A geometrical approach to computing free-energy landscapes from short-ranged potentials

Miranda Holmes-Cerfon¹, Steven J. Gortler, and Michael P. Brenner

School of Engineering and Applied Sciences and Kavli Institute for Bionano Science and Technology, Harvard University, Cambridge, MA 02138

Edited by José N. Onuchic, Rice University, Houston, TX, and approved November 7, 2012 (received for review July 10, 2012)

Particles interacting with short-ranged potentials have attracted increasing interest, partly for their ability to model mesoscale systems such as colloids interacting via DNA or depletion. We consider the free-energy landscape of such systems as the range of the potential goes to zero. In this limit, the landscape is entirely defined by geometrical manifolds, plus a single control parameter. These manifolds are fundamental objects that do not depend on the details of the interaction potential and provide the starting point from which any quantity characterizing the system—equilibrium or nonequilibrium—can be computed for arbitrary potentials. To consider dynamical quantities we compute the asymptotic limit of the Fokker–Planck equation and show that it becomes restricted to the low-dimensional manifolds connected by “sticky” boundary conditions. To illustrate our theory, we compute the low-dimensional manifolds for $n \leq 8$ identical particles, providing a complete description of the lowest-energy parts of the landscape including floppy modes with up to 2 internal degrees of freedom. The results can be directly tested on colloidal clusters. This limit is a unique approach for understanding energy landscapes, and our hope is that it can also provide insight into finite-range potentials.

sticky spheres | self-assembly | transition rates | sticky Brownian motion

The dynamics on free-energy landscapes are a ubiquitous paradigm for characterizing molecular and mesoscopic systems, from atomic clusters, to protein folding, to colloidal clusters (1–4). The predominant strategy for understanding the dynamics on an energy landscape has focused on the stationary points of the energy, the local minima and the transition states, and seeks the dynamical paths that connect these to each other, whereas more recent models generalize to metastable states connected by paths as a Markov state model (5). These techniques have proved to be extremely powerful, giving innumerable insights into the behavior of complex systems (6–13). On the other hand, a major issue has been the difficulty of finding the transition paths, connecting local minima or metastable states to each other, especially given a complex energy landscape in a high-dimensional space. A variety of creative methods have been developed in recent years for efficiently finding transition paths (14–24) but for a given system, there is no guarantee that all such paths have been found.

Here we present a different point of view for understanding an energy landscape that occurs when the range over which particles interact is much smaller than their size. Such is the case in certain mesoscale systems, for example, for C_{60} molecules (25, 26), or for colloids interacting via depletion (4) or coated with polymers or cDNA strands (27–30). We show that in this limit, the free-energy landscape is described entirely by geometry, plus a single control parameter κ that is a function of the temperature, depth, and curvature of the original potential. This limit is related to the sticky sphere limit of a square-well potential (31), which has been used to investigate thermodynamic properties of hard sticky spheres (32–34). The landscape can be thought of as a polygon living in a high-dimensional space, whose corners (zero-dimensional manifolds) are joined to each other by lines (one-dimensional manifolds) that in turn form the boundaries of faces (two-dimensional manifolds), and so on. These manifolds

are fixed functions of the particles’ geometries, independent of the details of the original interaction potential from which the limit was taken.

Once the geometrical manifolds composing the landscape are computed, nonequilibrium quantities characterizing the dynamics can be calculated by solving the Fokker–Planck equation or its adjoint on these manifolds. We show that in the short-ranged limit these equations acquire an effective boundary condition at the boundary of every p -dimensional manifold in the polygon. This makes the kinetics computationally tractable, because the stiff modes of a narrow potential become a set of boundary conditions.

The geometrical nature of the energy landscape does not mitigate its high dimensionality, but at low temperatures (high κ) both the free energies and the kinetics are dominated by the lowest-dimensional manifolds. This means that the description of the free-energy landscape and kinetics for short-ranged potentials reduces to a problem in discrete and computational geometry—to find all of the low-dimensional manifolds for a given set of interacting particles.

As an illustration of the framework, we characterize the geometrical landscape for $n \leq 8$ particles with identical potentials and demonstrate how these solutions lead to a complete description of the energy landscape and the kinetics of this system. This solution describes the geometrical limit of small atomic clusters as well as a direct prediction for colloidal clusters interacting with depletion forces (4, 6, 35, 36), where the predictions could be tested experimentally. The solution also provides a framework for understanding and extending simulations on clusters with finite range potentials (1). Our calculation of the energy landscape builds on the enumeration of all finite sphere packings of n particles with at least $3n - 6$ contacts (37, 38). With these as the starting point, we compute every one- and two-dimensional manifold of motions that contains a zero-dimensional manifold at its boundary, from which we can extract statistical quantities such as the relative entropies of the different types of motions. Then, we solve Fokker–Planck equations on these manifolds to obtain transition rates between the lowest energy states, the zero-dimensional manifolds.

Geometrical Landscape

We begin by showing how the geometrical free-energy landscape arises as a distinguished limit of particles interacting with arbitrary potentials of finite range. We consider a point $x \in \mathbb{R}^N$ in configuration space, with the potential energy given as a sum of potentials concentrated near different geometrical boundaries as

Author contributions: M.H.-C. and M.P.B. designed research; M.H.-C. and M.P.B. performed research; S.J.G. contributed new reagents/analytic tools; and M.H.-C. and M.P.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. E-mail: holmes@cims.nyu.edu.

See Author Summary on page 12 (volume 110, number 1).

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1211720110/-DCSupplemental.

$$U^\epsilon(x) = \sum_k C_\epsilon U(y_k(x)/\epsilon). \quad [1]$$

The functions $y_k : \mathbb{R}^N \rightarrow \mathbb{R}$, ($k = 1, \dots, k_{\max}$) represent the geometrical boundaries via their level sets $\{y_k(x) = 0\}$, and $U(y) : \mathbb{R} \rightarrow \mathbb{R}$ is the potential energy near each boundary. For concreteness, let us suppose this is a model for n spheres with centers at $x_i \in \mathbb{R}^3$, $i = 1, \dots, n$, so the configuration is $x = (x_{11}, x_{12}, x_{13}, x_{21}, \dots, x_{n3}) \in \mathbb{R}^{3n}$, and we take y_k to be the excess bond distance, as $y_k = \sqrt{|x_{i(k)} - x_{j(k)}|^2} - d$, where d is the particles' diameter. Then $U(y)$ is the pairwise interaction potential that we assume has minimum U_0 at $y = 0$ and is negligible beyond some cutoff r_c . For ease of exposition, the interaction potential is assumed to be identical between each pair of particles, but this is not a necessary restriction for the geometrical landscape to apply. For the total potential U^ϵ , the parameter ϵ characterizes the range of the potential, whereas C_ϵ is proportional to the depth. The geometrical limit requires taking $C_\epsilon \rightarrow \infty$ as $\epsilon \rightarrow 0$ in a manner that we specify momentarily.

We consider particles that evolve according to the overdamped Langevin dynamics (39) at temperature T ,

$$\frac{dx}{dt} = -\frac{1}{\gamma} \nabla U^\epsilon(x) + \sqrt{2D} \eta(t), \quad [2]$$

where γ is the friction coefficient, $D = (\beta\gamma)^{-1}$ is the diffusion coefficient, $\beta = (k_B T)^{-1}$, k_B is the Boltzmann constant, and $\eta(t)$ is a $3n$ -dimensional white noise. The equilibrium probability for this system is the Gibbs distribution (40),

$$d\rho^\epsilon(x) = (Z^\epsilon)^{-1} e^{-\beta U^\epsilon(x)} dx, \quad [3]$$

where Z^ϵ is the normalizing constant.

