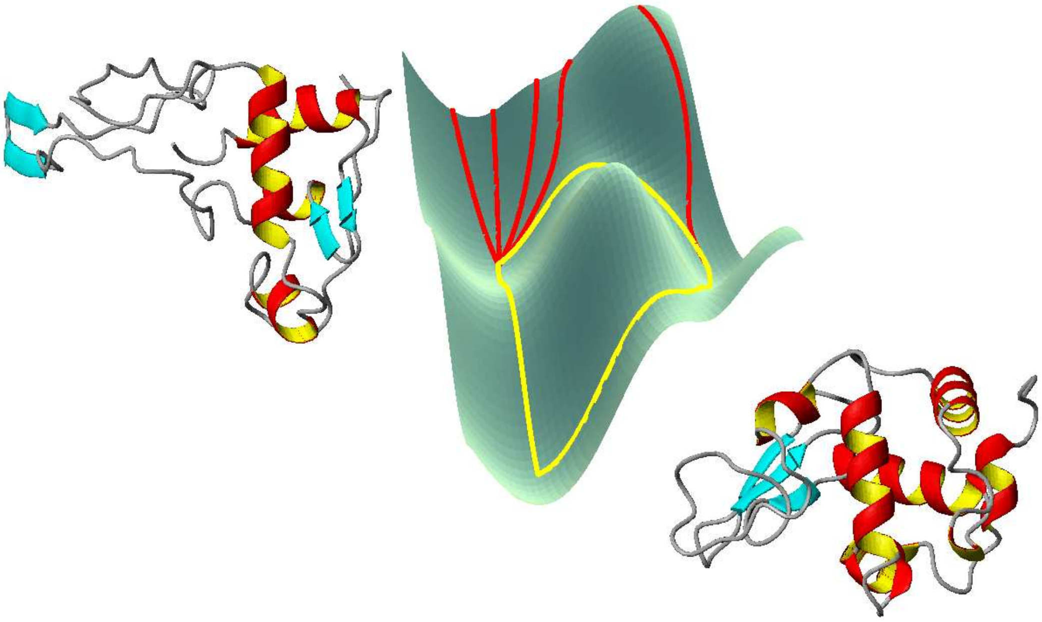
Energy Landscapes

Jonathan P. K. Doye

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



- Energy Landscapes: from Proteins to Glasses via Clusters
- The Born-Oppenheimer Approximation
- Stationary Points and Pathways
- Tunneling Splittings
- Exploring the Landscape
- Describing the Landscape
- Global Thermodynamic and Dynamic Properties

Books: ‘Energy Landscapes’, D. J. Wales, CUP, 2003.

1 Introduction

In classical mechanics the potential energy function, V, determines the structure, dynamics and thermodynamics of any system. Stable configurations occur at minima in V, defining the structure; the gradient of V gives (minus) the forces on the particles, which appear in the equations of motion; and the partition function, which determines the thermodynamics, is also a function of V. Analogous statements can be made in quantum mechanics within the Born–Oppenheimer approximation (§2), where the potential energy surface (PES) is the solution of the Schrödinger equation for frozen nuclear positions.

The idea of a PES should be a familiar concept to a chemist. The 'energy' profile of a reaction is typically introduced in A-level chemistry. In 1A kinetics the PES for a collinear three-atom reaction was depicted. However, for systems typically studied by chemists the PES is a very simple object. For an isolated chemical system there may be only a single minimum, or if there is some conformational freedom present there may be a few isomers, e.g. in n-butane there are two minima corresponding to trans and gauche isomers. For these small systems, there are well-developed methods to calculate partition functions and rate constants. But what happens when we wish to understand systems, such as clusters, biomolecules or liquids, which have a large number of degrees of conformational freedom?

When we are interested in the structure and physical properties of a chemical system we often mean the properties of the global minimum on the PES, or the properties at zero Kelvin. However, the dynamics of a system at temperatures or energies where it can escape from the global minimum depend on larger regions of the surface. The topology and topography of these extended regions determines the precise behaviour. When considering the wider features of the PES in this way, it has become common to refer to it as the 'potential energy landscape', or just the 'energy landscape'. The latter expression is also used to refer to free energy, though this should be clear from the context.

In this introductory material, I wish to illustrate how the energy landscape can be used to understand some of the properties of complex systems, by focussing on structural relaxation as an example.

Structural relaxation plays a key role in a diverse range of problems in chemical physics, including protein folding, glass formation, and the observation of 'magic number' peaks in the mass spectra of cluster ions.

1.1 Levinthal's Paradox

In the late 1960's the first attempts were being made to predict the three-dimensional structure of globular proteins from their amino acid sequence. Anfinsen had shown that some denatured proteins regained their native structure reliably on a laboratory time scale.

By coarse-graining configuration space, Levinthal realised that the number of possible conformations for a typical protein is astronomically large. If they were searched at random on the time scale of a typical vibration the time required to find the native state would probably exceed the lifetime of the universe (problem 1). (See http://www-wales.ch.cam.ac.uk/~mark/levinthal/levinthal.html.) This discrepancy has come to be known as Levinthal's paradox. It has now been largely resolved from a better understanding of energy landscapes.

The first attempts to address the paradox centred on the idea that not all the possible configurations have to be searched. Levinthal himself suggested that folding might occur via a specific pathway, initiated by the formation of a condensed nucleus and proceeding through a well-defined sequence of events. Alternatively, it was also suggested that different regions of secondary structure could start to form simultaneously or that the initial random coil might collapse to a compact state or molten globule, which would then need to rearrange.

Simulations of simple lattice models suggest that the initial collapse is indeed fast and insensitive to the residue sequence. However, it seems unlikely that a reduction in the search space is the solution to the paradox, since the number of states remaining would still be too large for a random search to succeed. However, such ideas also contain the implicit notion that, in some way, the search is not random. In terms of the energy landscape there are two reasons for this. Firstly, conformations have different statistical weights in the thermodynamic ensemble, and secondly, they are not arranged at random in configuration space.

Levinthal's analysis assumes that the energy landscape is flat, like a golf course with a single hole corresponding to the native state. For a simple model that includes an energetic bias towards the native structure the search time can be dramatically reduced to physically meaningful scales. In 1992 it was proposed that the landscape of a naturally occurring protein consists of a collection of convergent kinetic pathways that lead to a unique thermodynamically stable native state.

This landscape structure was termed a folding 'funnel' because it focuses the manifold of misfolded states towards the correct target. The random search element of Levinthal's 'paradox' is simply removed. In fact, we could

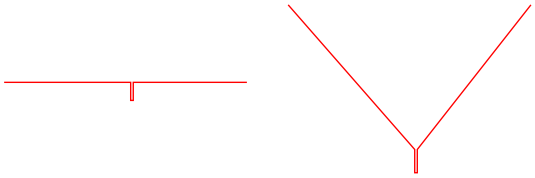


FIGURE 1: Flat (left) and biased (right) golf courses.

ask similar questions about crystallisation, which may be an even more statistically improbable event, and how it is that buckminsterfullerene is selected from a mass of vaporised small carbon chains when there are vast numbers of other possible isomers?

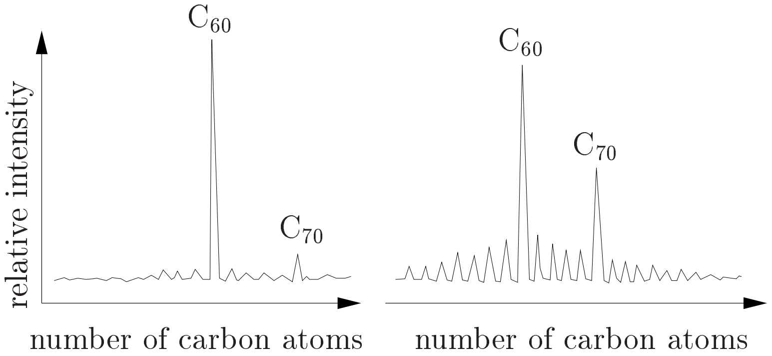


FIGURE 2: Time of flight mass spectra for carbon clusters prepared by laser vaporisation at high (left) and medium (right) buffer gas pressures.

