Defining and Quantifying Frustration in the Energy Landscape: Applications to Atomic and Molecular Clusters, Biomolecules, Jammed and Glassy Systems

V. K. de Souza,[1] J. D. Stevenson,[1] S. P. Niblett,[1] J. D. Farrell,[1] and D. J. Wales[1, a]

University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW,

UK

(Dated: 15 February 2017)

The emergence of observable properties from the organisation of the underlying potential energy landscape is analysed, spanning a full range of complexity from self-organising to glassy and jammed systems. The examples include atomic and molecular clusters, a β-barrel protein, the GNNQQNY peptide dimer, and models of condensed matter that exhibit structural glass formation and jamming. We have considered measures based on several different properties, namely the Shannon entropy, an equilibrium thermodynamic measure that uses a sample of local minima, and indices that require additional information about the connections between local minima in the form of transition states. A frustration index is defined that correlates directly with key properties that distinguish relaxation behaviour within this diverse set. The index uses the ratio of the energy barrier to the energy difference with reference to the global minimum. The contributions for each local minimum are weighted by the equilibrium occupation probabilities. Hence we obtain fundamental insight into the connections and distinctions between systems that cover the continuum from efficient structure-seekers, to landscapes that exhibit broken ergodicity and rare event dynamics.

[a]dw34@cam.ac.uk

The potential energy landscape approach can provide both novel insight and computational methodology in molecular and condensed matter science.[1] Some of the most powerful tools in the theory of energy landscapes and the associated numerical methodology are based upon descriptions of structure, dynamics and thermodynamics formulated in terms of local minima on the landscape, and the transition states and pathways that connect them.[1] Disconnectivity graphs[2,3] (see Section I A) can provide an insightful representation of many-dimensional landscapes and are constructed using information about minima and transition states, preserving the potential (or free) energy barriers between minima. In particular, such graphs have enabled us to identify characteristic patterns of organisation in the underlying potential energy surface,[3] which result in efficient self-organisation at one end of the spectrum, and glassy behaviour at the other extreme.[4,5] These are systems with different levels of frustration. Self-organising systems have low levels of frustration and undergo efficient relaxation to the lowest-energy minimum whereas systems that are highly frustrated may never find this minimum-energy structure.

There is a qualitative difference in the landscapes for systems with different degrees of frustration. All the landscapes visualised for good structure-seekers, including 'magic number' clusters such as buckminsterfullerene, C_{60}, naturally occurring proteins, bulk representations of crystalline materials, and self-assembling mesoscopic shell structures, have a unique low-lying free energy minimum, with no competing morphologies separated by high barriers.[3] The appearance of the corresponding disconnectivity graphs has been likened to a multilevel palm tree, where local minima are separated from the global minimum by small barriers.[3] The resulting 'funnelled' landscape effectively guides relaxation to the global minimum. An example of a palm tree disconnectivity graph is shown for a Lennard-Jones cluster of 13 atoms (LJ_{13}) in the first panel of Figure 1. At the other extreme, the landscape for structural glass formers supports an exponentially large number of amorphous minima, separated by high barriers.[5] An example of a glassy tree can be seen in Fig. 3.

The folding of naturally occurring proteins has been associated with a 'principle of minimal frustration'.[6] These proteins fold easily and as they fold, their energy decreases more than would be expected by chance. There is a strong energetic bias toward the native state. This bias overcomes the unfavourable entropy change and also any kinetic traps contained within the structure of the landscape. Geometrical frustration, for example associated with the fact that locally favoured structures such as icosahedrons cannot tile space, is considered

2

to play a key role in the glass transition.[7] Between the limits of efficient self-organisation and glassy behaviour lies a continuous range of complexity. Of particular interest are systems characterised by two competing low-energy structures with a large interconversion barrier. Such 'double funnel' landscapes may exhibit a separation of relaxation time scales and heat capacity features associated with a low temperature transition between the alternative morphologies.[3] Such competition between structures has been termed 'frustration',[8,9] and a number of examples have been analysed in detail, as benchmarks for global optimisation, enhanced thermodynamic sampling designed to overcome broken ergodicity, and rare event dynamics.[3,10–15]

A variety of disconnectivity graphs for contrasting energy landscapes will be presented in the following sections. Our aim is to show how the global organisation of the landscape can be described in a quantitative manner, by considering various indices that reflect different measures of the complexity or frustration. One particular frustration index is found to be particularly suitable for this purpose.

I.   THE POTENTIAL ENERGY LANDSCAPE

The potential energy landscape (PEL) is a function of all the relevant atomic or molecular coordinates, which we represent by the components of a vector $\mathbf{X}$. Here we coarse-grain by considering stationary points, where the energy gradient vanishes, using second derivatives to classify minima, transition states, and higher-index saddles. The PEL is then characterised by three quantities: the potential energy $V(\mathbf{X})$; $\mathbf{g}(\mathbf{X})$, the first derivative of the potential energy; and the second derivative matrix $\mathbf{H}(\mathbf{X})$. The elements of the Hessian $\mathbf{H}(\mathbf{X})$ define the curvature of the landscape.

Stationary points have a physical meaning. Local minima have no negative Hessian eigenvalues, and a small displacement in any of the internal coordinates increases the energy. Even apparently simple systems can support large numbers of local minima on the PEL; the lowest is the global minimum. Transition states are defined geometrically in the present work, as stationary points with a single negative Hessian eigenvalue. Most transition states connect two minima, which are identified by calculating (approximate) steepest-descent paths leaving parallel and antiparallel to the eigenvector corresponding to the negative eigenvalue; such a minimum—transition state—minimum triplet can be considered as an elementary

3

85 rearrangement.

86 Knowledge of all the stationary points in the system would in principle provide insight
87 into a variety of interesting properties. However, larger systems possess a huge number of
88 stationary points, and so some method of sampling the landscape to obtain a representative
89 selection is required. There are a number of possible approaches to this problem and the
90 most appropriate method will generally depend on the properties of interest. Although the
91 landscape was explored in different ways for the different systems considered in this paper,
92 in all cases the same geometry optimisation tools were employed for locating minima and
93 transition states, and for characterising pathways. The OPTIM program was used for all
94 these calculations.

95 Local minima were found by minimising the potential energy using the limited memory
96 Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm,[16] which has proved to be particu-
97 larly efficient.[17] Transition states were located using the doubly-nudged[18] elastic band[19,20]
98 (DNEB) method. Local maxima in the chain of states corresponding to the DNEB inter-
99 polation were taken as candidate transition states (TS), which were refined using hybrid
100 eigenvector-following.[21] To identify the two minima connected by each TS, a small step was
101 taken both parallel and antiparallel to the eigenvector corresponding to the unique negative
102 Hessian eigenvalue. LBFGS minimisation was then applied to each set of displaced coordi-
103 nates to find the two connected minima. If a complete pathway of elementary arrangements
104 between the two original minima was not found, a modified Dijkstra algorithm[22] was used
105 to choose a pair of intermediate minima to connect next, with the aim of completing the
106 pathway using the smallest number of connection attempts.

107 For most of the systems studied in the present work, the PATHSAMPLE program was
108 used as a driver for OPTIM to expand the stationary point databases. The discrete path
109 sampling (DPS) approach[13] implemented in PATHSAMPLE is a coarse-grained analogue
110 of the transition path sampling method.[23] PATHSAMPLE can systematically generate sets
111 of discrete paths from an initial connected path between two end points, creating a kinetic
112 transition network. These discrete paths are connected sequences of minima and intervening
113 transition states.[1,13] The PELE[24] package, a python library for finding global minima and
114 connecting minima, was also used for some of the systems. Further details are provided
115 below.

A. Disconnectivity Graphs

To construct a disconnectivity graph, we group minima that are mutually accessible at a given total energy, E. These minima are connected by single- or multi-step pathways where the energy never exceeds E. Minima are placed in a different group if the path between them contains a transition state that lies above the threshold energy, E. These groups can then be connected at a higher energy. Each group (or superbasin) is represented by a point, or node, on the horizontal axis, and the vertical axis corresponds to increasing energy. Nodes are joined at higher energy if they belong to the same superbasin and nodes continue down to the level of each single local minimum. The horizontal axis is usually arbitrary and the order and spacings between nodes can be chosen for clarity. The superbasin analysis is performed at a discrete series of total energies, E. There remains an adjustable parameter in the spacing between energy levels. If the spacing is too large, little topographical information is left, but if it is too small, any coarse-grained structure may be hidden. We emphasise that disconnectivity graphs are simply employed as a visualisation tool to provide insight into the organisation of the potential or free energy landscape. When thermodynamic or kinetic properties are required, we utilise additional information. Connected databases of minima and transition states define kinetic transition networks, which can be constructed systematically using the discrete path sampling (DPS) framework.[13] Information from the underlying kinetic transition network, specifically the local densities of states, allows us to estimate occupation probabilities, free energies and minimum-to-minimum rate constants.