The geometrical free-energy landscape occurs when the range $\epsilon \rightarrow 0$. The relationship between the depth and the range is critical to obtaining an interesting limit. If only the range goes to zero, then particles are bonded for shorter and shorter times so that in the limit they simply behave like hard spheres. On the other hand, if the depth goes to $-\infty$ too quickly, then the particles simply stick together and unbind only on exponentially long timescales (41). The interesting limit occurs when particles stick to each other but unbind on accessible timescales; for this we must choose C_ϵ so that the Boltzmann factor for two particles to be bonded approaches a finite, nonzero constant, $\kappa = \lim_{\epsilon \rightarrow 0} \frac{1}{d} \int_0^{r_c} e^{-C_\epsilon \beta U(x/\epsilon)} dx$, where we define Boltzmann factors nondimensionally by scaling by the diameter d . Evaluating the integral using Laplace's method then implies

$$\kappa = \lim_{\epsilon \rightarrow 0} \frac{\epsilon e^{-\beta C_\epsilon U_0}}{d \sqrt{c \beta C_\epsilon U''(0)}}. \quad [4]$$

We call the constant κ the sticky parameter. Note that κ is a function of both the potential and the temperature. The constant $c = 2/\pi$ for hard spheres.

In the geometrical limit, the probability measure ρ^ϵ becomes concentrated at the exact locations in configuration space where a bond forms, i.e., on the level sets $\{y_k(x) = 0\}$ and all possible multiway intersections. Thus, the limiting probability distribution will be a weighted sum of delta functions, each defined on a manifold corresponding to a different set of bond constraints. The weight of each delta function depends on the number of bonds and a geometrical factor associated with the entropy of the configuration. This gives a geometrical picture of the energy landscape: When κ is large, the occupation probabilities will be dominated by configurations where the number

of bonds m is large. For identical particles, with $n \leq 9$, the maximum number of contacts is $m = 3n - 6$ (37, 38): These are rigid structures that have no internal degrees of freedom, so they correspond to zero-dimensional manifolds or "points". The next lowest configurations in potential energy are manifolds with $m = 3n - 7$, which are obtained from rigid structures by breaking one bond. These have one internal degree of freedom so are one-dimensional manifolds or "lines". The lines form the boundaries of two-dimensional manifolds or "faces", when another bond is broken, and continuing up in dimensionality we obtain the entire energy landscape as the union of manifolds of different dimensions.

Fig. 1 shows a schematic contrasting this limiting case with the traditional picture of an energy landscape. The traditional picture is of an undulating surface, with local minima connected through

Fig. 1. (Top) schematic of a traditional depiction of an energy landscape. Local minima are separated by energy barriers (red line) that govern the transition rate from one basin to another. (Middle) Schematic of a geometrical energy landscape, showing the zero- and one-dimensional sets. Local minima are infinitesimally narrow, deep points, separated by long, nearly flat lines. Along these the dynamics are governed mainly by diffusion, so the length of the line (in red) determines the transition rate. (Bottom) Example of a one-dimensional manifold from the landscape for $n = 6$, showing the transition path from polytetrahedron (Left) to octahedron (Right). The black line is the free energy $F_a/k_B T = -\log n_a h_a l - 11 \log \kappa$ along the one-dimensional manifold in units of $k_B T$, where we chose $\kappa = 20$. Black solid circles are the free energy $F_a/k_B T = -\log n_a h_a l - 12 \log \kappa$ for the zero-dimensional endpoints. Red crosses mark locations of clusters that are plotted explicitly (they are plotted with one-half their actual diameter for clarity), and the horizontal axis is a parameterization of the manifold in the quotient-space distance s .

saddle points, whose heights provide an activation barrier that determines the transition rates between basins. In contrast, in the geometrical limit, the local minima are infinitely narrow and deep, separated by long, relatively flat spaces in between—the landscape can be thought of as a golf course punctuated with deep trenches and very deep holes. Kinetics on this landscape are determined partly by an activation barrier—the time it takes to climb out of the hole—and partly by diffusion, or the time to cross the horizontal barrier.

Fig. 1 also shows an explicit one-dimensional manifold taken from the landscape for $n = 6$, an example we return to throughout the text. There are two ground states, the polytetrahedron and the octahedron (38), and the manifold is the set of deformations corresponding to the transition path between these when a single bond is broken.

To explicitly calculate the equilibrium probabilities of the different states in the geometrical landscape we consider a configuration with m constraints or equivalently $p \equiv 3n - 6 - m$ bonds broken. The constraints are written as an ordered multiindex $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)$ so the manifold of configurations satisfying such constraints is

$$\Omega_\alpha = \left\{ x : y_{\alpha_1}(x) = y_{\alpha_2}(x) = \dots = y_{\alpha_m}(x) = 0, y_\beta(x) > 0, \beta \neq \alpha_i \right\}. \quad [5]$$

We write $\alpha = 0$ to mean the region where no constraints are active and let $\Omega = \cup_\alpha \Omega_\alpha$ be the full space of accessible configurations. The limiting partition function associated with these constraints is

$$z_\alpha = \lim_{\epsilon \rightarrow 0} \frac{1}{d^{3n}} \int_{\Omega_\alpha^\epsilon} e^{-\beta U^\epsilon(x)} dx, \quad [6]$$

where Ω_α^ϵ is the neighborhood surrounding the manifold where the potential U^ϵ associated with the constraints is active:

$$\Omega_\alpha^\epsilon = \left\{ x : 0 \leq y_{\alpha_1}(x), \dots, y_{\alpha_m}(x) \leq \epsilon r_c, y_\beta(x) > \epsilon r_c, \beta \neq \alpha_i \right\}. \quad [7]$$

This splits configuration space near each manifold into two parts—the fast variables y_{α_i} changing rapidly along directions associated with the constraints (sometimes called the vibrational modes) and the slow variables y_β that are the unconstrained configuration.

To compute the integral in Eq. 6 we need a parameterization of the manifolds associated with the constraints. It is convenient to parameterize the fast directions by the constraint variables themselves, $x \rightarrow y_{\alpha_1}, \dots, y_{\alpha_m}$. We choose the additional $3n - m$ variables $y \in \mathbb{R}^{3n-m}$ so that $\nabla y \cdot \nabla y_{\alpha_i} = 0$ on Ω_α ; i.e., the variables y, y_{α_i} are orthogonal on the manifold. As discussed in *SI Appendix*, it is possible to find such a parameterization locally as long as the manifold associated with the constraint variables is regular—i.e., the coordinate transformation for the constraint variables must be smooth and invertible. This happens when the Hessian of the potential energy (say at $\epsilon = 1$)

$$H_\alpha(x) = \nabla \nabla U^{\epsilon=1}(x) = U''(0) \sum_{i=1}^m \nabla y_{\alpha_i} (\nabla y_{\alpha_i})^T \quad [8]$$

has m nonzero eigenvalues.

We can now evaluate

$$z_\alpha = \frac{1}{d^{3n}} \int e^m e^{-\beta C_\epsilon \sum_{i=1}^m U(Y_i)} \sqrt{|g_{ij}|} dY dy, \quad [9]$$

where $Y = (Y_1, \dots, Y_m)$ with $Y_i = y_{\alpha_i}/\epsilon$, g_{ij} is the metric tensor associated with the transformation $(y_{\alpha_i}, y) \rightarrow x$ with components $g_{ij} = J_{ki} J_{kj} = J^T J$, where $J_{ij} = \frac{\partial x_i}{\partial y_j}$, and $|g_{ij}|$ is its determinant. Let the metric tensor separate into blocks as

$$g = \begin{pmatrix} g_{ab} & g_{av} \\ g_{ub} & g_{uv} \end{pmatrix}. \quad [10]$$

The first set of indexes runs from $1 \leq a, b \leq m$ and describes the fast variables, perpendicular to the manifold, whereas the second set runs from $m+1 \leq u, v \leq 3n$ and describes the slow variables, along the constraint manifold. Let $|g_{cd}|$ be the determinant of a particular block. It follows from the definition of the metric that $|g_{ab}|_{Y=0} = \prod_{i=1}^m \lambda_i^{-1}$, where the λ_i s are the nonzero eigenvalues of $H_\alpha(x)/U''(0)$, and the condition of orthogonality gives $|g_{av}|_{Y=0} = |g_{ub}|_{Y=0} = 0$. Evaluating the integral in Eq. 9 over the fast variables using Laplace's method, and letting $\epsilon \rightarrow 0$, shows the limit is

$$z_\alpha = \kappa^m \frac{1}{d^{p+6}} \int_{\Omega_\alpha} h_\alpha(y) \sqrt{|g_\alpha|} dy, \quad [11]$$

where

$$h_\alpha(y) = \prod_{i=1}^m \lambda_i^{-1/2} \quad [12]$$

is a geometrical factor (representing the “vibrational” degrees of freedom) that depends only on the level sets of the constraints, and $g_\alpha = g_{uv}|_{Y=0}$ is the metric on manifold Ω_α inherited from the ambient space by restriction. The integral Eq. 11 is simply the volume integral of $h_\alpha(y)$ over Ω_α .