Free energy funnels for proteins, which incorporate the entropy, have now been investigated for a variety of models. To do this it is necessary to express the free energy (usually Helmholtz) in terms of order parameters. The free energy landscape depends upon temperature and mass—the potential energy surface does not.

Often the dependence of the free energy, A, is expressed in terms of a single order parameter, Q, which is designed to somehow measure the similarity of the protein configuration to the native state. The function $A(Q)$ may exhibit a double minimum over a particular temperature range, allowing a thermodynamic 'transition state' to be defined between folded and unfolded states.

An important research goal is to understand why some amino acid sequences fold rapidly and reliably whilst others do not. Some simulations suggest that the folding ability can be measured by the ratio of the folding temperature, $T_{\rm f}$, below which the native state has the lowest free energy, to

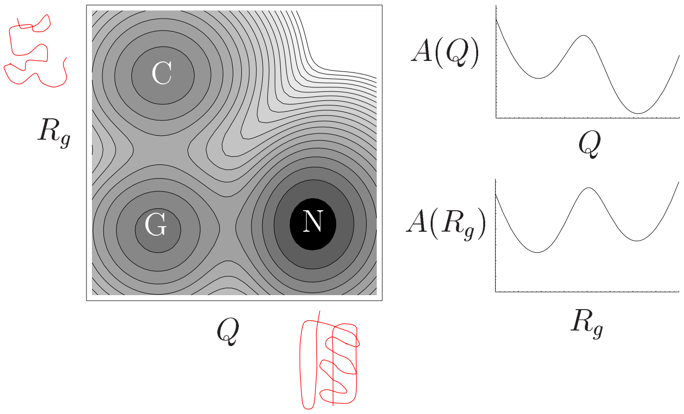


FIGURE 3: Protein free energy surface as a function of the radius of gyration, R_g, and the number of native contacts, Q. N=native state, C='random coil', G='molten globule'.

the 'glass' transition temperature, $T_{\rm g}$, where the kinetics exhibit a dramatic slowing down. Folding is easiest for large $T_{\rm f}/T_{\rm g}$, since the native state is then statistically populated at temperatures where it is kinetically accessible.

For a simple lattice model it was suggested that efficient folding occurs when the global potential energy minimum lies well below the other local minima. This situation certainly increases the value of $T_{\rm f}$, but the strength of this correlation has since been questioned. Relaxation from high-lying states must also depend on global properties of the potential energy surface, not just the region around the global minimum. In fact, surfaces exist with pronounced potential energy minima that are kinetically inaccessible (§6).

The energy of a protein depends on interactions between parts of the chain that come into contact as it folds. If the favourable interactions that define the native state involve bringing together other residues that interact unfavourably, or if two similar structures can only be interconverted by disrupting contacts that have been made elsewhere, the system is said to be 'frustrated'. Both these effects lead to roughening of the potential energy landscape. Proteins that avoid these problems as far as possible are said to obey the principle of 'minimum frustration'.

It is now generally believed, or at least assumed, that the native state of a protein is the global free energy minimum at the appropriate temperature (Anfinsen's thermodynamic principle). To the extent that protein folding is a search for the global potential energy minimum, it is an example of the more general problem of global optimisation (§10), which is of great interest in fields as different as economics and microelectronics. The success of any global optimisation algorithm is inextricably linked to the structure of the landscape, and an understanding of these connections has recently been

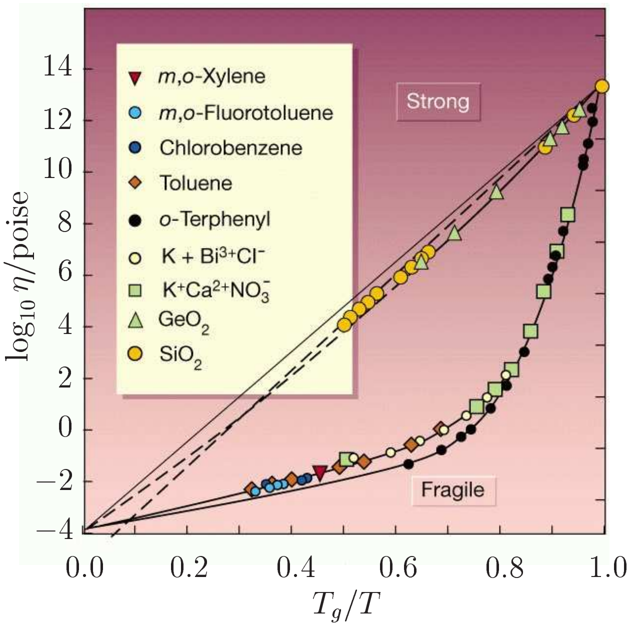


FIGURE 4: Arrhenius plots of viscosity data scaled by values of T_g for a variety of strong and fragile glass formers.

exploited to design more effective techniques (§10).

1.2 'Strong' and 'Fragile' Liquids

The potential energy landscape also plays an important role in determining the behaviour of both finite and bulk matter in the liquid state. Angell's scheme classifies liquids on a scale from 'strong' to 'fragile' using both thermodynamic criteria, such as the entropy relative to the crystal, and dynamic criteria, such as the viscosity. Dynamic criteria are generally related to a reorganisation or structural relaxation time scale.

A strong liquid is characterised by a diffusion constant and viscosity that follow the Arrhenius form ($\propto \exp[A/T]$). These are often liquids with strongly bonded open network structures like SiO_2, BeF_2 and P_2O_5. Fragile liquids exhibit non-Arrhenius behaviour and tend to have more isotropic interactions.

Strong materials also have small heat capacity differences between the supercooled liquid and crystal; fragile materials exhibit large heat capacity differences. The violation of the third law of thermodynamics by the extrapolated entropy of a supercooled liquid is known as the Kauzmann entropy paradox. The 'paradox' is avoided by the intervention of the glass transition where the viscosity increases rapidly and liquid-like diffusion ceases.

Whether we consider the viscosity or a general relaxation time, τ, the

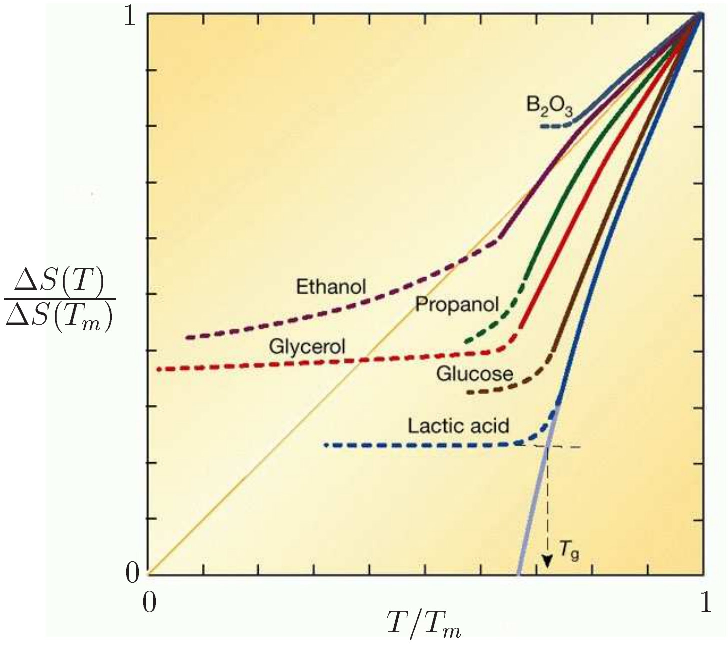


FIGURE 5: Schematic variation of the excess entropy, $\Delta S(T) = S_{\text{liquid}}(T) - S_{\text{crystal}}(T)$, divided by $\Delta S(T_{\text{m}})$, the value at the melting point, as a function of T/T_{m} for some strong and fragile glass formers. The extrapolation leading to paradoxical behaviour is indicated for lactic acid.

temperature dependence can often be described over a certain range by the empirical Vogel-Tammann-Fulcher (VTF) equation:

$$\tau = \tau_0 \exp[DT_0/(T - T_0)], \tag{1}$$

where τ_0, D and T_0 are fitted parameters. Arrhenius behaviour is recovered for small T_0 and large D.