II. SYSTEMS

A. Atomic Clusters

Many small atomic clusters exhibit efficient relaxation to the global minimum. However, clusters of particular sizes can exhibit two (or more) competing morphologies, distinct struc- tures with similar energy to the global minimum, leading to a frustrated landscape.[8,9] The interatomic potential considered is the pairwise additive, isotropic Lennard-Jones (LJ) form:

$$V = 4\epsilon \sum_{i<j} \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right], \tag{1}$$

143 where ϵ and $2^{1/6}\sigma$ are the pair equilibrium well depth and separation, respectively, and we
144 employ reduced units of energy and distance defined by ϵ and σ.

145   Disconnectivity graphs for five different cluster sizes are shown in Figure 1. For clusters
146 with 13 and 55 atoms, denoted LJ_{13} and LJ_{55}, the global minimum is a complete Mackay
147 icosahedron, and the landscapes have a single funnel 'palm tree' form, with no competing
148 alternative low energy morphologies.[3] In contrast, for each of LJ_{31}, LJ_{38} and LJ_{75} there is a
149 double funnel structure. For LJ_{31}, the competing funnel is characterised by an alternative
150 surface structure. For LJ_{38} and LJ_{75}, structures in the competing funnels have fundamen-
151 tally different packing schemes, separated from higher energy incomplete icosahedral minima
152 by relatively large barriers.[3,10] The aim of this paper is to find a useful measure of this frus-
153 tration, which should identify the LJ_{31}, LJ_{38} and LJ_{75} double-funnel systems as significantly
154 more frustrated than LJ_{13} and LJ_{55}. When searching for the global minimum, the mean first
155 encounter time is significantly longer for systems with a double funnel structure.[1,25] Hence,
156 these double funnel landscapes provide important benchmarks for global optimisation, en-
157 hanced sampling schemes, and rare event methods.[11–15]

158   Stationary point databases were all obtained by initially connecting low-lying minima
159 encountered in basin-hopping global optimisation runs.[26,27] Further refinement was based on
160 locating kinetically relevant discrete paths between competing morphologies for the double-
161 funnel systems, or increasing the number of connections per minimum for LJ_{13} and LJ_{55}. In
162 each case the database was also enlarged using a scheme to remove artificial frustration.[28]
163 Aside from LJ_{13}, the samples obtained are necessarily incomplete. These databases were
164 therefore extended until the appearance of the low-energy region of the landscape appeared
165 to have converged, as judged by inspection of the disconnectivity graphs. The convergence
166 of the various measures of frustration defined in Section III for the lower temperature range
167 of interest was also monitored.

B. $(H_2O)_{20}$

169   The energy landscape of the 'nanodroplet' cluster $(H_2O)_{20}$ displays signatures of hier-
170 archical organisation.[3] This complexity has been associated with the additional rotational
171 degrees of freedom for the water molecule, which lead to a wide range of energies and barrier
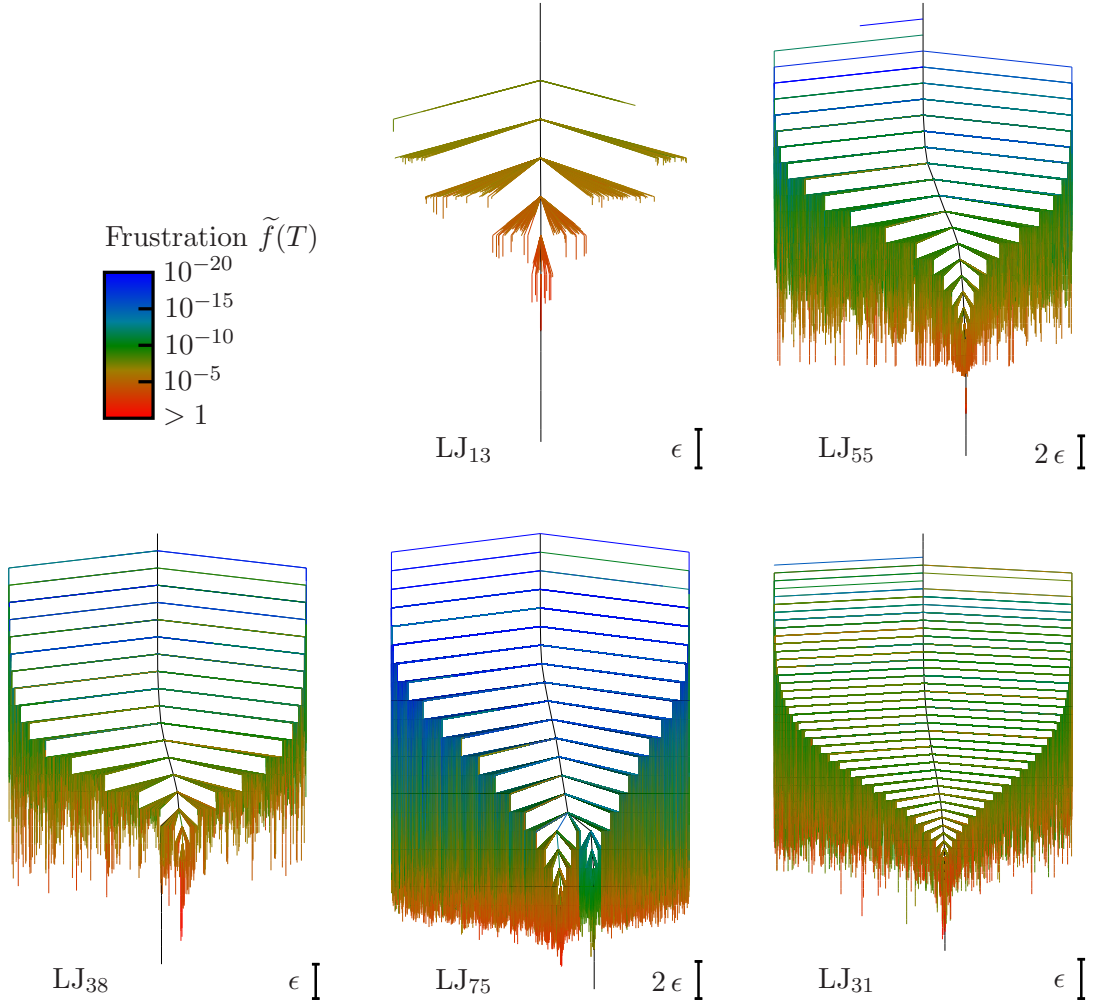172 heights separating structures with similar arrangements of the oxygen atoms, but alternative

FIG. 1. Disconnectivity graphs for five Lennard-Jones clusters. The graphs are arranged in order of increasing frustration from left to right and top to bottom: LJ_{13}, LJ_{55}, LJ_{38}, LJ_{75} and LJ_{31}. The global minimum is shown in black, and branches leading to every other minimum are coloured according to their contribution to the $\widetilde{f}(T)$ frustration index at the melting temperature of the system. The scalebar for each graph defines the energy spacing.

hydrogen-bonding patterns.[3] Understanding how these effects result in qualitative differences from the atomic clusters considered above is of great interest in view of the central role of water in solvation of biomolecular systems.