The manifold Ω_α contains 6 degrees of freedom representing translation and rotation of the cluster, and the partition function integral can be further simplified by integrating over the subspace spanned by these motions. This introduces a factor $I(x)$ in the integral, the square root of the moment of inertia tensor (40). If we let Ω_α^Q be the quotient space formed by identifying points $x \sim z$ if x can be mapped to z by a combination of translations and rotations, i.e., $\Omega_\alpha^Q = \Omega_\alpha / SE(3)$, where $SE(3)$ is the special Euclidean group, then we can write

$$z_\alpha = \kappa^m \zeta_\alpha = \kappa^m \frac{1}{d^p} \int_{\Omega_\alpha^Q} h_\alpha(x) I(x) \sqrt{|g_\alpha|} dx, \quad [13]$$

where $\bar{g}_\alpha$ is the metric on Ω_α^Q , and we have dropped constants (such as free volume) that are the same for all configurations. For convenience later on, we define ζ_α as the part of the partition function that is independent of κ . *SI Appendix* has a detailed discussion for how to construct $\bar{g}_\alpha$, which is critical for using the formalism developed here for practical calculations. Fig. 1 (Bottom) is an example of such a quotient manifold, where each point on the manifold represents the six-dimensional space of clusters in configuration space that are related by rotations or translations.

The particles in Fig. 1 are different colors to identify the different transitions that occur when moving around configuration space. However, when (as imagined here) the particles are identical, permuting the colors of any particular structure yields a geometrically isomorphic structure on a separate part of the quotient space. The free energy must account for the number n_α of distinct manifolds that are geometrically isomorphic to Ω_α^Q . When $p = 0$, this is $n_\alpha = C_0 n! / \sigma$, where σ is the symmetry number, i.e., the number of particle permutations that are

equivalent to a rotation, and $C_0 = 2$ if the structure is chiral and $C_0 = 1$ otherwise (42). For $p > 0$, we count the multiplicities by counting how many times a mode occurs from the perspective of each zero-dimensional “corner” of the mode and dividing by the total number of corners. (This is a combinatorial argument; it is equivalent to considering the molecular symmetry group for nonrigid molecules as in ref. 1, section 3.4.) For example, Fig. 1 has corners from two different ground states, the polytetrahedron and the octahedron, which occur with multiplicities n_1, n_2 respectively. For each polytetrahedron, there is a $\nu_1 = 1$ line coming out of it that is isomorphic to this transition, and for the octahedron there are $\nu_2 = 12$ distinct lines. (The numbers ν_1, ν_2 are indicated on the arrows connecting red circles to blue circles in Fig. 2, where the transition under consideration is mode 7.) Therefore, the total multiplicity of the line is $(n_1\nu_1 + n_2\nu_2)/2$. Consider a transition connecting a polytetrahedron to a distinct copy of itself, say mode 5. Here there are $\nu_1 = 4$ such lines connected to each polytetrahedron, so the multiplicity of the line is $n_1\nu_1/2$. In general, each p -dimensional manifold Ω_α contains a total of n_c corners from $N \leq n_c$ nonisomorphic ground states, each ground state having multiplicity n_i , and such that each single ground state is connected to ν_i distinct manifolds isomorphic to Ω_α , so the multiplicity is $n_\alpha = \sum_{i=1}^N n_i \nu_i / n_c$.

Putting this all together, the total partition function of all structures isomorphic to a given constraint manifold Ω_α^Q is $n_\alpha z_\alpha$, and the free energy of these structures is $F_\alpha = -k_B T \log(n_\alpha z_\alpha)$. We can separate this free energy into $F_\alpha = -mk_B T \log \kappa - k_B T \log(n_\alpha \zeta_\alpha)$, using the definition of ζ_α in Eq. 13. The first term ($-mk_B T \log \kappa$) depends on the temperature, bond energy, and width of the potential, whereas the second term with $S_\alpha = -k_B \log(n_\alpha \zeta_\alpha)$ is entirely geometrical and essentially is the entropy of structures corresponding to the constraint set Ω_α^Q . As an example, we have plotted $F_\alpha/k_B T$ along the polytetrahedral-octahedral transition in Fig. 1 (Bottom). This varies smoothly along the manifold as $I(x), h(x)$ vary. The endpoints of the manifold are the ground states, where the free energy changes discontinuously because the number of bonds has changed—this causes a jump in the energetic part via m and in the entropic part via h_α .

With these results in hand, we can now compare the total entropies of floppy manifolds of different dimensions, to understand the temperature range in which the different manifolds

occur. Let the total geometrical partition function of manifolds of dimension p be

$$Z_p = \sum_{\dim(\Omega_\alpha^Q)=p} n_\alpha \zeta_\alpha, \quad Z = \sum_p \kappa^{3n-6-p} Z_p. \quad [14]$$

Here Z_p is independent of the temperature and potential, whereas Z combines everything to obtain the entire landscape. Note that lower-dimensional manifolds have more bonds and thus are favored in the partition function when the temperature (or equivalently κ) is small. As temperature increases, κ shrinks and higher-dimensional manifolds become more highly populated. Eventually clusters will fall apart into single particles. The temperature dependence of how clusters fall apart is encoded in the relative sizes of the Z_p s. The temperature where the landscape transitions from having more p -dimensional structures than $p+1$ -dimensional structures is found by solving $\kappa Z_p = Z_{p+1}$ for κ , which gives roughly $(k_B T_p)^{-1} \approx \ln Z_{p+1}/Z_p + \text{const.}$

Free-Energy Landscape for Identical Particles

To illustrate our asymptotic calculations with a concrete example we have computed the geometrical manifolds up to dimension $p = 2$ for $n = 5, 6, 7, 8$ identical particles with diameter $d = 1$. To do this, we begin with the set of clusters with $\geq 3n - 6$ bonds derived in Arkus et al. (38) For every rigid cluster we break each single bond in turn and move along the internal degree of freedom until we form another bond. This is the set of one-dimensional manifolds. For the 2D manifolds, we break each pair of bonds from the rigid clusters in turn and move along the internal degrees of freedom to compute the boundaries, corners, and interior of the 2D manifolds (details in SI Appendix). This algorithm ensures we have every floppy manifold that can eventually access one of the rigid clusters in our list only by forming bonds, but never breaking them. Our analysis makes three assumptions that we believe to be true, but await rigorous proof: First, we are assuming that the list of clusters in Arkus et al. (38) is the complete set of rigid (zero-dimensional) clusters; this is true as long as there are no rigid clusters with $3n - 7$ bonds or less, a condition that was not checked. (The existence of such pathological examples is not ruled out purely by rank constraints on the Jacobian as there could be singular structures; see e.g., the

Fig. 2. Free-energy landscape for $n = 6$ with zero, one, and two bonds broken. Red circles are zero-dimensional modes, blue circles are one-dimensional modes, and yellow circles are two-dimensional modes. The area of each circle is proportional to the geometrical partition function $n_\alpha \zeta_\alpha$ of each mode and hence to its probability in equilibrium relative to modes of the same dimension. Modes are identified by numbers and arrows show the connectivity: An arrow from mode i to mode j indicates that mode i is part of the boundary of mode j . The number on each arrow indicates the number of different manifolds of type j that are connected to a single manifold of type i . The computed parameterizations are shown for each of the two-dimensional modes.