For strong liquids the relaxation behaviour of the property described by τ is usually found to fit a simple exponential form, $p(t) = p(0)e^{-t/\tau}$ (Debye relaxation). For fragile liquids a variety of alternative functional forms have been suggested, such as the Kohlrausch-Williams-Watts stretched exponential, $p(t) = p(0)\exp[-a\, t^{\theta}]$. $\theta = 1$ corresponds to Debye relaxation. To define a general relaxation time we can use

$$\int_0^\infty \frac{p(t) - p(\infty)}{p(0) - p(\infty)} dt, \tag{2}$$

which gives the expected result for Debye relaxation (problem 2).

As the temperature is lowered the absorption spectrum of a supercooled liquid or glass corresponding to dielectric relaxation often exhibits a peak that moves to lower frequency as the temperature decreases. Once this primary or α peak has disappeared a weaker β absorption may be revealed, which in turn moves to lower frequency as the temperature decreases. The α peak frequency as a function of temperature often exhibits non-Arrhenius behaviour and has been associated with slower molecular rearrangements.

Adam and Gibbs provided a heuristic connection between thermodynamics and dynamics by considering 'cooperative relaxation' to obtain

$$\tau = \tau_0 \exp(A\Delta\mu/TS_{\mathrm{c}}), \tag{3}$$

where A is a constant, $\Delta\mu$ is the free energy barrier per molecule in the cooperative group and $S_{\mathrm{c}} =$ is the excess configurational entropy of the liquid over the crystal. The VTF equation (1) can be recovered if $C_{p,\mathrm{liquid}} - C_{p,\mathrm{crystal}} \propto 1/T$ (problem 3).

In fact it is not always possible to classify liquids unambiguously as strong or fragile, and some strong liquids become fragile at higher densities. For example, the diffusion of both Si and O atoms in silica becomes strongly non-Arrhenius when the density is increased by 30%.

Glasses often (not always) exhibit other characteristic features such as a low temperature heat capacity that varies linearly with T. This phenomenon may be due to quantum mechanical tunnelling (§4) between nearly degenerate minima separated by tiny barriers. Many glasses also exhibit an excess vibrational density of states at low frequency compared to the usual Debye dependence ($g(\nu) \propto \nu^2$). This feature is termed the 'boson peak'.

Although much has been measured and fitted in the field of supercooled liquids and glasses, rather less is actually understood. However, the fundamental basis for all such behaviour must lie in the underlying PES, as Goldstein realised as early as 1969. For example, the slow α and fast β processes are generally presumed to be associated with mechanisms involving larger and smaller barriers.

2 The Born-Oppenheimer Approximation

The Born-Oppenheimer approximation is a key element in the notion of a potential energy surface. The Schrödinger equation for a molecule with n electrons, mass m_e, and N nuclei, masses M_t, is

$$\left[-\sum_{t=1}^{N}\frac{\hbar^2}{2M_t}\nabla_t^2-\sum_{i=1}^{n}\frac{\hbar^2}{2m_\mathrm{e}}\nabla_i^2+V(\mathbf{x},\mathbf{X})\right]\psi(\mathbf{x},\mathbf{X})=E_\mathrm{total}\psi(\mathbf{x},\mathbf{X}), \qquad (4)$$

where $\mathbf{x}$ and $\mathbf{X}$ represent the electronic and nuclear coordinates. The potential energy is

$$V(\mathbf{x},\mathbf{X})=\frac{e^2}{4\pi\epsilon_0}\left[-\sum_{\mathrm{all}\ i,t}\frac{Z_t}{r_{it}}+\sum_{i<j}\frac{1}{r_{ij}}+\sum_{t<s}\frac{Z_tZ_s}{r_{ts}}\right], \qquad (5)$$

where Z_t is the atomic number (nuclear charge) of nucleus t and e is the modulus of the electronic charge. $r_{it}=|\mathbf{x}_i-\mathbf{X}_t|$ where $\mathbf{x}_i$ and $\mathbf{X}_t$ are the position vectors of electron i and nucleus t, etc.

It is more convenient to use atomic units where $e=1$, $m_\mathrm{e}=1$, $4\pi\epsilon_0=1$ and $\hbar=1$. The units of energy and length are then the hartree, $E_h=e^2/4\pi\epsilon_0 a_0$, and the Bohr radius, $a_0=4\pi\epsilon_0\hbar^2/m_\mathrm{e}e^2$.

Denote the total Hamiltonian operator in equation (4) by $\hat{\mathcal{H}}$, and the nuclear kinetic energy operator [the first term in equation (4)] by $\hat{T}_\mathrm{n}$. The Born-Oppenheimer electronic wavefunction is the solution of

$$\left(\hat{\mathcal{H}}-\hat{T}_\mathrm{n}\right)\psi_\mathrm{e}=V_\mathrm{e}\psi_\mathrm{e}. \qquad (6)$$

ψ_e is a function of the electronic coordinates $\mathbf{x}$, but only depends upon the nuclear positions $\mathbf{X}$ parametrically, because equation (6) is solved for some particular nuclear geometry. Hence we write $\psi_\mathrm{e}(\mathbf{x};\mathbf{X})$ and $V_\mathrm{e}(\mathbf{X})$ to show that different electronic wavefunctions and energies are obtained for different atomic structures.

The potential energy surface defines the variation of the electronic energy, $V_\mathrm{e}(\mathbf{X})$, with the nuclear geometry. In a diatomic molecule we simply consider $V_\mathrm{e}(R)$ as the bond-length, R, varies. A classical mechanical justification for the Born-Oppenheimer approximation is that the ratio of the mass of a proton to the mass of an electron is about 1,836. Hence, nuclear velocities are expected to be very small in comparison with electronic velocities.

Assuming that $V_\mathrm{e}(\mathbf{X})$ defines an effective potential in which the nuclei move, the appropriate Schrödinger equation for the nuclear wavefunction, $\psi_\mathrm{n}(\mathbf{X})$, is

$$\left(\hat{T}_\mathrm{n}(\mathbf{X})+V_\mathrm{e}(\mathbf{X})\right)\psi_\mathrm{n}(\mathbf{X})=E_\mathrm{total}\psi_\mathrm{n}(\mathbf{X}). \qquad (7)$$

Write $\psi(\mathbf{x},\mathbf{X}) = \psi_{\rm e}(\mathbf{x};\mathbf{X})\psi_{\rm n}(\mathbf{X})$, substitute this into equation (4) and neglect all the terms involving derivatives of $\psi_{\rm e}(\mathbf{x};\mathbf{X})$ with respect to nuclear coordinates, i.e. $\psi_{\rm n}\nabla_t^2\psi_{\rm e}$ and $\nabla_t\psi_{\rm n} \bullet \nabla_t\psi_{\rm e}$. Separating the variables gives (divide by $\psi_{\rm e}(\mathbf{x};\mathbf{X})\psi_{\rm n}(\mathbf{X})$, problem 4)

$$-\sum_{i=1}^{n}\frac{\nabla_i^2\psi_{\rm e}(\mathbf{x};\mathbf{X})}{2\psi_{\rm e}(\mathbf{x};\mathbf{X})} + V(\mathbf{x},\mathbf{X}) = \sum_{t=1}^{N}\frac{\nabla_t^2\psi_{\rm n}(\mathbf{X})}{2M_t\psi_{\rm n}(\mathbf{X})} + E_{\rm total} = V_{\rm e}. \tag{8}$$

Hence we recover equation (6) and equation (7). The key point is that the latter two equations are much more tractable than equation (4).