For the TIP4P rigid molecule pair potential[29] the global potential energy minimum can be described in terms of three pentagonal prisms sharing approximately square faces.[30,31] There are two other morphologies with competing low energy minima based on face-sharing

179 cuboids ('box-kite' structure[32]) and three stacked pentagonal prisms sharing pentagonal
180 faces.[31] Minima based on dodecahedral structures generally lie rather higher in energy.[31]
181 The landscape for this system was explored as follows. Four low-energy structures corre-
182 sponding to the different morphologies were connected and then the pathway between the
183 two lowest structures was further refined using a short-cutting procedure.[28,33] The resulting
184 disconnectivity graph is shown in Fig. 2.
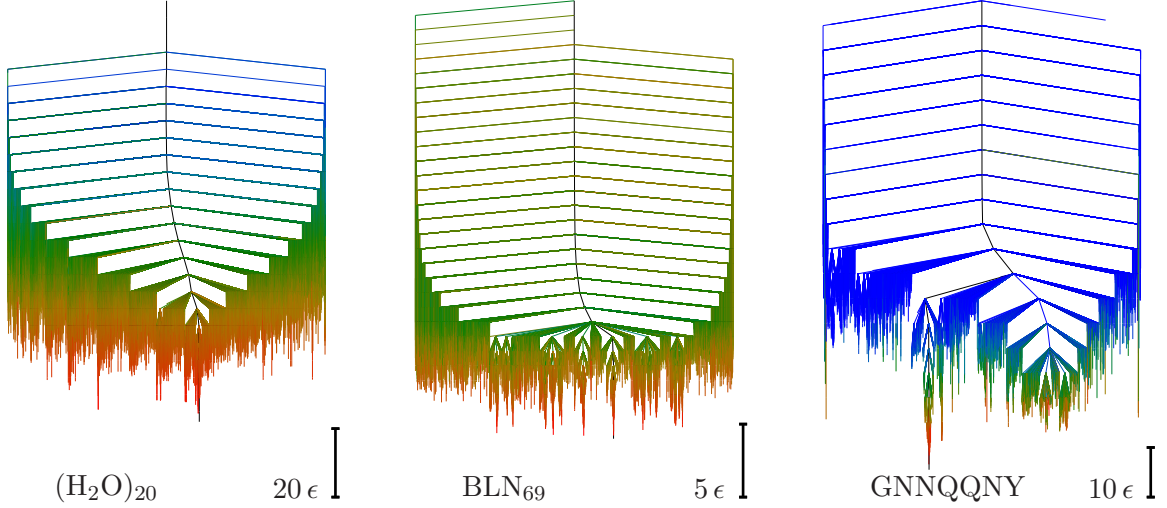
FIG. 2. Disconnectivity graphs for $(H_2O)_{20}$, the 69-residue BLN model protein, and the heptapep-
tide GNNQQNY. The global minimum for each system is shown in black and branches leading to
every other minimum are coloured according to their contribution to the $\widetilde{f}(T)$ frustration index
at the melting temperature of the system. The colour scale is identical to that used in Figure 1,
where minima corresponding to red branches make the largest contribution to frustration. The
scalebar for each graph defines the energy spacing.

185 ## C.   A Model β-Barrel Protein

186   Here we considered a coarse-grained BLN model for a 69-residue protein,[34] where each
187 amino acid is represented by a hydrophobic (B), hydrophilic (L) or neutral (N) bead. Stiff
188 harmonic springs were employed to restrain the bond lengths.[35] The global minimum of
189 the sequence $B_9N_3(LB)_4N_3B_9N_3(LB)_4$ (BLN$_{69}$) is a six-stranded β-barrel, with three chains
190 forming a hydrophobic core. However, we have previously shown that there are a num-

8

ber of low-lying minima with alternative arrangements of the β-strands, separated by high barriers.[36]

Figure 2 shows a disconnectivity graph for BLN_{69}. The database was again produced by initially connecting low-lying minima obtained using basin-hopping global optimisation,[26,27] and then refining to locate kinetically relevant paths between alternative morphologies and remove artificial frustration caused by undersampling.

D. The Amyloidogenic GNNQQNY Peptide Dimer

The polar amyloidogenic heptapeptide GNNQQNY is a key component of the N-terminal prion-determining domain of the yeast protein Sup35. This peptide is noteworthy for its individual stability,[37] and a crystal structure has been characterised, where the monomers align to give in-register parallel β-sheets.[38] Here we employ a DPS database created for the GNNQQNY dimer, which we previously found to support four competing conformations, one that is intermediate between compact and extended structures, and three β-sheets corresponding to in-register parallel, off-register parallel, and antiparallel arrangements.[28] These results were obtained with the CHARMM19[39] potential and EEF1 implicit solvent.[40] The corresponding potential energy landscape might be expected to exhibit an intermediate level of frustration, and this intuition is consistent with the new results reported below. The time scale for interconversion between the four low-lying structures was estimated as hours or longer at 298 K.[28] Replica exchange molecular dynamics[41] (REMD) was used for an initial exploration of the conformational space. The resulting database of minima and transition states was then grown using DPS with the focus on the paths making the largest contributions to the steady state rate constants between the alternative morphologies. Shortcutting procedures[28,33] were followed by the removal of artificial traps.[28] A disconnectivity graph for this system is shown in Fig. 2.

E. Binary Structural Glass Formers

Binary Lennard-Jones (BLJ) potentials are extensively used in studies of structural glass phenomenology,[42–45] since with suitable parameterisations they do not crystallise on time scales usually accessible to molecular dynamics. Here we consider BLJ systems modelled

9

using periodic boundary conditions containing two types of atom, A and B, in the ratio 80:20, which was originally proposed to model the metallic glass $Ni_{0.8}P_{0.2}$.[46] Although this model possesses mixed and phase separated crystalline structures,[47,48] these can easily be excluded from the structural databases employed for landscape analysis. The parameterisation employed was $\sigma_{AA} = 1$, $\sigma_{AB} = 0.8$, $\sigma_{BB} = 0.88$, $\epsilon_{AA} = 1$, $\epsilon_{AB} = 1.5$, and $\epsilon_{BB} = 0.5$.[42] The resulting disconnectivity graphs[5] exhibit many low-lying minima corresponding to amorphous structures, separated by high barriers of order $30k_BT_g$, where T_g corresponds to a glass transition temperature. Grouping local minima according to whether they can be interconverted without a cage-breaking rearrangement, which would be necessary for diffusion, has revealed a higher order structure in the energy landscape,[5] which may yield a useful way to define 'metabasins'.[49,50]

To avoid finite-size effects, the system must not be too small, and 60 atoms was found to be a good compromise in previous work.[43] Our results are for a 60-atom BLJ mixture of 48 type A and 12 type B particles (BLJ_{60}) and a 256-atom mixture of 204 A and 52 B particles (BLJ_{256}). Both systems have a number density of 1.3 in reduced units of σ_{AA}^{-3} (used throughout) and were simulated with periodic boundary conditions. Full details are given in Ref. 51 and disconnectivity graphs for BLJ_{60} and BLJ_{256} are shown in Figure 3. The BLJ_{60} database contains over 11000 minima, while the database for BLJ_{256} has 2500. The PELs for both systems show typical glassy organisation, with many minima of similar energies separated by high barriers, and a hierarchical structure. For both BLJ systems, initial samples of minima were obtained from locally ergodic canonical MD trajectories.[51] Local minimisation was applied for each configuration of the trajectory, producing a sequential series of minima. Adjacent minima were then connected, giving a single kinetically relevant discrete path.[5,52] We expect to find lower frustration for such a connected path of minima and transition states, compared to a more extensive search for a particular region of the energy landscape, as discussed below.
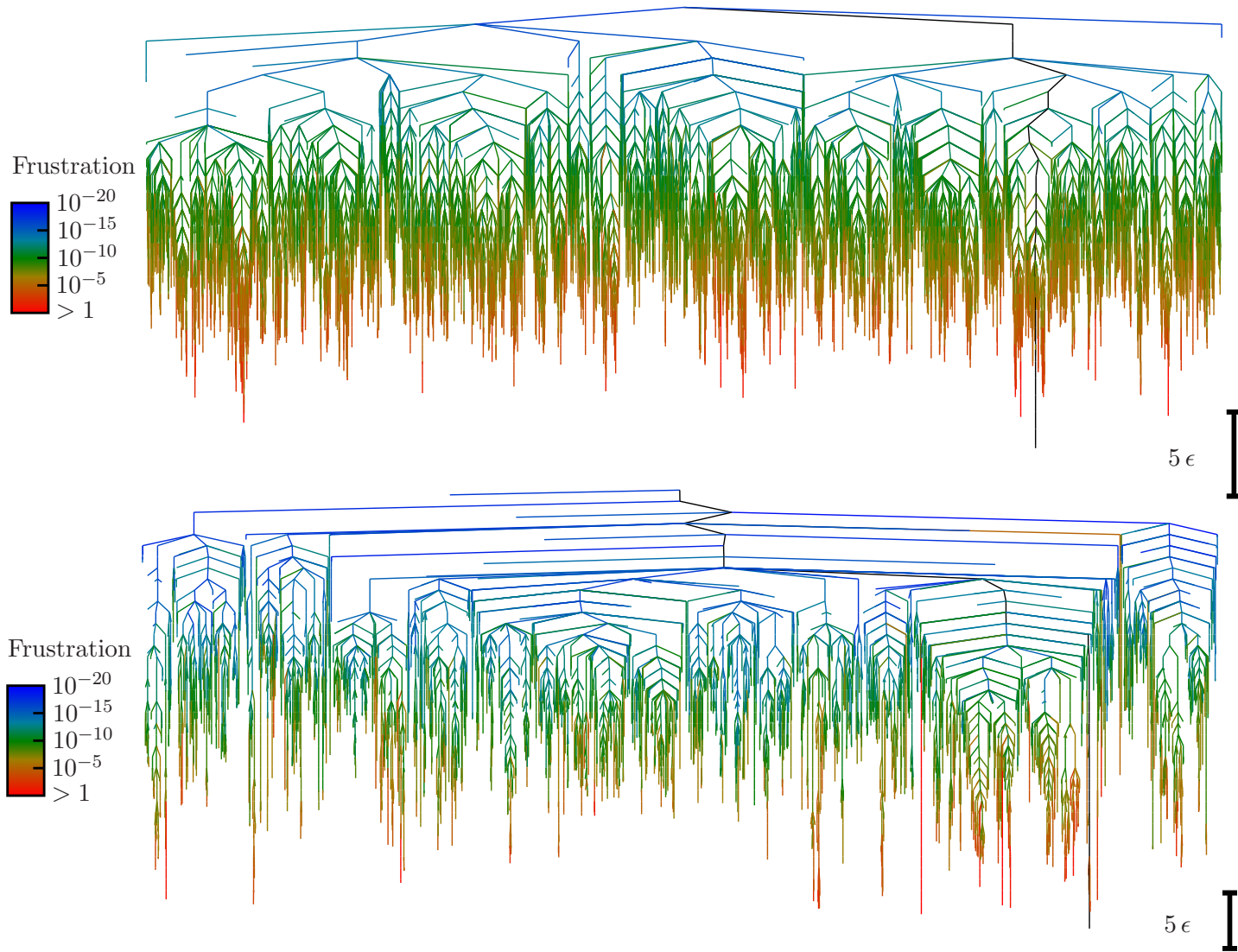
FIG. 3. Disconnectivity graphs for BLJ_{60} (top) and BLJ_{256} (bottom). The colour scale shows the contributions to the frustration index $\widetilde{f}(T)$ at the melting temperature, and the energy scale bar defines the spacing.