Fig. 3. Two-dimensional manifold (mode 18, $n = 6$), parameterized in the plane, with selected points on the manifold plotted as clusters. The vertex of each triangle represents a cluster and black or red solid circles indicate the ones that are plotted. The corners (black solid circles) are rigid structures, or zero-dimensional manifolds. The edges are one-dimensional manifolds and points on these are obtained from rigid structures by breaking one bond, whereas points in the interior form a two-dimensional manifold and are obtained by breaking two bonds. The one-dimensional manifolds, beginning at the octahedron (left) and moving clockwise, are 7, 5, 5, 7. The text indicates the type of bond that breaks or forms as one moves along each edge.

examples in ref. 43.) Second, in the calculations of the entropy of the 2D floppy manifolds we assume that the manifolds are topologically equivalent to a disk. The fact that our parameterization algorithm works is evidence for this claim, although we have not proved this rigorously. Third, we assume that all floppy manifolds can eventually access a rigid mode and are not, for example, circles. We mention these caveats because although we are confident that they do not apply in the low n examples described here, it is possible that potentially significant exceptions arise at higher n .

The landscape for $n = 6$ is shown in Fig. 2. (Quantitative summaries are given in the *SI Appendix*.) There are two ground states, denoted by the red circles, each with $3n - 6 = 12$ contacts; the area of the red circles is proportional to the probability of each state, with the polytetrahedral ground state ≈ 25 times more likely than the octahedral. The light blue circles denote the 5 topologically unique structures that are missing a single bond in the ground states—such structures correspond to a one-dimensional manifold, with continuous deformation along the direction of the missing bonds. The yellow circles denote the 13 unique structures

that are missing two bonds from the ground state. Again, the area of the circles is proportional to the occurrence probability of these modes. These structures correspond to 2D manifolds, with continuous deformations allowed along both of the directions between the missing bonds. The connections between the different modes are denoted by arrows in Fig. 2, with structures missing (say) two bonds generally arising from breaking a single bond from several different pathways.

Each element of a mode with two bonds broken can be mapped to a polygon in $\mathbb{R}^2$, and these parameterizations are also shown in Fig. 2. We have chosen one parameterization (mode 18) to illustrate in detail in Fig. 3. The interior of the manifold represents structures with 10 bonds, with each point representing a different set of coordinates for the particles. The edges correspond to structures with 11 bonds, whereas the corners are structures with a full $3n - 6 = 12$ bonds. Of the four corners, three correspond to polytetrahedra with different permutations of the particles, and the final one is an octahedron. The one-dimensional edges connecting the corners are possible transition paths that can be followed by breaking only one bond.

In general, the number of corners varies among modes, with no a priori way to determine this without solving the full geometry problem. It ranges from three to six for $n = 6, 7$ and from three to seven for $n = 8$. Many two-dimensional modes contain several permutations of a given zero-dimensional mode as a corner.

Table 1 summarizes the partition function data. The number of different modes increases combinatorially with n , as do the geometrical partition functions Z_p . Strikingly, the ratios Z_1/Z_0 , Z_2/Z_1 remain virtually constant as n increases. This implies that the temperature dependence of the landscape is independent of n . Indeed, Fig. 4 shows the relative probabilities of zero-, one-, and two-dimensional modes as the temperature varies, for $n = 6, 7, 8$, with the yield of a p -dimensional mode given by $y_p = \kappa^{2-p} Z_p / (\kappa^2 Z_0 + \kappa Z_1 + Z_2)$. As an illustration, we have chosen $U_0/k_B = -4$, $(U''(0)/k_B)(2/\pi) = 15$, so that $\kappa(T) = e^{4/T} / \sqrt{15/T}$. Because the ratios Z_{p+1}/Z_p are nearly the same, these graphs are essentially indistinguishable for different numbers of particles. Moreover, the critical temperature for transitioning from mostly zero-dimensional structures to one-dimensional structures ($\approx 1/\ln Z_1/Z_0$) is quite close to that for transitioning from one-dimensional structures to two-dimensional structures ($\approx 1/\ln Z_2/Z_1$). If this remains true as p increases, it would imply that clusters melt explosively at some critical temperature, rather than incrementally: Clusters would occupy either mostly the zero-dimensional modes or a gaseous, no- or few-bond-state, but not the chain-like floppy configurations in between.

Kinetics on the Geometrical Landscape

We now consider kinetics on the geometrical landscape. The concentration of equilibrium probabilities on manifolds with varying dimensions also applies to nonequilibrium quantities, such as transition rates, first-passage times, the evolution of probability density, etc. Such quantities can be computed from the time-dependent transition probabilities, which for dynamics

Table 1. Geometric partition functions Z_p , and numbers of different modes, for the set of $p = 0, 1, 2$ -dimensional manifolds as the number of spheres n varies

n	No. 0D	No. 1D	No. 2D	Z_0	Z_1	Z_2	Z_1/Z_0	Z_2/Z_1
5	1	2	4	10.7	73.8	545	6.9	7.4
6	2	5	13	36.1	256	1,140	7.1	4.5
7	5	16	51	1.1×10^3	8.5×10^3	39×10^3	7.6	4.6
8	13	75	281	49×10^3	396×10^3	1.87×10^6	8.1	4.7

The Z_p are geometrical quantities that do not depend on the temperature or potential. The total partition function of the p -dimensional manifolds, which includes both of these dependencies via κ , is $\kappa^m Z_p$.

given by Eq. 2 are obtained from the corresponding forward or backward Fokker–Planck (FP) equations (44). We show that in the geometrical limit, the FP equation asymptotically approaches a hierarchy of FP equations, one on each manifold of each dimension. These equations are coupled to each other, with equations on manifolds with dimension p serving as boundary conditions for the equations on manifolds with dimension $p + 1$.

The idea behind the derivation is quite natural: If a potential is deep and narrow, then it equilibrates much more rapidly in the directions along the bonds than along a cluster's internal degrees of freedom. Therefore, the probability density near a cluster with p bonds broken is approximately $p(y, t)e^{-\beta \sum_i U^\epsilon(y_{a_i})}$, where y parameterizes the internal degrees of freedom of the cluster. The “constant” $p(y, t)$ evolves slowly in the transverse directions according to the Fokker–Planck dynamics (which include an “effective” potential that arises from the curvature of the manifold), and it also changes due to the flux of probability out of the $p + 1$ -dimensional manifolds for which it forms part of their boundary. This gives a hierarchy of coupled FP equations.

