceedings of the National Academy of Sciences, 18 (1932), pp. 70–82.

[106] L. Onsager, *Crystal statistics. i. a two-dimensional model with an order-disorder transition*, Physical Review, 65 (1944), p. 117.

[107] L. Peliti, *Statistical mechanics in a nutshell*, vol. 10, Princeton University Press, 2011.

[108] O. Penrose, *Foundations of statistical mechanics: A deductive treatment*, Courier Corporation, 2005.

[109] Y. B. Pesin, *Characteristic lyapunov exponents and smooth ergodic theory*, Dynamical Systems: Collection of Papers, 1 (1991), p. 117.

[110] S. Pieprzyk, M. N. Bannerman, A. C. Brańka, M. Chudak, and D. M. Heyes, *Thermodynamic and dynamical properties of the hard sphere system revisited by molecular dynamics simulation*, Physical Chemistry Chemical Physics, 21 (2019), pp. 6886–6899.

[111] F. P. Preparata and M. I. Shamos, *Computational geometry: an introduction*, Springer Science & Business Media, 2012.

[112] K. Ridderbos, *The coarse-graining approach to statistical mechanics: how blissful is our ignorance?*, Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, 33 (2002), pp. 65–77.

[113] K. Ritchey, *Computational Topology for Configuration Spaces of Disks in a Torus*, PhD thesis, Ohio State University, Electronic Thesis or Dissertation, 2019.

[114] O. Sackur, *Die anwendung der kinetischen theorie der gase auf chemische probleme*, Annalen der Physik, 341 (1911), pp. 958–980.

[115] R. Seidel, *The upper bound theorem for polytopes: an easy proof of its asymptotic version*, Computational Geometry, 5 (1995), p. 115116.

[116] R. Shankar, *Principles of quantum mechanics*, Springer Science & Business Media, 2012.

[117] D. R. Sheehy, *Linear-size approximations to the vietoris–rips filtration*, Discrete & Computational Geometry, 49 (2013), pp. 778–796.

[118] R. Talmon, I. Cohen, S. Gannot, and R. R. Coifman, *Diffusion maps for signal processing: A deeper look at manifold-learning techniques based on kernels and graphs*, IEEE signal processing magazine, 30 (2013), pp. 75–86.

[119] J. B. Tenenbaum, V. d. Silva, and J. C. Langford, *A global geometric framework for nonlinear dimensionality reduction*, Science, 290 (2000), p. 23192323.

[120] H. Tetrode, *Die chemische konstante der gase und das elementare wirkungsquantum*, Annalen der Physik, 343 (1912), pp. 434–442.

[121] J. Tian, H. Jiang, and A. Mulero, *Equations of state for the hard disk fluids*, Molecular Physics, 118 (2020), p. e1687948.

[122] R. C. Tolman, *The principles of statistical mechanics*, Courier Corporation, 1979.

[123] H. Touchette, *When is a quantity additive, and when is it extensive?*, Physica A: Statistical Mechanics and its Applications, 305 (2002), pp. 84–88.

[124] ———, *Equivalence and nonequivalence of ensembles: thermodynamic, macrostate, and measure levels*, Journal of Statistical Physics, 159 (2015), pp. 987–1016.

[125] J. Uffink, *Probabilities in physics*, Oxford University Press, 2011, ch. Subjective probabilit yand statistical physics.

[126] L. Vietoris, *ber den hheren zusammenhang kompakter rume und eine klasse von zusammenhangstreuen abbildungen*, Mathematische Annalen, 97 (1927), p. 454472.

[127] U. Von Luxburg, A. Radl, and M. Hein, *Hitting and commute times in large random neighborhood graphs*, The Journal of Machine Learning Research, 15 (2014), pp. 1751–1798.

[128] H. Weber, D. Marx, and K. Binder, *Melting transition in two dimensions: A finite-size scaling analysis of bond-orientational order in hard disks*, Physical Review B, 51 (1995), p. 14636.

[129] E. Weinan, W. Ren, and E. Vanden-Eijnden, *String method for the study of rare events*, Physical Review B, 66 (2002), p. 052301.

[130] ———, *Simplified and improved string method for computing the minimum energy paths in barrier-crossing events*, Journal of Chemical Physics, 126 (2007), p. 164103.

[131] M. T. Wells, G. Casella, and C. P. Robert, *Generalized accept-reject sampling schemes*, in A Festschrift for Herman Rubin, Institute of Mathematical Statistics, 2004, pp. 342–347.

[132] C. N. Yang, *The spontaneous magnetization of a two-dimensional ising model*, Physical Review, 85 (1952), p. 808.

[133] A. Young, *Melting and the vector coulomb gas in two dimensions*, Physical Review B, 19 (1979), p. 1855.

[134] M. Yu, D. R. Trinkle, and R. M. Martin, *Energy density in density functional theory: Application to crystalline defects and surfaces*, Physical Review B, 83 (2011), p. 115113.