F. Soft Spheres: Landscapes for Jamming

Jamming corresponds to a transition to rigidity in disordered matter. In the jammed state, materials respond essentially elastically to small applied shear stresses. However, such systems can easily be unjammed and made to flow by tuning various control parameters. The jamming transition can be induced by varying thermodynamic variables, such as temperature or density, and also by mechanical variables, such as applied stress.[53]

Packings of soft repulsive spheres at zero temperature exhibit a sharp jamming transition in the thermodynamic limit. The transition is induced by applied pressure or enforced packing fraction, causing deformation of the particles. When deformations vanish, the system loses rigidity. At low packing fraction, no particles interact, a state described as an unjammed 'mechanical vacuum'. As the packing fraction is increased, there is a transition to a jammed, rigid structure. Such frictionless spheres with finite-range repulsions have been studied extensively.[53–55]

The two systems considered here consist of $N = 60$ and $N = 256$ spheres of equal mass, M, and variable radii, R_i, with periodic boundary conditions. The radii are distributed evenly within the range 0.416 to 0.583. When they overlap, the spheres interact via a hertzian soft sphere potential. Denoting the centre-to-centre distance as r_{ij}, a dimensionless overlap parameter, δ_{ij}, can be defined as

$$\delta_{ij} = 1 - \frac{r_{ij}}{R_i + R_j}. \tag{2}$$

The interaction potential is zero unless $\delta_{ij} > 0$ and then has the form

$$V_{ij} = \epsilon_{ij}\delta_{ij}^{2.5}. \tag{3}$$

For the soft sphere systems, a particular region of the potential energy landscape was explored in detail using PELE. Basin-hopping steps were taken from an initial starting minimum to find further minima, and then pathways were obtained using DNEB. The resulting database was refined with the untrapping method designed to remove artificial frustration, as for some of the other models discussed above.[28] The untrapping scheme was designed to locate missing lower energy pathways in the low energy part of the landscape. However, for soft spheres, it also results in significant further exploration, finding many more minima with similar energies, which contribute to the frustration. As this sampling method

may overestimate frustration, for comparison we have performed an identical analysis for 274

BLJ$_{256}$. Disconnectivity graphs are shown in Fig. 4 for the 60 and 256 sphere systems and 275
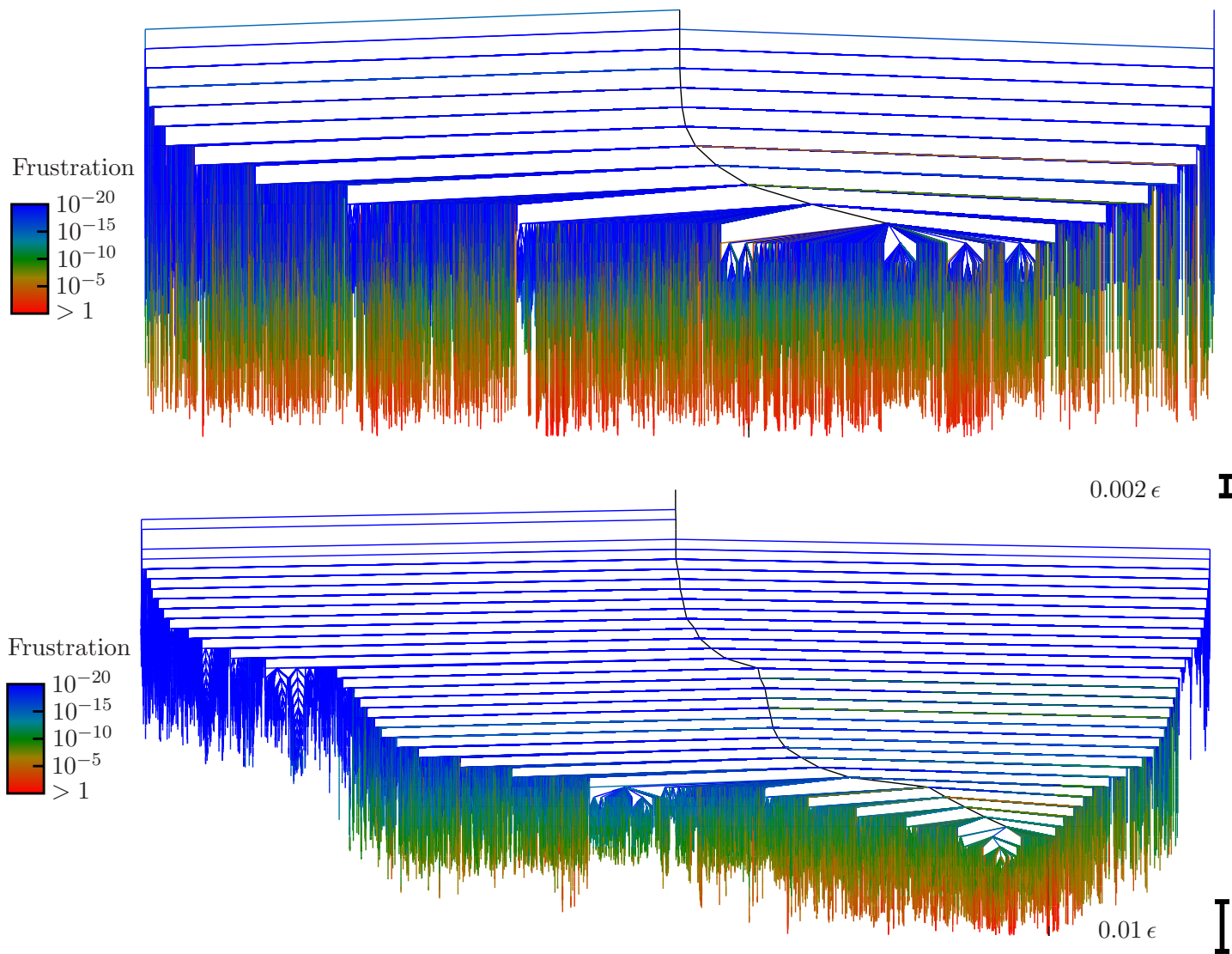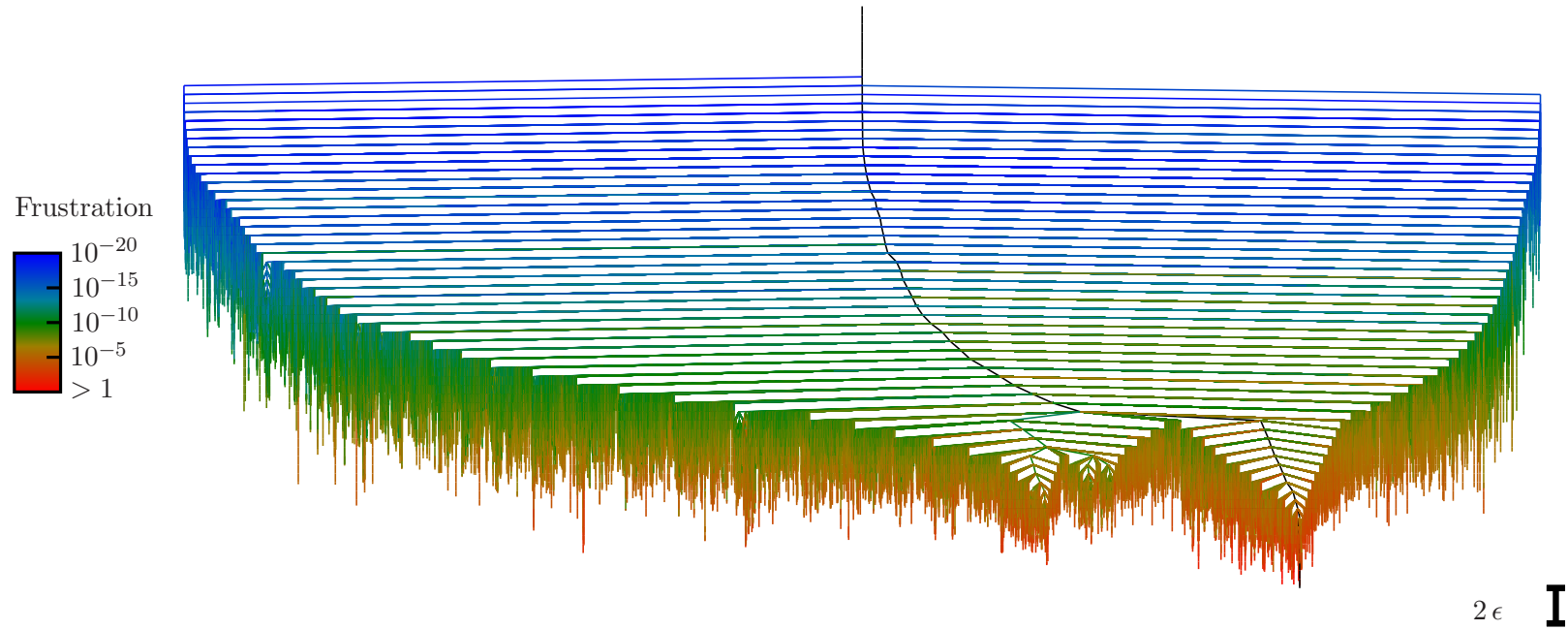
in Fig. 5 for BLJ$_{256}$. 276