To see how this comes about in detail, we examine solutions to the Fokker–Planck equation for the evolution of the probability density $p^\epsilon(x, t)$. Given a parameterization of configuration space $\mathbb{R}^{3n}$ with metric tensor g , the nondimensional Fokker–Planck equation corresponding to Eq. 2 is

$$\begin{aligned} \partial_t p^\epsilon &= \text{div}(p^\epsilon \text{grad } U^\epsilon + \text{grad } p^\epsilon) \\ &= \frac{1}{\sqrt{|g|}} \partial_i \left(\sqrt{|g|} (p^\epsilon g^{ij} \partial_j U^\epsilon + g^{ij} \partial_j p^\epsilon) \right), \end{aligned} \quad [15]$$

with boundary conditions at each level set $\{y_i(x) = 0\}$

$$(p^\epsilon \text{grad } U^\epsilon + \text{grad } p^\epsilon) \cdot \hat{n}^{(i)} = (p^\epsilon g^{ij} \partial_j U^\epsilon + g^{ij} \partial_j p^\epsilon) \hat{n}_j^{(i)} = 0,$$

where $\hat{n}^{(i)}$ is the outward normal to the boundary. We have nondimensionalized lengths by d , times by d^2/D , and energy by $k_B T$. Away from all boundaries there is no force, so in Ω_0 the limiting probability p evolves only by diffusion as

$$\partial_t p = -\text{div } j_0, \text{ with } j_0 = -\text{grad } p. \quad [16]$$

Now consider the evolution near a manifold Ω_α , a “wall”. The dynamics in the directions orthogonal to the wall, where bond distances are changing, are much faster than those along it, so near the wall the probability density will rapidly approach a multiple of the equilibrium probability. Parameterizing the region Ω_α^ϵ near the wall as (y_{a_i}, y) as in the previous section, we obtain

$$p^\epsilon(y_{a_i}, y, t) = p(0, y, t) e^{-\sum_{i=1}^m U^\epsilon(y_{a_i})} + \epsilon p_1 e^{-\sum_{i=1}^m U^\epsilon(y_{a_i})}, \quad [17]$$

where $\epsilon p_1(y, y_{a_i}, t)$ is the correction to the leading-order formula. This satisfies the matching condition that p^ϵ be asymptotically continuous. This ansatz can also be derived from a consistent asymptotic expansion of Eq. 15 after the change of variables $Y_{a_i} = y_{a_i}/\epsilon$.

Substituting Eq. 17 into the Fokker–Planck Eq. 15 gives

$$\partial_t(e^{-U^\epsilon} p) = \text{div}(e^{-U^\epsilon} \text{grad } p) + O(\epsilon). \quad [18]$$

We want to integrate out the fast variables so we need to separate these from the slow ones. This is most conveniently done using the metric g with block decomposition (Eq. 10). Making the change of variables $Y_i = y_{a_i}/\epsilon$ shows that Eq. 18 can be written as

$$\begin{aligned} \partial_t \left(\sqrt{|g|} e^{-U_\alpha} (p + p_1) \right) &= \frac{1}{\epsilon^2} \partial_a \left(\sqrt{|g|} e^{-U_\alpha} g^{ab} \partial_b (p + \epsilon p_1) \right) \\ &+ \frac{1}{\epsilon} \partial_a \left(\sqrt{|g|} e^{-U_\alpha} g^{av} \partial_v (p + \epsilon p_1) \right) \\ &+ \frac{1}{\epsilon} \partial_u \left(\sqrt{|g|} e^{-U_\alpha} g^{ub} \partial_b (p + \epsilon p_1) \right) \\ &+ \partial_u \left(\sqrt{|g|} e^{-U_\alpha} g^{uv} \partial_v (p + \epsilon p_1) \right), \end{aligned} \quad [19]$$

where we abbreviate $U_\alpha = \sum_{i=1}^m C_\epsilon U(Y_i)$.

We now integrate Eq. 19 over the fast variables $\epsilon^m dY_1 dY_2 \dots dY_m$ and keep the leading-order parts. The terms α_ϵ^{-1} vanish in the limit because $g^{av}|_{Y=0} = g^{ub}|_{Y=0} = 0$. The $O(1)$ terms require evaluating an integral similar to Eq. 9, which converts $\sqrt{|g|} e^{-U_\alpha}$ to the factor $\sqrt{|g_{uv}|} \kappa^m h_\alpha$ at each point on the manifold. The first term is the most interesting. This is the divergence in the fast variables, and although p does not depend on these, the unknown p_1 might and contributes at $O(\epsilon^{-1})$. However, we can use the divergence theorem to replace this term with an integral of the flux through the p -dimensional fast-variable boundary, at $\{x : Y_i(x) = r_c\}$. We then introduce a second matching condition that requires the flux to be continuous, so we can replace it with the sum of fluxes from each $p + 1$ -dimensional manifold Ω_β that has Ω_α as part of its boundary and evaluate the limit of the boundary integral for these matched fluxes. [Specifically, we require that $\epsilon \tilde{j}_\alpha \cdot \hat{n}^{(c)}|_{Y_i=r_c} = \tilde{j}_\beta \cdot \hat{n}^{(i)}|_{y_{a_i}=0}$, where $\hat{n}^{(c)}$ is the normal to level set $Y_i = r_c$ and $\hat{n}^{(i)}$ is the normal to level set $y_i = 0$, and $\tilde{j}_\alpha = -e^{-U_\alpha} g^{ab} \partial_b (p + \epsilon p_1)$ with a similar expression for $\tilde{j}_\beta$.]

Finally, we integrate over the space of rotations and translations of each point, assuming $p(x, 0)$ is constant on orbits. The divergence in these directions will disappear by Stokes' theorem, and the remaining directions provide dynamics on the quotient space. Combining with the previous calculations yields

Fig. 4. Relative yields of zero-dimensional, one-dimensional, and two-dimensional modes (neglecting all higher-dimensional modes) for $n = 6, 7, 8$. The yield for p -dimensional modes is calculated as $y_p = \kappa^{2-p} Z_p / (\kappa^2 Z_0 + \kappa Z_1 + Z_2)$ with $\kappa(T) = e^{-(U_0/k_B)/T} / \sqrt{(U''(0)/k_B)(2/\pi)/T}$. We used $U_0/k_B = -4$, $(U''(0)/k_B)(2/\pi) = 15$, but the numbers do not change the qualitative shape. Note that modes with dimensions greater than two will become important at the higher temperatures.

$$\begin{aligned}\partial_t(\kappa_\alpha p) &= \frac{1}{\sqrt{|\bar{g}_{uv}|}} \partial_u \left(\sqrt{|\bar{g}_{uv}|} \kappa_\alpha \bar{g}^{uv} \partial_v p \right) + \sum_{\beta \rightarrow \alpha} j_\beta \cdot \hat{n}^{\beta\alpha} \\ &= \text{div}_\alpha(\kappa_\alpha \text{grad}_\alpha p) + \sum_{\beta \rightarrow \alpha} j_\beta \cdot \hat{n}^{\beta\alpha},\end{aligned}\quad [20]$$

where the fluxes have leading-order part

$$j_\alpha = -\kappa_\alpha \text{grad}_\alpha p \quad \text{on } \Omega_\alpha^Q. \quad [21]$$

Here $\kappa_\alpha \equiv \kappa^n h_\alpha I$ combines the sticky parameter and the geometric factor at the wall, and the metric tensor $\bar{g}$ is the quotient metric obtained from the metric g_α on Ω_α^Q , which in turn is inherited from the original metric g in the ambient space by restriction: $g_\alpha = g|_{Y_1=\dots=Y_m=0}$. The sum is over β such that Ω_α^Q is part of the boundary of the $p+1$ -dimensional manifold Ω_β^Q and where $\hat{n}^{\beta\alpha}$ is an outward normal vector to Ω_β^Q at Ω_α^Q . We define $\text{div}_\alpha, \text{grad}_\alpha$ to be the differential operators on the quotient manifold Ω_α^Q .

Eq. 20 has a more intuitive interpretation as the evolution of the probability along the wall. It is clear from the derivation that the total probability density (with respect to the wall coordinates) of being on the wall is $P_\alpha \equiv \kappa_\alpha p$. This satisfies

$$\begin{aligned}\partial_t P_\alpha &= \frac{1}{\sqrt{|\bar{g}_{uv}|}} \partial_u \left(\sqrt{|\bar{g}_{uv}|} \left(-P_\alpha \bar{g}^{uv} \partial_v \log \kappa_\alpha + \bar{g}^{uv} \partial_v P_\alpha \right) \right) + j_\beta \cdot \hat{n}^{\beta\alpha} \\ &= \text{div}_\alpha \left(\underbrace{-P_\alpha \text{grad}_\alpha \log \kappa_\alpha}_{\text{effective force}} + \underbrace{\text{grad}_\alpha P_\alpha}_{\text{diffusion}} \right) + \underbrace{j_\beta \cdot \hat{n}^{\beta\alpha}}_{\text{flux to/from wall}}.\end{aligned}\quad [22]$$

The dynamics along the wall are therefore a combination of diffusion, plus drift due to an effective potential $-\log \kappa_\alpha$, plus a flux in and out of the wall.