Potential energy surfaces only exist within the Born-Oppenheimer approximation. If the approximation were exactly true then H−D would have no dipole moment, because the extra neutron in the frozen deuterium nucleus would not affect the electrons. This is why we usually assume that the (electronic) potential energy curve (or the effective force constant) of a diatomic molecule is unchanged by isotopic substitution. In fact H−D has a dipole moment of order 10^{-4} D (1 Debye= 3.336×10^{-30} Cm); for comparison, CH_3F has a dipole moment of 1.85 D. Hence the approximation is very good for H−D, and even better for heavier atoms.

2.1 Normal Modes

Various approaches to the nuclear dynamics problem summarised in equation (7) are commonly employed. The simplest is normal mode analysis, which requires an orthogonal transformation of coordinates.

2.1.1 Orthogonal Transformations

In general, we interpret the elements of an n-dimensional (column) vector, $\mathbf{P} = (P_1, P_2, \ldots, P_n)^T$ as the components of a vector quantity in a particular basis of orthonormal (i.e. orthogonal and normalised) unit vectors, $\hat{\mathbf{E}}_\alpha$, defined by

$$\hat{\mathbf{E}}_\alpha = \frac{\partial\vec{\mathbf{R}}}{\partial u_\alpha}\Big/\left|\frac{\partial\vec{\mathbf{R}}}{\partial u_\alpha}\right|, \tag{9}$$

where $\vec{\mathbf{R}}$ is a position vector that depends upon the coordinate u_α. We must distinguish between the components, denoted by the column matrix $\mathbf{P}$, and the vector

$$\vec{\mathbf{P}} = \sum_{\alpha=1}^{n} P_\alpha\,\hat{\mathbf{E}}_\alpha. \tag{10}$$

If we change basis to a new set of orthonormal vectors, $\hat{\mathbf{e}}_\alpha$, where

$$\hat{\mathbf{e}}_\alpha = \sum_{\beta=1}^{n} C_{\alpha\beta}\,\hat{\mathbf{E}}_\beta,$$

$$\text{then}\qquad \vec{\mathbf{p}} = \vec{\mathbf{P}} = \sum_{\alpha=1}^{n} p_\alpha\,\hat{\mathbf{e}}_\alpha = \sum_{\alpha,\beta=1}^{n} C_{\alpha\beta}\,p_\alpha\,\hat{\mathbf{E}}_\beta,$$

$$\text{so that}\qquad P_\beta = \sum_{\alpha=1}^{n} C_{\alpha\beta}\,p_\alpha \qquad \text{or} \qquad \mathbf{P} = \mathbf{C}^T\mathbf{p}. \tag{11}$$

For transformations between orthonormal basis sets $\mathbf{C}$ is an orthogonal matrix (see problem 8) and distances and scalar products are conserved:

$$(\mathbf{p}_1)^T(\mathbf{p}_2) = (\mathbf{p}_1)^T\mathbf{C}\,\mathbf{C}^T(\mathbf{p}_2) = (\mathbf{C}^T\mathbf{p}_1)^T(\mathbf{C}^T\mathbf{p}_2) = (\mathbf{P}_1)^T(\mathbf{P}_2). \tag{12}$$

Components of a vector quantity in different orthonormal bases are connected by the same transformation matrix $\mathbf{C}$ as the basis vectors themselves.

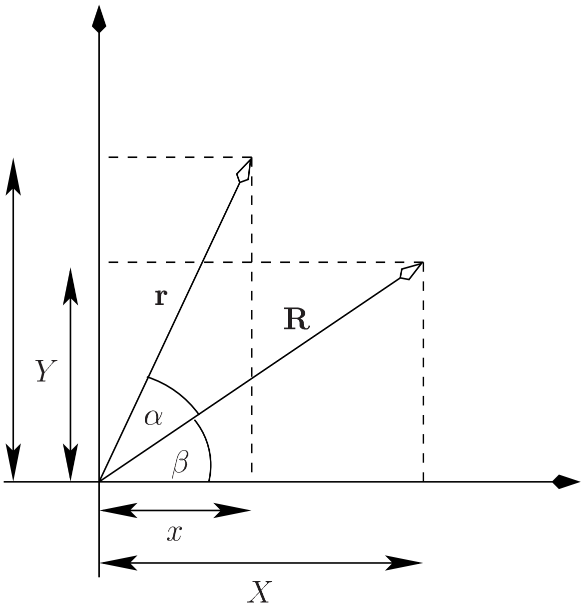


FIGURE 6: For an anticlockwise rotation of position vector $\mathbf{R}$, length R, through angle α the initial and final coordinates are related by

$$\begin{aligned} x &= R\cos(\alpha+\beta) = R\cos\alpha\cos\beta - R\sin\alpha\sin\beta = X\cos\alpha - Y\sin\alpha, \\ y &= R\sin(\alpha+\beta) = R\sin\alpha\cos\beta + R\cos\alpha\sin\beta = X\sin\alpha + Y\cos\alpha. \end{aligned}$$

If the components of $\mathbf{P}$ and $\mathbf{p}$ correspond to different coordinates then it may be helpful to define a function of the new coordinates, f, such that

$f(\mathbf{p}) \equiv F(\mathbf{P}) = F(\mathbf{C}^{-1}\mathbf{p})$ at any point. The new function, f, evaluated at $\mathbf{p}$ gives the same result as the old function, F, evaluated at $\mathbf{P}$. For example, suppose that $\mathbf{C}$ represents an anticlockwise rotation through 45° about the z axis (Figure 6). The components of the initial and final position vectors of an arbitrary point, $\mathbf{R}$ and $\mathbf{r}$, are related by:

$$\mathbf{r} = \mathbf{C}\,\mathbf{R} \quad \text{so that} \quad \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{C} \begin{pmatrix} X \\ Y \end{pmatrix},$$

$$\text{where} \quad \mathbf{C} = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}. \tag{13}$$

The potential energy function is independent of the coordinate system, but the functional form will generally change; for example, if

$$V(\mathbf{R}) \equiv V(X, Y) = X^2 + 2Y^2, \tag{14}$$

then for the rotation considered in equation (13)

$$\begin{aligned} v(\mathbf{r}) \equiv v(x, y) &= V(\mathbf{C}^{-1}\mathbf{r}) = V\left(\frac{(x+y)}{\sqrt{2}}, \frac{(y-x)}{\sqrt{2}}\right) \\ &= \frac{3}{2}(x^2 + y^2) - xy = X^2 + 2Y^2, \end{aligned} \tag{15}$$

so that $V(\mathbf{R}) = v(\mathbf{r})$ as required. The magnitude of the position vector is also unaffected by the coordinate transformation:

$$\mathbf{r} \cdot \mathbf{r} = x^2 + y^2 = X^2 + Y^2 = \mathbf{R} \cdot \mathbf{R}. \tag{16}$$