FIG. 4. Disconnectivity graphs for 60 (top) and 256 (bottom) soft spheres. The colour scale shows the contributions to the frustration index $\widetilde{f}(T)$ at the melting temperature, and the scalebar defines the energy spacing.

FIG. 5. Disconnectivity graph corresponding to extensive exploration of a local region for BLJ_{256}. The colour scale shows the contributions to the frustration index $\widetilde{f}(T)$ at the melting temperature, and the scalebar defines the energy spacing.

III. QUANTIFYING FRUSTRATION IN THE LANDSCAPE

We have considered various different indices to analyse the degree of frustration in different systems. Firstly, the Shannon entropy[56,57], $s(T)$, depends upon equilibrium thermodynamic properties, and therefore only requires information about local minima.

$$s(T) = -\sum_\alpha p_\alpha^{\text{eq}}(T) \ln p_\alpha^{\text{eq}}(T), \tag{4}$$

where T is the temperature and $p_\alpha^{\text{eq}}(T)$ is the equilibrium occupation probability of minimum α, calculated using harmonic vibrational densities of states.[58] If information about transition states is not available, this index can provide a useful measure of frustration, as shown in Figure 6. To compare results for very different systems we have used a dimensionless temperature, obtained using an estimate for the melting temperature, T_m. This estimate comes from the temperature at the corresponding heat capacity maximum, obtained consistently within the harmonic superposition approximation.[58]

The three indices defined below require databases where the connectivity of the local minima is defined by transition states.

$$N(T) = \sum_{\alpha=1}^{m} \sum_{\gamma < \alpha} \left[p_\alpha^{\text{eq}}(T) + p_\gamma^{\text{eq}}(T) \right] n_{\alpha\gamma}(T),$$

$$K(T) = \sum_{\alpha=1}^{m} \sum_{\gamma < \alpha} \left[p_\alpha^{\text{eq}}(T) k_{\gamma\alpha}(T) + p_\gamma^{\text{eq}}(T) k_{\alpha\gamma}(T) \right]$$

$$f(T) = \sum_{\alpha \neq \text{gmin}} p_\alpha^{\text{eq}}(T) \left(\frac{V_\alpha^\dagger - V_{\text{gmin}}}{V_\alpha - V_{\text{gmin}}} \right) \tag{5}$$

$N(T)$ uses $n_{\alpha\gamma}$, the number of steps (transition states) in the fastest path between minima γ and α, and $K(T)$ uses $k_{\alpha\gamma}$, the rate constant for transitions between minima γ and α. $n_{\alpha\gamma}$ and $k_{\alpha\gamma}$ were both obtained from Dijkstra shortest path analysis[59], with edge weights of $-\ln P_{\alpha\gamma}$, where $P_{\alpha\gamma}$ is the branching probability of stepping to minimum α from among the direct connections of minimum γ. The branching probabilities and minimum-to-minimum rate constants $k_{\alpha\gamma}$ were estimated from transition state theory,[60–62] using harmonic vibrational densities of states. In contrast, $f(T)$ uses information about the potential energy of the minima, V_α, and transition states. V_{gmin} signifies the potential energy of the lowest minimum in the database and $V_\alpha^\dagger$ is the potential energy of the highest transition state on the lowest energy path between minimum α and the lowest minimum. The lowest minimum

16

is the global minimum for most of the databases considered, except for the condensed matter 304
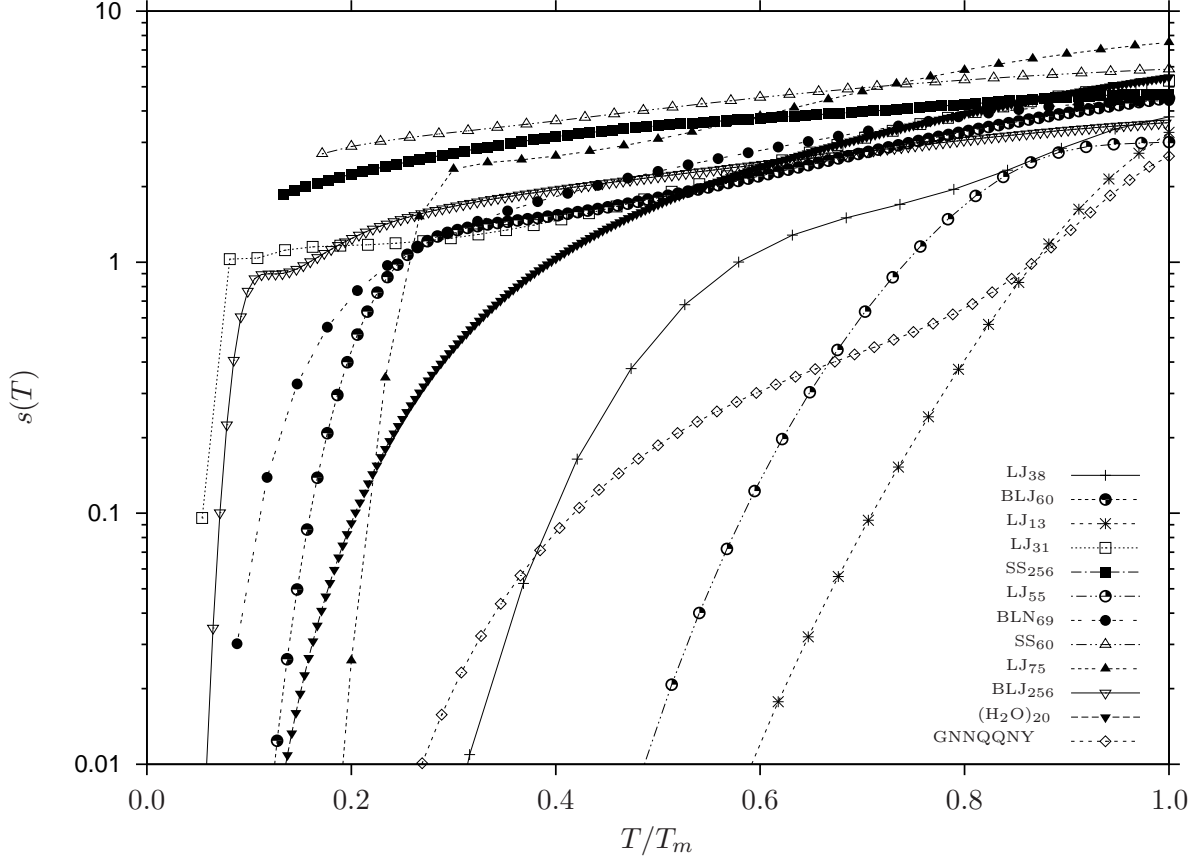systems, where the crystalline region of configuration space is intentionally excluded. 305

FIG. 6. The shannon entropy, $s(T)$, for all the models considered.