The effective potential is entropic in nature and comes from the changing wall curvature, which makes the potential look wider in some places than in others. A particle will spend more time in the wide places than in the narrow ones, and because it reaches equilibrium much more quickly in the transverse directions than in the along-wall directions, it looks like there is an effective force pushing it to the wider areas. This is the same equation one obtains by letting the depth of the potential become infinite without changing the width (41), but with the addition of the flux in and out of the wall.

This flux term is illustrated schematically in Fig. 5 for a simple case where the configuration space $(x, y) \in \mathbb{R}^2$ has a single con-

Fig. 5. Schematic showing the asymptotic boundary condition near a one-dimensional boundary. (Left) Gray shading indicates the depth of the potential, and the dashed line indicates the boundary of the outer and inner solutions. By considering the total probability flux in/out of each small-volume element near the boundary (red box), one can replace the detailed dynamics near the boundary in the limit as $\epsilon r_c \rightarrow 0$ with an effective boundary condition (Right).

straint, $y_1(x, y) = x$ —a wall at the horizontal axis. The probability integrated over a small box (red) with length Δx near the wall changes in a time increment Δt due to two processes: probability fluxing along the wall in the x -coordinate, which contributes a change of $\Delta t(p_x(x + \Delta x, t) - p_x(x, t)) \int e^{-U(y)} dy$, and the flux of probability from the wall to the interior, which contributes a change of $\Delta t \Delta x p_y(x, 0, t)$. Equating with $\Delta t p_t(x, t) \int e^{-U(y)} dy$ gives the effective boundary condition.

To summarize, the limiting FP equation and boundary conditions are Eq. 16, plus Eq. 20 on every Ω_α^Q . Substituting for the time derivatives shows that the boundary conditions are second order, and this is why conditions are needed on every boundary and not only on those with codimension 1. We call this set of equations the “sticky” equations because they are the Fokker–Planck equations for sticky Brownian motion (45), a stochastic process that has a probability atom on the boundary of its domain.

Transition Rates for Sticky Brownian Motion

To transition between the rigid configurations in the geometrical limit, a cluster must break one (or more) bonds and then diffuse across the line segment (or face, etc.) until it hits the other endpoint. The time it takes to do this depends on the length of the line and on how the entropy of the configuration varies along the line, and we can find this by solving an equation on the full line segment (face, etc.). In our asymptotic limit there are no meaningful transition “states”—rather, the entire line segment can be thought of as a transition state. Once the energetic barrier of breaking a bond has been overcome, transitions are dominated by diffusion.

We now consider how to compute transition rates using the sticky equations (Eq. 20), supposing we have a stochastic process $X(t)$ whose probability evolution is well approximated by these. (Note that we have not actually constructed a process that satisfies this in the strong sense. However, we began with a process satisfying Eq. 2 and showed the sticky equations describe its probability evolution asymptotically so we use these to compute rates.) Consider the transition rate between sets $A \subset \Omega$, $B \subset \Omega$, and for simplicity let us focus only on the case where these are both (disjoint) subsets of the zero-dimensional manifolds. For example, we may be interested in the transition rate between the octahedron and the polytetrahedron (introduced in Fig. 1), in which case A would contain all points in the quotient space representing the octahedron and B all those representing the polytetrahedron. These rates can be computed using transition path theory, which provides a mathematical framework for computing transition rates directly from the Fokker–Planck equations. We simply state the facts that are relevant to our example and refer the reader to other resources for more details (46–48).

The committor function $q(x)$ is the probability, starting from point x , of reaching set B before set A . This solves the stationary backward Fokker–Planck equations with boundary conditions $q(A) = 0$, $q(B) = 1$, plus any other boundary conditions remaining from the equations. As shown in SI Appendix, the forward sticky equations (Eq. 20) are self-adjoint (with respect to the invariant measure) so these are also the backward sticky equations.

A reactive trajectory is a segment of the path $X(t)$ that hits B before A going forward in time and A before B going backward in time. The probability current of reactive trajectories is a vector field that, when integrated over a surface element, gives the net flux of reactive trajectories through it. Because our process is time reversible, this current is (47)

$$J(x) = \rho(x) \nabla q(x), \quad [23]$$

where $\rho(x)$ is the equilibrium probability measure. From Eq. 13 we find this is

$$\rho(x) = Z^{-1} \sum_{\alpha} \kappa_{\alpha}(x) \delta_{\alpha}(x), \quad [24]$$

where $\delta_{\alpha}(x)$ is the singular measure on Ω_{α} ; i.e., it satisfies $\int_{\Omega} f(x) \delta_{\alpha}(x) = \int_{\Omega_{\alpha}} f(x)$. Finally, the transition rate k^{AB} is calculated by integrating this flux over a surface S dividing the two states, giving

$$k^{AB} = \int_S J \cdot \hat{n} \, dS, \quad [25]$$

where $\hat{n}$ is the normal pointing from the side containing A into the side containing B .

Computing this exactly requires solving the backward equations on a high-dimensional space and integrating over a high-dimensional surface—a computationally infeasible proposition. However, when the sticky parameter κ is large, most of the probability is concentrated on the lowest-dimensional manifolds so we expect these to contribute the most to the rates. Therefore, let us expand the equations in powers of κ^{-1} . We suppose all variables have an asymptotic expansion as

$$k^{AB} = k_0^{AB} + \kappa^{-1} k_1^{AB} + \dots, \quad [26]$$

and similarly for q, μ, ρ, J , etc. Expanding ρ shows that to first order it is a measure on points, to second order it is a measure on points and lines, etc. Measures on points will not contribute to the rate because the dividing surface S can be chosen to avoid them, so $k_0^{AB} = 0$ and the leading-order part of the rate is $O(\kappa^{-1})$, computed from $\rho_1 \nabla q_0$.

Expanding the backward sticky equations in powers of κ^{-1} gives a set of equations for q_0 ,

$$\text{div}_{\alpha} h_{\alpha} I \text{grad}_{\alpha} q_0 = 0 \quad \text{on } \Omega_{\alpha} \quad (p > 0), \quad [27]$$

with boundary conditions $q_0(A) = 0$, $q_0(B) = 1$, and $\sum_{\beta \rightarrow \alpha} (h_{\beta} I \text{grad}_{\beta} q_0) \cdot \hat{n}^{\beta\alpha} = 0$ on all other zero-dimensional manifolds.

To solve these equations we can first find the solution on the zero- and one-dimensional manifolds and then use this as a boundary condition for the solution on the two-dimensional manifolds, which becomes in turn a boundary condition for the solution on the three-dimensional manifolds, etc. The leading-order rate requires only the solution for $p = 0, 1$. If we enumerate the lines connecting a point $a_k \in A$ to a point $b_k \in B$ and use an arc-length parameterization for the k th line whose total length is s_k , then this is given analytically as $q_0(s) = Q_k^{-1} \int_{s=0}^s (h_{\alpha}(s') I(s'))^{-1} ds'$, where

$$Q_k = \int_{s=0}^{s_k} (h_{\alpha}(s') I(s'))^{-1} ds' \quad [28]$$

is the normalization factor. On all other lines $q_0(s) = 0$.