The gradient vector and its components in the two coordinate systems are

$$\begin{aligned} \mathbf{G}(X, Y) &= (\nabla_X V(X, Y), \nabla_Y V(X, Y))^T = (2X, 4Y)^T \\ \text{and} \quad \vec{\mathbf{G}}(X, Y) &= 2X\,\hat{\mathbf{E}}_X + 4Y\,\hat{\mathbf{E}}_Y, \end{aligned} \tag{17}$$

$$\begin{aligned} \mathbf{g}(x, y) &= (\nabla_x v(x, y), \nabla_y v(x, y))^T = (3x - y, 3y - x)^T \\ \text{and} \quad \vec{\mathbf{g}}(x, y) &= (3x - y)\hat{\mathbf{e}}_x + (3y - x)\hat{\mathbf{e}}_y. \end{aligned} \tag{18}$$

These components are related in the same way as the components of the position vector, $\mathbf{R}$:

$$\begin{aligned} \mathbf{C}\,\mathbf{G}(X, Y) &= \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 2X \\ 4Y \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 2X - 4Y \\ 2X + 4Y \end{pmatrix} \\ &= \begin{pmatrix} 3x - y \\ 3y - x \end{pmatrix} = \mathbf{g}(x, y). \end{aligned} \tag{19}$$

We can also verify that the gradient at any point is unaffected by the transformation

$$\vec{\mathbf{g}}(x,y) = (3x-y)\hat{\mathbf{e}}_x + (3y-x)\hat{\mathbf{e}}_y = 2X\,\hat{\mathbf{E}}_X + 4Y\,\hat{\mathbf{E}}_Y = \vec{\mathbf{G}}(X,Y), \tag{20}$$

but the function itself changes, and must therefore be represented by a new symbol.

In general, components of a vector quantity in different orthonormal bases are connected by the same transformation matrix $\mathbf{C}$ as the basis vectors themselves. Here we have considered the rotation according to the *active* convention, where position vectors and functions are rotated. There is an alternative, *passive*, convention, where the coordinate axes are rotated instead. The two conventions are equivalent if we account for the fact that a clockwise rotation of position vectors corresponds to an anticlockwise rotation of the axes.

The unit vectors in the (x,y) coordinate system are defined by

$$\begin{aligned}
\hat{\mathbf{e}}_x &= \frac{\partial\vec{\mathbf{R}}}{\partial x}\Big/\left|\frac{\partial\vec{\mathbf{R}}}{\partial x}\right| = \left(\frac{\partial X}{\partial x}\hat{\mathbf{E}}_X + \frac{\partial Y}{\partial x}\hat{\mathbf{E}}_Y\right)\Big/\left|\frac{\partial\vec{\mathbf{R}}}{\partial x}\right| = \frac{1}{\sqrt{2}}(\hat{\mathbf{E}}_X - \hat{\mathbf{E}}_Y)\\
\hat{\mathbf{e}}_y &= \frac{\partial\vec{\mathbf{R}}}{\partial y}\Big/\left|\frac{\partial\vec{\mathbf{R}}}{\partial y}\right| = \left(\frac{\partial X}{\partial y}\hat{\mathbf{E}}_X + \frac{\partial Y}{\partial y}\hat{\mathbf{E}}_Y\right)\Big/\left|\frac{\partial\vec{\mathbf{R}}}{\partial y}\right| = \frac{1}{\sqrt{2}}(\hat{\mathbf{E}}_X + \hat{\mathbf{E}}_Y).
\end{aligned}$$

If we use this result to evaluate $C_{\alpha\beta} = \hat{\mathbf{e}}_\alpha \cdot \hat{\mathbf{E}}_\beta$ from (11) the answer is consistent with equation (13).

Once we know how to transform between different coordinates using equation (11), derivatives of a function such as the potential energy can easily be obtained from the chain rule:

$$\frac{\partial}{\partial P_i} = \sum_j \frac{\partial p_j}{\partial P_i}\frac{\partial}{\partial p_j} = \sum_j C_{ji}\frac{\partial}{\partial p_j}, \tag{21}$$

because $\mathbf{p} = \mathbf{C}\,\mathbf{P}$, so that $p_j = C_{ji}\,P_i$.

2.1.2 The Normal Mode Transformation

We first expand $V_{\mathrm{e}}(\mathbf{X})$ in a Taylor series around a configuration corresponding to stable equilibrium at $\mathbf{X}^0$:

$$V_{\mathrm{e}}(\mathbf{X}) \approx V_{\mathrm{e}}(\mathbf{X}^0) + \frac{1}{2}\sum_{\alpha,\beta=1}^{3N} \frac{\partial^2 V_{\mathrm{e}}(\mathbf{X}^0)}{\partial X_\alpha \partial X_\beta}(X_\alpha - X_\alpha^0)(X_\beta - X_\beta^0). \tag{22}$$

The first derivatives of $V_{\mathrm{e}}(\mathbf{X})$ vanish if $\mathbf{X}^0$ is a local minimum. Substituting the truncated series expansion of V_{e} into equation (7) gives

$$\hat{T}_{\mathrm{n}}(\mathbf{X}) + V_{\mathrm{e}}(\mathbf{X}) = -\sum_{t=1}^{N} \frac{\nabla_t^2}{2M_t} + \frac{1}{2}(\mathbf{X} - \mathbf{X}^0)^T \mathbf{H}(\mathbf{X}^0)(\mathbf{X} - \mathbf{X}^0), \tag{23}$$

where $H_{\alpha\beta}(\mathbf{X}^0) = \partial^2 V_{\mathrm{e}}(\mathbf{X}^0)/\partial X_\alpha \partial X_\beta$, and we have chosen the energy origin such that $V_{\mathrm{e}}(\mathbf{X}^0) = 0$.

The first coordinate transformation required is a simple rescaling, which is governed by the atomic mass according to $\mathbf{q}_t = \mathbf{X}_t \sqrt{M_t}$, where $\mathbf{X}_t$ is the three-component position vector of nucleus t. Only the diagonal elements of the corresponding transformation matrix are nonzero, but this matrix is not orthogonal. The kinetic energy operator now takes on a more symmetrical form:

$$\hat{t}_{\mathrm{n}}(\mathbf{q}) + v_{\mathrm{e}}(\mathbf{q}) = -\frac{1}{2} \sum_{\alpha=1}^{3N} \frac{\partial^2}{\partial q_\alpha^2} + \frac{1}{2}(\mathbf{q} - \mathbf{q}^0)^T \, \mathbf{h}(\mathbf{q}^0) \, (\mathbf{q} - \mathbf{q}^0), \tag{24}$$

where $h_{\alpha\beta}(\mathbf{q}^0) = \partial^2 v_{\mathrm{e}}(\mathbf{q}^0)/\partial q_\alpha \partial q_\beta$, or $h_{\alpha\beta}(\mathbf{q}^0) = H_{\alpha\beta}(\mathbf{X}^0)/\sqrt{M_\alpha M_\beta}$, with M_α the nuclear mass corresponding to the component q_α. Both $\mathbf{H}$ and $\mathbf{h}$ are commonly referred to as Hessian or second derivative matrices.