The most insightful quantity was found to be $f(T)$, which employs barrier thresholds $V_\alpha^\dagger$ 306
in the same way as a scheme that we introduced in earlier work to help refine databases to 307
remove artificial frustration.[28] This approach was, in turn, derived from an analogous index 308
based on free energies that we used earlier,[63] which extends measures based on stability[64] 309
and energy gaps[65,66] by including explicit barrier information. A further advantage of $f(T)$ 310
over the other measures that include connectivity, $N(T)$ and $K(T)$, is that it does not require 311
additional characterisation of discrete paths or rate information and hence, it is quick and 312
easy to calculate. 313

We calculated $V_\alpha^\dagger$ in tandem with the superbasin analysis[2] that yields the disconnectivity 314
graphs, identifying the energy threshold below which the lowest minimum is no longer ac- 315
cessible from minimum α. Results for $f(T)$, which we will refer to as the *frustration index*, 316
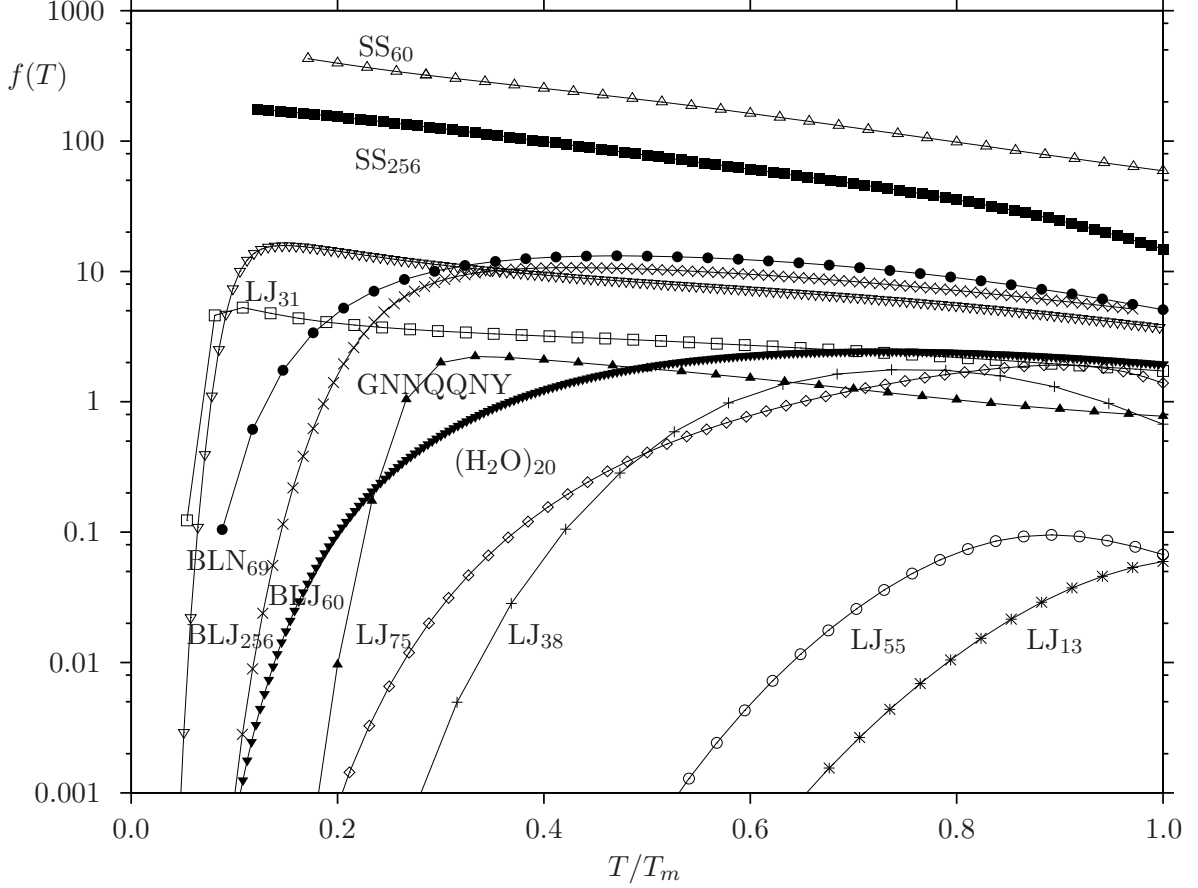are shown in Figure 7. 317

17

FIG. 7. The frustration index, $f(T)$, is shown for all the models considered.

Analogous indices $\widetilde{s}(T)$, $\widetilde{N}(T)$, $\widetilde{K}(T)$, and $\widetilde{f}(T)$ were calculated from the probabilities $\widetilde{p}_\alpha^{\text{eq}} = p_\alpha^{\text{eq}}/(1 - p_{\text{gmin}}^{\text{eq}})$. These quantities reflect the renormalised relative populations of the minima when the temperature dependence of the global minimum is removed. As the temperature decreases, the equilibrium occupation probability of the global minimum can become very large. As the global minimum does not itself contribute to frustration, this high occupation probability causes the frustration measures to decrease rapidly at low temperature. This decrease is an erroneous effect as relaxation to the global minimum becomes more difficult and hence frustration increases at low temperature. The renormalisation is shown for LJ$_{38}$ in Figure 8, where $f(T)$ and $\widetilde{f}(T)$ are compared. Full results for $\widetilde{f}(T)$ are shown in Figure 9.
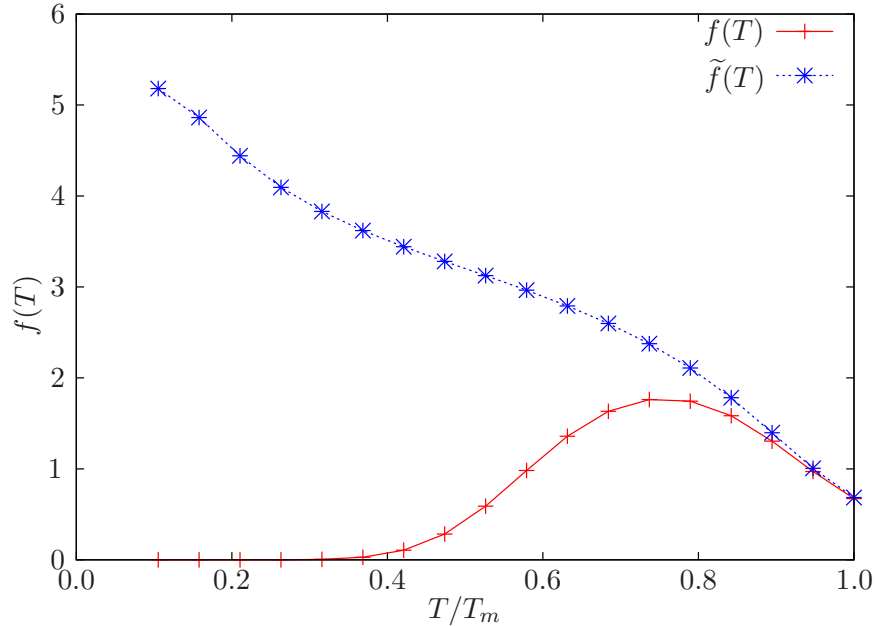
FIG. 8. $f(T)$ and $\widetilde{f}(T)$ for LJ$_{38}$.

The classification of systems is probably clearest for $\widetilde{f}(T)$, which provides a relatively stable index over quite a wide range of temperature, thanks to removal of the temperature dependence of the lowest minimum. We see that for most of the range $0 < T/T_m < 1$ structure-seekers have $\widetilde{f}(T) < 1$, frustrated landscapes including structural glasses have $1 > \widetilde{f}(T) > 50$, and jammed systems have $\widetilde{f}(T) > 100$. We would not wish to overinterpret smaller differences in $\widetilde{f}(T)$; it is the order of magnitude differences that are really significant here.