Any dividing surface S hits each line at a single point, so the leading-order rate (in dimensional units) is

$$k_{AB} = \kappa^{-1} k_1^{AB} = \kappa^{-1} \frac{D}{d^2} Z_0^{-1} \sum_k Q_k^{-1}, \quad [29]$$

where the sum is over all connecting lines. This is asymptotically equivalent to the rate one would obtain simply by restricting the

Fig. 6. Rates (geometrical components) at leading order, for $n = 6, 7, 8$. Dimensional rates are $\kappa^{-1} D/d^2$ times the above. These rates indicate the total number of each type of transition one expects to see, per unit time.

full committor function and invariant measure to the set of zero- and one-dimensional manifolds.

Transition Rates for Hard Spheres. We used the formalism to compute the leading-order rates for $n = 6, 7, 8$ hard spheres with diameter $d = 1$. To do this we took the set of one-dimensional solutions computed as part of the free-energy landscape and computed the factor Q_k from Eq. 29 on each manifold. Summing over all of the one-dimensional manifolds that connect a zero-dimensional manifold numbered a to a zero-dimensional mode numbered b gives the transition rate between a and b . We also include transitions between different ground states belonging to the same mode (e.g., $a \rightarrow a$), by multiplying the previous calculation by 2 because transitions can go in either direction along the line.

Fig. 6 shows the network of zero-dimensional states and the reaction rates between each state. The numbers reported are the dimensionless, purely geometrical parts of the rates $Z_0^{-1} \sum_k Q_k^{-1}$ and should be multiplied by $\kappa^{-1} D/d^2$ to give the dimensional rate. These rates, when multiplied by the total time of a simulation or experiment, give the average number of transitions one would expect to observe, so they are equal for both directions $a \rightarrow b$ and $b \rightarrow a$ because our system is time reversible. To obtain the rate relative to a particular state, i.e., the rate at which one leaves state a to visit state b next, given that the last state visited was state a , one should divide by the so-called reactive probability of a (49). To leading order, this is equivalent to dividing by the equilibrium probability of a .

Simulations. We have verified our results by performing Brownian dynamics simulations of Eq. 2 with a short-range Morse potential. The results agree very well with our calculations of both free-energy and transition rates. Fig. 7 shows a comparison of the simulated probabilities vs. theoretical probabilities of each mode for $n = 6$ (see SI Appendix for $n = 7, 8$), for particles interacting with a Morse potential with range parameter $\rho = 30$, a range of $\sim 5\%$ of the particle diameter. The agreement is nearly perfect. Fig. 7 also compares the number of each type of transition we saw in the simulations to that predicted from theory. The theory slightly overpredicts the total number of transitions—this is what one

should expect from the geometrical picture, as these leading-order rates neglect the time the system spends in the floppy manifolds, which would tend to slow it down.

These results are encouraging partly because they are evidence that we have executed these calculations correctly, but also because they suggest the asymptotic limit may apply for experimental systems, such as Meng et al. (4) where the potential reportedly had roughly this width.

Comparison with Other Numerical Approaches to the Free-Energy Landscape. We have compared our results with those obtained by a numerical study that directly searched an energy landscape of a short-ranged potential (a Morse potential, with range roughly 0.05 particle diameter) for local minima and transition states (42). The numerical method found fewer local minima than there are zero-dimensional modes and fewer transition states than one-dimensional modes, suggesting that the asymptotic theory has the roughest landscape. By computing adjacency matrices for the numerical states with a cutoff bond distance of $1 + 1 \times 10^{-2}$, we verify that for $n = 6, 7, 8$ each local minimum corresponds to a unique zero-dimensional mode, and each transition state lies on a unique one-dimensional manifold.

We identify the point on the one-dimensional manifold that is closest to each transition state. The transition states are very close to the local maxima of the vibrational factor $-\log h(x)$ (see SI Appendix for plot), consistent with it being a saddle point of the potential energy. We believe the small discrepancy in location can be attributed to the finite width of the Morse potential used in the numerical procedure.

The missing zero-dimensional modes occur for manifolds that are very close to each other in the quotient space metric (separation ≈ 0.08), which is the case for modes $\{1, 4\}$ ($n = 7$) and modes $\{1, 2, 3, 4\}$ ($n = 8$). For these only one local minimum was found for the entire group. We hypothesize that because the separation is within the range of the potential, the zero-dimensional modes merge to form a single local minimum.

The missing one-dimensional modes often, but not always, correspond to self-self transitions—these transitions do not matter when particles are identical; however, they will account for transitions between different states when the particles are not all the same (e.g., ref. 50). For example, for $n = 8$ the numerical procedure identifies 45 transition states, compared with our 75 one-dimensional manifolds. Of the missing manifolds, 16 are self-self transitions, 9 are nonself-nonself transitions within the group $\{1, 2, 3, 4\}$, and 5 are nonself-nonself transitions for endpoints not both in the group.

Discussion/Conclusions

We have developed a unique framework for understanding energy landscapes when particles interact with a short-ranged potential. We show that in the limit as the range goes to zero and the depth goes to $-\infty$, the energy landscape becomes entirely governed by geometry, with a single parameter κ encapsulating details about the potential and temperature. When κ is large, only the lowest-dimensional geometrical manifolds contribute significantly to the landscape and this makes a computational approach tractable. To illustrate the limit, we have computed the set of low-dimensional manifolds for $n \leq 8$ hard, spherical particles. This solution to a nontrivial problem in statistical mechanics can be used to compute equilibrium or nonequilibrium quantities for any potential whose range is short enough.

We were able to calculate this set of low-dimensional manifolds because we began with the set of rigid clusters and made the conjecture that all floppy modes can be accessed from these by breaking bond constraints. Solving for the complete set of rigid clusters is a difficult problem in discrete geometry that has been done only for $n \leq 11$ (38, 51), but with current computational power and novel approaches (51, 52) one can anticipate reaching larger n . Very large

Fig. 7. (Lower) Theoretical probabilities of each mode (relative to modes of the same dimension), vs. probabilities computed from simulations using a potential with width $\approx 5\%$ of particle diameter, for $n = 6$. The red line is theory = simulation, indicating a perfect match. (Upper) Number of transitions of each type, for simulation and theoretical prediction. The theoretical prediction indicates expected 95% confidence intervals, computed using a normal approximation to a binomial. The sticky parameter was $\kappa = 16$ and the total running time was 2.3×10^3 units.

n will eventually require making approximations to the geometry problem. We speculate that as n increases, structures with extra bonds, as well as “singular” structures whose Jacobians have extra zero eigenvalues, will come to dominate the landscape—these have not yet been considered in our asymptotic framework but they are observed with high probability in experiments (4).

We have compared our results to those obtained by numerically searching the free-energy landscape of a short-ranged Morse potential for local minima and transition states. Our method finds more minima and transition regions of the potential energy than the numerical search procedure, and this points to a potentially useful extension of our theory—one can imagine starting with the limiting geometrical manifolds and following these in some way as the range of the potential is increased, to obtain a low-dimensional approximation to the free-energy landscape for finite-width potentials, such as Lennard–Jones or Van der Waals clusters. This method would

overcome a major issue with numerical searches, which is that there is no way to ensure that all important parts of the landscape have been found—we claim that our manifolds are the complete set of low-energy states so they will remain so under small enough perturbations. In addition, this would provide a way to deal with the increasing ruggedness of energy landscapes with short-ranged potentials, which are a challenge for numerical methods—we start with the most rugged landscape and would need only to smooth it.

ACKNOWLEDGMENTS. We thank David Wales for generously providing data and Vinodhan Manoharan and Eric Vanden-Eijnden for helpful discussions. This research was funded by the National Science Foundation through the Harvard Materials Research Science and Engineering Center (DMR-0820484), the Division of Mathematical Sciences (DMS-0907985), and the Kavli Institute for Bionano Science and Technology at Harvard University. M.P.B. is an investigator of the Simons Foundation.