The symmetric $3N \times 3N$ matrix $\mathbf{h}$ can be diagonalised using a matrix, $\mathbf{A}$, whose columns are the eigenvectors of $\mathbf{h}$ with associated eigenvalues ω_γ^2:

$$\sum_{\beta=1}^{3N} h_{\alpha\beta} \, A_{\beta\gamma} = \omega_\gamma^2 \, A_{\alpha\gamma}. \tag{25}$$

The γth eigenvector of $\mathbf{h}$ therefore has elements $A_{\beta\gamma}$ with β running from 1 to $3N$:

$$(A_{1\gamma}, \; A_{2\gamma}, \; A_{3\gamma}, \ldots, \; A_{3N\,\gamma})^T. \tag{26}$$

The quantities ω_γ correspond to normal mode angular frequencies, as shown below. The matrix $\mathbf{h}$ is symmetric if the potential energy has continuous second derivatives, and we can then choose the eigenvectors to be orthonormal so that

$$\sum_{\beta=1}^{3N} A_{\beta\gamma} \, A_{\beta\epsilon} = \sum_{\beta=1}^{3N} A_{\gamma\beta}^T \, A_{\beta\epsilon} = \delta_{\gamma\epsilon}, \tag{27}$$

or $\mathbf{A}^T\mathbf{A} = \mathbf{I} = \mathbf{A}\,\mathbf{A}^T$, and $\mathbf{A}$ is therefore an orthogonal matrix.

We now define new coordinates, Q_γ, by the orthogonal transformation

$$q_\alpha = \sum_{\gamma=1}^{3N} A_{\alpha\gamma} \, Q_\gamma, \quad \text{so that} \quad Q_\beta = \sum_{\alpha=1}^{3N} A_{\alpha\beta} \, q_\alpha. \tag{28}$$

This transformation leads to the desired separation of variables in the Hamiltonian. If we move the origin of coordinates to the position of the stationary point then in terms of the new variables we have

$$\begin{aligned}
\frac{1}{2}\sum_{\alpha,\beta=1}^{3N} h_{\alpha\beta}\, q_\alpha\, q_\beta &= \frac{1}{2}\sum_{\alpha,\beta,\gamma,\delta=1}^{3N} h_{\alpha\beta}\, A_{\alpha\gamma}\, A_{\beta\delta}\, Q_\gamma\, Q_\delta \\
&= \frac{1}{2}\sum_{\alpha,\gamma,\delta=1}^{3N} A_{\alpha\gamma}\, \omega_\delta^2\, A_{\alpha\delta}\, Q_\gamma\, Q_\delta = \frac{1}{2}\sum_{\delta=1}^{3N} \omega_\delta^2\, Q_\delta^2. \\
-\frac{1}{2}\sum_{\alpha=1}^{3N} \frac{\partial^2}{\partial q_\alpha^2} &= -\frac{1}{2}\sum_{\alpha,\beta,\gamma=1}^{3N} \left(\frac{\partial Q_\beta}{\partial q_\alpha}\frac{\partial}{\partial Q_\beta}\right)\left(\frac{\partial Q_\gamma}{\partial q_\alpha}\frac{\partial}{\partial Q_\gamma}\right) \qquad (29) \\
&= -\frac{1}{2}\sum_{\alpha,\beta,\gamma=1}^{3N} \frac{\partial^2}{\partial Q_\beta \partial Q_\gamma}\, A_{\alpha\beta}\, A_{\alpha\gamma} = -\frac{1}{2}\sum_{\beta=1}^{3N} \frac{\partial^2}{\partial Q_\beta^2},
\end{aligned}$$

where we have used the chain rule to transform the derivative $\partial/\partial q_\alpha$ and the orthogonality of $\mathbf{A}$ to simplify the sums, as well as equation (25).

The overall transformation, $Q_\beta = \sum_{\alpha=1}^{3N} A_{\alpha\beta}\, X_\alpha \sqrt{M_\alpha}$, therefore reduces the Hamiltonian to an uncoupled sum of $3N$ simple harmonic oscillator Hamiltonians, one for each normal coordinate:

$$\hat{\mathcal{T}}_{\mathrm{n}}(\mathbf{Q}) + \mathcal{V}_{\mathrm{e}}(\mathbf{Q}) = \frac{1}{2}\sum_{\alpha=1}^{3N}\left[-\frac{\partial^2}{\partial Q_\alpha^2} + \omega_\alpha^2\, Q_\alpha^2\right]. \qquad (30)$$

For such a separable Hamiltonian the wavefunction is a simple product of solutions for the separate degrees of freedom and the energy is a simple sum; the normal mode transformation is now complete.

The corresponding classical Hamiltonian is

$$\frac{1}{2}\sum_{\alpha=1}^{3N}[\dot{Q}_\alpha^2 + \omega_\alpha^2\, Q_\alpha^2], \qquad (31)$$

where $\dot{Q}_\alpha$ denotes the time derivative. Classically, we must therefore solve $3N$ uncoupled equations, each with the form

$$\dot{Q}_\alpha^2 = -\omega_\alpha^2\, Q_\alpha^2 + \text{constant}. \qquad (32)$$

Differentiating with respect to time produces the familiar differential equation for simple harmonic motion:

$$\ddot{Q}_\alpha = -\omega_\alpha^2\, Q_\alpha, \qquad (33)$$

and the solution can be written as

$$Q_\alpha = c_\alpha \cos\left(\omega_\alpha t + \epsilon_\alpha\right), \tag{34}$$

where c_α and ϵ_α are two constants that would be fixed by two initial conditions, as expected for a second-order differential equation, and t is the time.

To obtain the general solution for the classical harmonic motion of a nuclear coordinate, X_α, we must now invert the transformation:

$$\begin{aligned} X_\alpha = q_\alpha/\sqrt{M_\alpha} &= \sum_{\gamma=1}^{3N} A_{\alpha\gamma}\, Q_\gamma/\sqrt{M_\alpha} \\ &= \sum_{\gamma=1}^{3N} A_{\alpha\gamma}\, c_\gamma \cos\left(\omega_\gamma t + \epsilon_\gamma\right)/\sqrt{M_\alpha}. \end{aligned} \tag{35}$$

The Q_α are the normal mode coordinates, with fundamental angular frequencies ω_α related to the conventional frequencies by $\omega_\alpha = 2\pi\nu_\alpha$. Motion corresponding to a particular normal mode γ corresponds to the appropriate term in the sum in equation (35). All the displacements X_α then oscillate in phase with frequency ω_γ and relative amplitudes $A_{\alpha\gamma}/\sqrt{M_\alpha}$. A general vibration is a superposition of all $3N$ modes.

3 Stationary Points and Pathways

The most interesting points of a potential energy surface are stationary points where the gradient vanishes. Non-linear molecules in field-free space have six zero normal mode frequencies at a stationary point, corresponding to overall translations and rotations that do not affect the energy; linear molecules have five. It is worth investigating the origin of these zero eigenvalues properly to appreciate the fundamental difference between the translational and rotational degrees of freedom.

3.1 Zero Hessian Eigenvalues

The potential energy must be invariant to an overall displacement of any magnitude, Δ, in the X direction, which we denote by $\Delta\,\hat{\mathbf{E}}_X$ where

$$\hat{\mathbf{E}}_X = \sqrt{\frac{1}{N}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}, \tag{36}$$

is the unit vector in the X direction. Taylor expansion gives

$$\begin{aligned} V(\mathbf{X}^0) &= V(\mathbf{X}^0 + \Delta\,\hat{\mathbf{E}}_X) \\ &= V(\mathbf{X}^0) + \Delta\,\mathbf{G}(\mathbf{X}^0)^T\,\hat{\mathbf{E}}_X + \frac{\Delta^2}{2}\hat{\mathbf{E}}_X^T\,\mathbf{H}(\mathbf{X}^0)\,\hat{\mathbf{E}}_X + \ldots, \end{aligned}$$

where $\mathbf{G}(\mathbf{X}^0) = \nabla V(\mathbf{X}^0)$, or $G_\alpha(\mathbf{X}^0) = \partial V(\mathbf{X}^0)/\partial X_\alpha$, and $H_{\alpha\beta}(\mathbf{X}^0) = \partial^2 V(\mathbf{X}^0)/\partial X_\alpha \partial X_\beta$.