A.   Comparison of Sampling Protocol

For most of the models compared here, the underlying kinetic transition networks correspond to samples of stationary points designed to describe the dynamics associated with specific pathways or regions of the landscape. The degree of frustration could therefore exhibit some degree of sample dependence, and we have therefore provided details of the landscape exploration methods for each system in Section II. However, if all the landscapes were sampled in the same way, we expect the key trends to be unaffected. A representative sample should provide us with occupation probabilities and barrier heights that provide a converged value for the chosen measure of frustration, $\widetilde{f}(T)$. The most frustrated systems we have considered here are the soft sphere packings, which were explored through a more
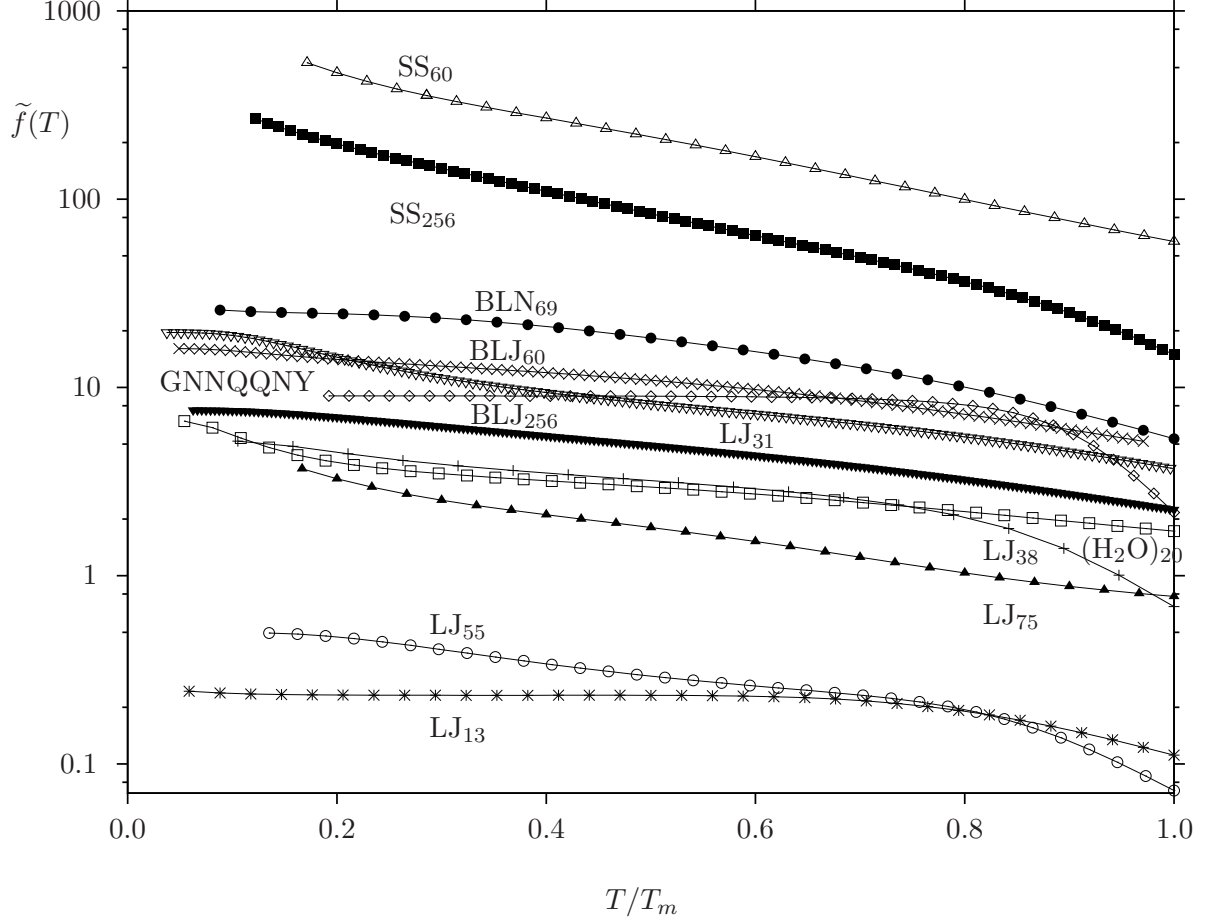
19

FIG. 9. The frustration index calculated from renormalised probabilities, $\widetilde{f}(T)$, for each of the models considered.

extensive local search. The same local exploration was therefore repeated for the binary Lennard-Jones system of 256 particles and continued until we obtained the same number of minima as for the soft sphere example. Fig. 10 shows that the results for this scheme with around 40000 minima are very similar to those obtained for a single kinetically relevant discrete path of around 3600 minima, based upon an initial molecular dynamics trajectory. The frustration index is still significantly lower than for the soft sphere systems. In fact, the more extensive local searches may produce lower frustration measures as they do not encounter many of the alternative low-energy minima separated by high barriers that are present in the landscape.
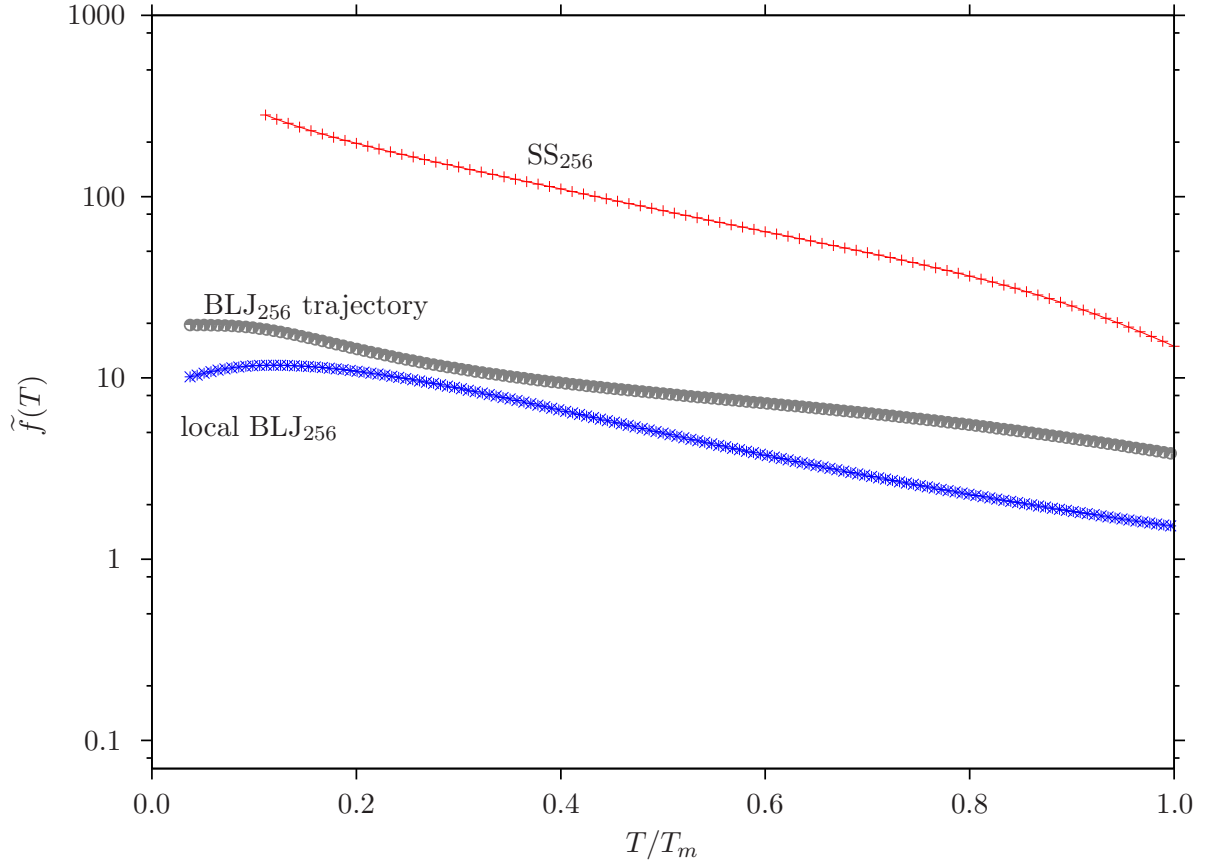
20

FIG. 10. $\widetilde{f}(T)$ for a 256-atom binary Lennard-Jones mixture, where the energy landscape has been explored by two different schemes. In the first approach, the sample is based on a pathway through the landscape corresponding to a locally-ergodic molecular dynamics trajectory. In contrast, the second scheme employs a more extensive local exploration, as for the 256-atom soft sphere system.