1. Wales DJ (2003) *Energy Landscapes* (Cambridge Univ Press, Cambridge, UK).
2. Stillinger F, Weber T (1984) Packing structures and transition in liquids and solids. *Science* 225:983–989.
3. Beberg AL, Ensign DL, Jayachandran G, Khaliq S, Pande VS (2009) Folding@home: Lessons from eight years of distributed computing. *IEEE International Symposium on Parallel & Distributed Processing*, May 23–29, 2009, IEEE Computer Society.
4. Meng G, Arkus N, Brenner MP, Manoharan VN (2010) The free-energy landscape of clusters of attractive hard spheres. *Science* 327(5965):560–563.
5. Bowman GR, Pande VS (2010) Protein folded states are kinetic hubs. *Proc Natl Acad Sci USA* 107(24):10890–10895.
6. Doye JPK, Wales DJ (1996) The structure and stability of atomic liquids: From clusters to bulk. *Science* 271:484–487.
7. Stillinger FH (1995) A topographic view of supercooled liquids and glass formation. *Science* 267(5206):1935–1939.
8. Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Theory of protein folding: The energy landscape perspective. *Annu Rev Phys Chem* 48:545–600.
9. Liwo A, Lee J, Ripoll DR, Pillardy J, Scheraga HA (1999) Protein structure prediction by global optimization of a potential energy function. *Proc Natl Acad Sci USA* 96(10):5482–5485.
10. Wolynes PG (2005) Recent successes of the energy landscape theory of protein folding and function. *Q Rev Biophys* 38(4):405–410.
11. Rothmund PW (2006) Folding DNA to create nanoscale shapes and patterns. *Nature* 440(7082):297–302.
12. Clementi C (2008) Coarse-grained models of protein folding: Toy models or predictive tools? *Curr Opin Struct Biol* 18(1):10–15.
13. Maragliano L, Cottone G, Cicotti G, Vanden-Eijnden E (2010) Mapping the network of pathways of CO diffusion in myoglobin. *J Am Chem Soc* 132(3):1010–1017.
14. Swendsen RH, Wang JS (1986) Replica Monte Carlo simulation of spin glasses. *Phys Rev Lett* 57(21):2607–2609.
15. Elber R, Karplus M (1987) A method for determining reaction paths in large molecules: Application to myoglobin. *Chem Phys Lett* 139:375–380.
16. Weinan E, Ren W, Vanden-Eijnden E (2002) String method for the study of rare events. *Phys Rev B* 66:052301.
17. Bolhuis PG, Chandler D, Dellago C, Geissler PL (2002) Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu Rev Phys Chem* 53:291–318.
18. Hummer G (2004) From transition paths to transition states and rate coefficients. *J Chem Phys* 120(2):516–523.
19. Abrams CF, Vanden-Eijnden E (2010) Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics. *Proc Natl Acad Sci USA* 107(11):4961–4966.
20. Wales DJ, Scheraga HA (1999) Global optimization of clusters, crystals, and biomolecules. *Science* 285(5432):1368–1372.
21. Carr JM, Trygubenko SA, Wales DJ (2005) Finding pathways between distant local minima. *J Chem Phys* 122(23):234903.
22. Wales DJ (2006) Energy landscapes: Calculating pathways and rates. *Int Rev Phys Chem* 25:237–282.
23. Henkelman G, Uberuaga B, Jonsson H (2000) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J Chem Phys* 113:9901.
24. Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR (2009) Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc Natl Acad Sci USA* 106(45):19011–19016.
25. Hagen M, Jeijer E, Mooij G, Frenkel D, Lekkerkerker H (1993) Does C_{60} have a liquid phase? *Nature* 365:425–426.
26. Doye JPK, Wales DJ (1996) The structure of $(C_{60})_n$ clusters. *Chem Phys Lett* 262:167–174.
27. Gazzillo D, Giacometti A, Fantoni R, Sollich P (2006) Multicomponent adhesive hard sphere models and short-ranged attractive interactions in colloidal or micellar solutions. *Phys Rev E Stat Nonlin Soft Matter Phys* 74(5 Pt 1):051407.
28. Dreyfus R, et al. (2009) Simple quantitative model for the reversible association of DNA coated colloids. *Phys Rev Lett* 102(4):048301.
29. Macfarlane R, et al. (2011) Nanoparticle superlattice engineering with DNA. *Science* 334(6053):204–208.
30. Rogers WB, Crocker JC (2011) Direct measurements of DNA-mediated colloidal interactions and their quantitative modeling. *Proc Natl Acad Sci USA* 108(38):15687–15692.
31. Baxter R (1968) Percus-Yevick equation for hard spheres with surface adhesion. *J Chem Phys* 49:2770–2774.
32. Stell G (1991) Sticky spheres and related systems. *J Stat Phys* 63(5–6):1203–1221.
33. Miller MA, Frenkel D (2003) Competition of percolation and phase separation in a fluid of adhesive hard spheres. *Phys Rev Lett* 90(13):135702.
34. Gazzillo D, Giacometti A (2004) Analytic solutions for Baxter’s model of sticky hard sphere fluids within closures different from the Percus-Yevick approximation. *J Chem Phys* 120(10):4742–4754.
35. Malins A, Williams SR, Eggers J, Tanaka H, Royall CP (2009) Geometric frustration in small colloidal clusters. *J Phys Condens Matter* 21(42):425103.
36. Wales DJ (2010) Energy landscapes of clusters bound by short-ranged potentials. *ChemPhysChem* 11(12):2491–2494.
37. Arkus N, Manoharan VN, Brenner MP (2009) Minimal energy clusters of hard spheres with short range attractions. *Phys Rev Lett* 103(11):118303.
38. Arkus N, Manoharan V, Brenner M (2011) Deriving finite sphere packings. *SIAM J Discrete Math* 25:1860–1901.
39. Allen MP, Tildesley DJ (1989) *Computer Simulation of Liquids* (Oxford Univ Press, Oxford).
40. Landau LD, Lifshitz EM (1978) *Statistical Physics, Course of Theoretical Physics* (Pergamon, Oxford), Vol 5.
41. Fatkullin I, Kovacic G, Vanden-Eijnden E (2010) Reduced dynamics of stochastically perturbed gradient flows. *Commun Math Sci* 8:439–461.
42. Calvo F, Doye JPK, Wales DJ (2012) Energy landscapes of colloidal clusters: Thermodynamics and rearrangement mechanisms. *Nanoscale* 4(4):1085–1100.
43. Asimow L, Roth B (1978) The rigidity of graphs. *Trans Am Math Soc* 245:279–289.
44. Gardiner C (2009) *Stochastic Methods: A Handbook for the Natural and Social Sciences* (Springer, Berlin), 4th Ed.
45. Ikeda N, Watanabe S (1981) *Stochastic Differential Equations and Diffusion Processes* (Elsevier, Amsterdam).
46. Metzner P, Schütte C, Vanden-Eijnden E (2006) Illustration of transition path theory on a collection of simple examples. *J Chem Phys* 125(8):084110.
47. Vanden-Eijnden E (2006) Transition path theory. *Lect Note Phys* 703:453–493.
48. Weinan E, Vanden-Eijnden E (2010) Transition-path theory and path-finding algorithms for the study of rare events. *Annu Rev Phys Chem* 61:391–420.
49. Schütte C, Noé F, Lu J, Sarich M, Vanden-Eijnden E (2011) Markov state models based on milestoning. *J Chem Phys* 134(20):204105.
50. Hormoz S, Brenner M (2011) Design principles for self-assembly with short-range interactions. *Proc Natl Acad Sci USA* 108:19885–19890.
51. Hoy R, Harwayne-Gidansky J, O’Hern C (2012) Structure of finite sphere packing via exact enumeration: Implications for colloidal crystal nucleation. *Phys Rev E Stat Nonlin Soft Matter Phys* 85(5):051403.
52. Sommese A, Wampler C (2005) *Numerical Solution of Polynomial Systems Arising in Engineering and Science* (World Scientific, Singapore).