The terms in each power of Δ must vanish for any $\mathbf{X}^0$, so

$$\mathbf{G}(\mathbf{X}^0)^T\,\hat{\mathbf{E}}_X = 0, \quad \text{and} \quad \hat{\mathbf{E}}_X^T\,\mathbf{H}(\mathbf{X}^0)\,\hat{\mathbf{E}}_X = 0. \tag{37}$$

The first condition gives a conservation law for the X component of the centre of mass linear momentum, and the second implies that $\hat{\mathbf{E}}_X$ is an eigenvector of $\mathbf{H}$ with eigenvalue zero. The corresponding eigenvector of $h_{\alpha\beta} = H_{\alpha\beta}/\sqrt{M_\alpha M_\beta}$ is therefore $(\sqrt{M_1}, 0, 0, \sqrt{M_2}, 0\ldots)/\sqrt{M}$, where M is the total mass, and gives the components of the corresponding normal mode, $Q_X^{\rm trans}$, in terms of the mass-weighted coordinates q_α. Hence

$$\begin{aligned} Q_X^{\rm trans} &= \left(\sqrt{M_1}\,q_{1X} + \sqrt{M_2}\,q_{2X} + \cdots + \sqrt{M_N}\,q_{NX}\right)/\sqrt{M}, \\ &= \left(M_1 X_1 + M_2 X_2 + \cdots + M_N X_N\right)/\sqrt{M}, \end{aligned} \tag{38}$$

where X_t is the X coordinate of atom t. Analogous results apply for translation in the Y and Z directions.

Overall rotations, however, are more complicated. If the gradient does not vanish then overall rotations are coupled to the intramolecular vibrations and we obtain no further zero eigenvalues. However, at a stationary point we find that

$$\begin{aligned} Q_X^{\rm rot} &= \left(\sqrt{M_1}\left(Z_1^0 q_{1Y} - Y_1^0 q_{1Z}\right) + \sqrt{M_2}\left(Z_2^0 q_{2Y} - Y_2^0 q_{2Z}\right)\ldots\right)/\sqrt{I_{XX}}, \\ &= \left(M_1\left(Z_1^0 Y_1 - Y_1^0 Z_1\right) + M_2\left(Z_2^0 Y_2 - Y_2^0 Z_2\right)\ldots\right)/\sqrt{I_{XX}}, \end{aligned} \tag{39}$$

is a normal mode with zero eigenvalue, and analogous arguments apply to rotations about the Y and Z axes. The diagonal component of the inertia tensor, I_{XX}, appears due to normalisation. The above translational and rotational normal mode vectors are orthonormal if the body-fixed coordinate axes coincide with the principal axes and the centre of mass lies at the origin. The accuracy of the ‘zero’ eigenvalues can be a useful measure of how well a geometry optimisation has converged.

3.2 Classification of Stationary Points

The characteristics of any stationary point are determined by the corresponding Hessian. The number of negative Hessian eigenvalues is termed the Hessian index, and this is usually the same as the order defined as the number of mutually orthogonal directions in which the energy is maximal. Minima are stationary points with no negative Hessian eigenvalues. Note that a negative Hessian eigenvalue also corresponds to a negative force constant and an imaginary normal mode frequency. A displacement along such a coordinate lowers the energy.

The Murrell-Laidler theorem defines a transition state as a stationary point with a single imaginary normal mode frequency. Consider an index two saddle with two negative force constants, k_1 and k_2, for normal modes Q_1 and Q_2. With all the other coordinates fixed the change in potential energy for small displacements away from the saddle is:

$$\Delta\mathcal{V} = \frac{1}{2}(k_1\, Q_1^2 + k_2\, Q_2^2) < 0. \tag{40}$$

Hence any displacement in this space lowers the energy. The saddle is therefore a 'hill' in these two dimensions and we can always find a lower energy path by walking around it.

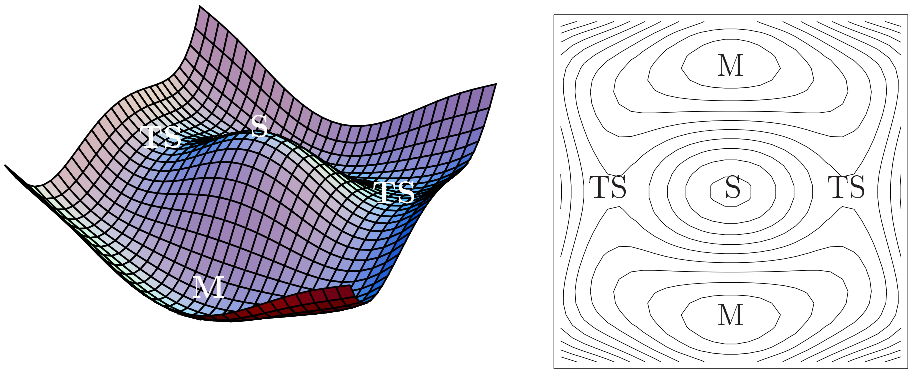


FIGURE 7: Surface and contour plots illustrating the Murrell-Laidler theorem. M, TS and S denote minima, transition states and the index 2 saddle, respectively.

Hence if there is a path connecting two minima via a saddle point of index > 1 then it is (almost) always possible to find a lower energy path via one or more true transition states. Rate theory is therefore usually concerned with passage over true transition states.

3.3 Pathways

We can define a steepest-descent path or gradient line, $\mathbf{X}(s)$, in terms of a parameter s such that $|d\mathbf{X}(s)/ds| = 1$. $\mathbf{X}(s)$ as a function of s defines a continuous path through nuclear configuration space. Furthermore,

$$\mathbf{X}(s_0 + ds) = \mathbf{X}(s_0) + \frac{d\mathbf{X}(s_0)}{ds} ds + \ldots \tag{41}$$

Hence, for $ds \to 0$ we have $\mathbf{X}(s_0 + ds) - \mathbf{X}(s_0) = (d\mathbf{X}(s_0)/ds)ds$ and the corresponding path length is $|d\mathbf{X}(s_0)/ds|ds$. Since we have defined s such that $|d\mathbf{X}(s_0)/ds| = 1$ the integrated path length between $\mathbf{X}(s_1)$ and $\mathbf{X}(s_2)$ is

$$\int_{s_1}^{s_2} \left| \frac{d\mathbf{X}(s)}{ds} \right| ds = s_2 - s_1. \tag{42}$$

The parameter s therefore corresponds to the path length.

For a steepest-descent path the tangent to the curve must be antiparallel to the gradient vector, and the constraint $|d\mathbf{X}(s_0)/ds| = 1$ defines the constant of proportionality:

$$\frac{d\mathbf{X}}{ds} = -\frac{\mathbf{G}(\mathbf{X})}{G}. \tag{43}$$

The minus sign ensures that the potential energy decreases with s.

3.4 Properties of Steepest-Descent Pathways

When a reaction path is defined in terms of the steepest-descent equation (43) a number of rigorous conditions can be deduced about its symmetry properties. These geometrical symmetry selection rules complement the orbital symmetry results of a Woodward-Hoffmann analysis. The geometrical rules must be rigorously obeyed by a reaction path, whereas ‘forbiddenness’ in the Woodward-Hoffmann sense implies a large but finite barrier, and can be overcome.