IV.   DISCUSSION AND CONCLUSIONS

We have explored measures to quantify the global organisation of the potential energy landscape. A frustration index, $\widetilde{f}(T)$, including renormalised approximate equilibrium occupation probabilities and barrier information is found to be particularly useful. If transition state information is not available, the Shannon entropy, based on equilibrium occupation probabilities, can also provide a measure of frustration, albeit without allowing for dynamical effects.

We studied a number of different systems and found soft sphere packings to be the most frustrated, followed by binary Lennard-Jones models of supercooled liquids and a 69-residue

model β-barrel protein. In all these very frustrated systems, there are a number of low-energy competing structures close in energy to the global minimum. Lower frustration is found for systems with a well-defined global minimum. LJ_{31} is the most frustrated Lennard-Jones cluster, while the amyloidogenic GNNQQNY heptapeptide and the $(H_2O)_{20}$ cluster exhibit comparable values for $\widetilde{f}(T)$. The least frustrated systems we compared are the LJ_{13} and LJ_{55} clusters, where the landscapes have a single funnel with no competing low-energy structures.

The frustration index provides a quantity to characterise aspects of the potential energy landscape that have previously been described more qualitatively. It is easy to calculate, and we have shown that it reliably distinguishes different degrees of frustration for a variety of systems. The key features are also relatively robust in terms of how the underlying landscape has been explored. These results should provide a useful tool for quantifying frustration across very different systems of interest, potentially providing new insight into our understanding of emergent properties.

ACKNOWLEDGMENTS

REFERENCES

[1] D. J. Wales, *Energy Landscapes* (Cambridge University Press, Cambridge, 2003).

[2] O. M. Becker and M. Karplus, J. Chem. Phys. **106**, 1495 (1997).

[3] D. J. Wales, M. A. Miller, and T. R. Walsh, Nature **394**, 758 (1998).

[4] D. J. Wales, Phil. Trans. Roy. Soc. A **363**, 357 (2005).

[5] V. K. de Souza and D. J. Wales, J. Chem. Phys. **129**, 164507 (2008).

[6] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **84**, 7524 (1987).

[7] G. Tarjus, S. A. Kivelson, Z. Nussinov, and P. Viot, J. Phys. Cond. Matt. **17**, 114 (2005).

[8] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Proteins: Struct., Func. and Gen. **21**, 167 (1995).

[9] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).

[10] J. P. K. Doye, M. A. Miller, and D. J. Wales, J. Chem. Phys. **110**, 6896 (1999).

[11] J. P. Neirotti, F. Calvo, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **112**, 10340 (2000).

[12] F. Calvo, J. P. Neirotti, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **112**, 10350 (2000).

[13] D. J. Wales, Mol. Phys. **100**, 3285 (2002).

[14] V. A. Mandelshtam, P. A. Frantsuzov, and F. Calvo, J. Phys. Chem. A **110**, 5326 (2006).

[15] F. Calvo, Phys. Rev. E **82**, 046703 (2010).

[16] J. Nocedal, Mathematics of Computation **35**, 773 (1980).

[17] S. T. Chill, J. Stevenson, V. Ruehle, C. Shang, P. Xiao, J. D. Farrell, D. J. Wales, and G. Henkelman, J. Chem. Theor. Comput. **10**, 5476 (2014).

[18] S. A. Trygubenko and D. J. Wales, J. Chem. Phys. **120**, 2082 (2004).

[19] G. Henkelman and H. Jónsson, J. Chem. Phys. **113**, 9978 (2000).

[20] G. Henkelman, B. P. Uberuaga, and H. Jónsson, J. Chem. Phys. **113**, 9901 (2000).

[21] L. J. Munro and D. J. Wales, Phys. Rev. B **59**, 3969 (1999).

[22] J. M. Carr, S. A. Trygubenko, and D. J. Wales, J. Chem. Phys. **122**, 234903 (2005).

[23] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, J. Chem. Phys. **108**, 1964 (1998).

[24] "Pele: Python energy landscape explorer, https://github.com/pele-python/pele,".

[25] D. J. Wales and H. A. Scheraga, Science **285**, 1368 (1999).

[26] Z. Li and H. A. Scheraga, Proc. Natl. Acad. Sci. USA **84**, 6611 (1987).

[27] D. J. Wales and J. P. K. Doye, J. Phys. Chem. A **101**, 5111 (1997).

[28] B. Strodel, C. S. Whittleston, and D. J. Wales, J. Amer. Chem. Soc. **129**, 16005 (2007).

[29] W. L. Jorgensen, J. Chem. Phys. **77**, 4156 (1982).

[30] D. J. Wales and I. Ohmine, J. Chem. Phys. **98**, 7257 (1993).

[31] D. J. Wales and M. P. Hodges, Chem. Phys. Lett. **286**, 65 (1998).

[32] C. J. Tsai and K. D. Jordan, Chem. Phys. Lett. **213**, 181 (1993).

[33] J. M. Carr and D. J. Wales, J. Chem. Phys. **123**, 234901 (2005).

[34] J. D. Honeycutt and D. Thirumalai, Proc. Nat. Acad. Sci. **87**, 3526 (1990).

[35] R. S. Berry, N. Elmaci, J. P. Rose, and B. Vekhter, Proc. Natl Acad. Sci. USA **94**, 9520

23

423 (1997).

36 M. T. Oakley, D. J. Wales, and R. L. Johnston, J. Phys. Chem. B **115**, 11525 (2011).

37 M. Balbirnie, R. Grothe, and D. S. Eisenberg, Proceedings of the National Academy of Sciences **98**, 2375 (2001).

38 R. Nelson, M. R. Sawaya, M. Balbirnie, A. O. Madsen, C. Riekel, R. Grothe, and D. Eisenberg, Nature **435**, 773 (2005).

39 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comp. Chem. **4**, 187 (1983).

40 T. Lazaridis and M. Karplus, Proteins: Struct., Funct., Genet. **35**, 133 (1999).

41 Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).

42 W. Kob and H. C. Andersen, Phys. Rev. Lett. **73**, 1376 (1994).

43 S. Büchner and A. Heuer, Phys. Rev. Lett. **84**, 2168 (2000).

44 F. Sciortino, W. Kob, and P. Tartaglia, Phys. Rev. Lett. **83**, 3214 (1999).

45 S. Sastry, Nature **409**, 164 (2001).

46 T. A. Weber and Stillinger, Phys. Rev. B **31**, 1954 (1985).

47 T. F. Middleton, J. Hernández-Rojas, P. N. Mortenson, and D. J. Wales, Phys. Rev. B **64**, 184201 (2001).

48 J. R. Fernández and P. Harrowell, Phys. Rev. E **67**, 011403 (2003).

49 F. H. Stillinger, Science **267**, 1935 (1995).

50 B. Doliwa and A. Heuer, J. Phys. Cond. Matt. **15**, S849 (2003).

51 V. K. de Souza and D. J. Wales, Phys. Rev. B **74**, 134202 (2006).

52 V. K. de Souza and D. J. Wales, J. Chem. Phys. **130**, 194508 (2009).

53 A. Liu and S. R. Nagel, Annu. Rev. Condens. Matter Phys **1**, 347 (2010).

54 M. van Hecke, J. Phys.: Condens. Matter **22**, 033101 (2009).

55 C. S. O'Hern, L. E. Silbert, A. J. Liu, and S. R. Nagel, Phys. Rev. E **68**, 011306 (2003).

56 C. E. Shannon, Bell System Technical Journal **27**, 379423 (1948).

57 C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (Univ. of Illinois Press, Urbana, IL, 1949).

58 D. J. Wales, Mol. Phys. **78**, 151 (1993).

59 E. W. Dijkstra, Numerische Math. **1**, 269 (1959).

60 H. Pelzer and E. Wigner, Z. Phys. Chem. **B15**, 445 (1932).

61 W. Forst, *Theory of Unimolecular Reactions* (Academic Press, New York, 1973).

24

[62]K. J. Laidler, *Chemical Kinetics* (Harper & Row, New York, 1987).

[63]J. M. Carr and D. J. Wales, J. Phys. Chem. B **112**, 8760 (2008).

[64]A. Godzik, A. Kolinski, and J. Skolnick, J. Comput.-Aided Mol. Des. **7**, 397 (1993).

[65]D. U. Ferreiro, J. A. Hegler, E. A. Komives, and P. G. Wolynes, Proc. Nat. Acad. Sci. USA **104**, 19819 (2007).

[66]D. U. Ferreiro, J. A. Hegler, E. A. Komives, and P. G. Wolynes, Proc. Nat. Acad. Sci. USA **108**, 3499 (2011).