3.4.1 Uniqueness

If we assume that all the first and second derivatives of $V(\mathbf{X})$ exist and are continuous then every point in the nuclear configuration space, $\mathbf{X}$, lies on one and only one solution of (43). This result follows immediately from the uniqueness theorem of first order ordinary differential equations. Suppose we have two different paths, $\mathbf{X}(s)$ and $\mathbf{X}'(s)$, which coincide at some point s_0. Then at this point we have

$$\frac{d\mathbf{X}(s_0)}{ds} = -\frac{\mathbf{G}(\mathbf{X}(s_0))}{G(\mathbf{X}(s_0))} = \frac{d\mathbf{X}'(s_0)}{ds}. \tag{44}$$

Hence the curves are coincident:

$$\begin{aligned} \mathbf{X}(s_0 + ds) &= \mathbf{X}(s_0) + \frac{d\mathbf{X}(s_0)}{ds} ds = \mathbf{X}'(s_0) + \frac{d\mathbf{X}'(s_0)}{ds} ds \\ &= \mathbf{X}'(s_0 + ds). \end{aligned} \tag{45}$$

A steepest-descent path is therefore uniquely specified by a single point that lies on it. This result is not true at a stationary point where $G(\mathbf{X}) = 0$ and equation (43) becomes undefined. In general there are infinitely many solutions to (43) that terminate at any given minimum.

We will now prove that for a true transition state there are only two solutions to (43) with decreasing potential energy starting at $\mathbf{X}_{\mathrm{TS}}$. These paths set off in opposite directions along the eigenvector of $\mathbf{H}$ corresponding to negative curvature.

Consider an orthogonal transformation using the matrix $\mathbf{B}$ that satisfies

$$\sum_{\beta=1}^{3N} H_{\alpha\beta} B_{\beta\gamma} = \varepsilon_\gamma^2 B_{\alpha\gamma}. \tag{46}$$

$B_{\beta\gamma}$ with β running from 1 to $3N$ corresponds to the γ'th eigenvector of $\mathbf{H}$, just as $A_{\beta\gamma}$ corresponds to the γ'th eigenvector of $\mathbf{h}$. Now define new coordinates $W_\alpha = \sum_{\beta=1}^{3N} B_{\beta\alpha} X_\beta$. Since $\mathbf{B}$ is an orthogonal matrix $X_\alpha = \sum_{\beta=1}^{3N} B_{\alpha\beta} W_\beta$. The potential energy in the Taylor expansion about a stationary point at the origin becomes

$$\begin{aligned} \frac{1}{2} \sum_{\alpha,\beta=1}^{3N} H_{\alpha\beta}\, X_\alpha\, X_\beta &= \frac{1}{2} \sum_{\alpha,\beta,\gamma,\delta=1}^{3N} H_{\alpha\beta}\, B_{\alpha\gamma}\, B_{\beta\delta}\, W_\gamma\, W_\delta \qquad (47) \\ &= \frac{1}{2} \sum_{\alpha,\gamma,\delta=1}^{3N} B_{\alpha\gamma}\, \varepsilon_\delta^2\, B_{\alpha\delta}\, W_\gamma\, W_\delta = \frac{1}{2} \sum_{\delta=1}^{3N} \varepsilon_\delta^2\, W_\delta^2. \end{aligned}$$

$$\text{Since} \quad \frac{\partial}{\partial X_\alpha} = \sum_{\beta=1}^{3N} B_{\alpha\beta} \frac{\partial}{\partial W_\beta}, \tag{48}$$

the steepest-descent equation (43) at the stationary point in these new coordinates is equivalent to the $3N$ equations (problem 14):

$$\frac{dW_\alpha}{ds} = -\frac{\varepsilon_\alpha^2 W_\alpha}{G}. \tag{49}$$

In fact, the path $\mathbf{W}'(u)$ defined by

$$\frac{dW'_\alpha}{du} = -\varepsilon_\alpha^2 W'_\alpha, \tag{50}$$

without the factor of G, coincides with $\mathbf{W}(s)$ if they share a common point where $G \neq 0$. Then if $\mathbf{W}(s_0) = \mathbf{W}'(u_0)$ expanding to first-order in the small displacements ds and du gives

$$\begin{aligned} W_\alpha(s_0 + ds) &= W_\alpha(s_0) + \frac{dW_\alpha(s_0)}{ds}\,ds = W'_\alpha(u_0) + \frac{dW'_\alpha(u_0)}{du}\,\frac{ds}{G(s_0)} \\ &= W'_\alpha\left(u_0 + \frac{ds}{G(u_0)}\right) = W'_\alpha(u_0 + du). \end{aligned}$$

Hence the solutions close to the stationary point (at $\mathbf{W}'(u) = \mathbf{0}$) can be written in parametric form as

$$W'_\alpha(u) = W'_\alpha(0)\exp\left(-\varepsilon_\alpha^2 u\right). \tag{51}$$

Now suppose that the stationary point corresponding to $\mathbf{W}'(u) = \mathbf{0}$ is a transition state, with ε_1^2 the unique negative Hessian eigenvalue (or curvature). Since the steepest-descent paths are now parameterised by u, with $du = ds/G$, and G vanishes at a stationary point, we must consider $u \to \pm\infty$ to determine how the paths approach the point $\mathbf{W}'(u) = \mathbf{0}$. There are only two independent paths leading to the transition state as $u \to -\infty$; they have $W'_\alpha(u) = 0$ for $\alpha \neq 1$ and $W'_1(0)$ either positive or negative. An infinite number of steepest-descent paths approach the transition state as $u \to \infty$, all with $W'_1(u) = 0$; these paths lie in the hyperplane that is perpendicular to the coordinate W_1. The energy rises if we step off the transition state in any of these directions, because the change in potential energy is approximately

$$\frac{1}{2}\sum_{\alpha=2}^{3N} \varepsilon_\alpha^2 [dW'_\alpha(u)]^2 > 0, \tag{52}$$

for sufficiently small displacements $dW'_\alpha(u)$, where $\alpha > 1$. There are therefore only two distinct downhill steepest-descent paths leading from the transition state, and they begin parallel and antiparallel to the eigenvector corresponding to the unique negative Hessian eigenvalue.

In fact, further restrictions are needed to guarantee that there are only two steepest-descent paths leading down from a transition state. Zero Hessian eigenvalues other than those corresponding to overall translation and rotation can lead to all sorts of interesting surfaces. For example, a ‘monkey saddle’

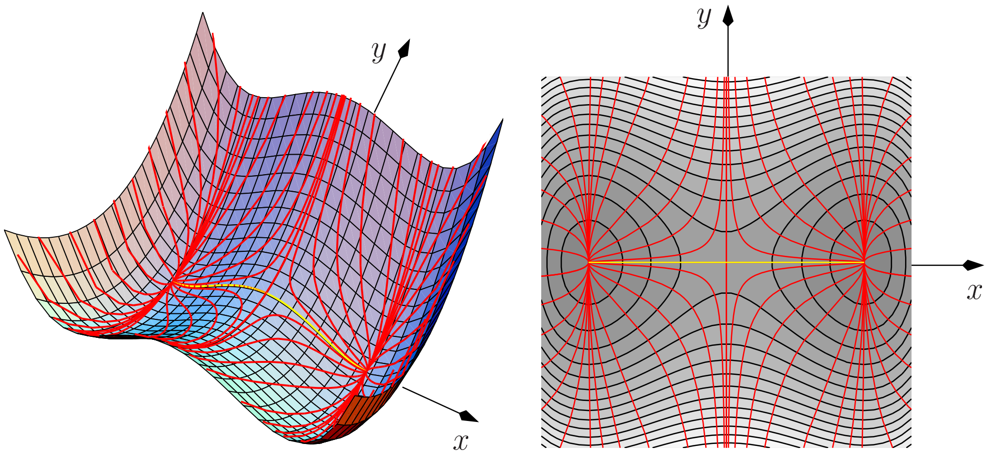


FIGURE 8: Surface and contour plots of the potential energy surface $V = x^4 - x^2 + y^2$. Steepest-descent paths are marked in red and the shading of the contour plot is darkest at the two minima. The two paths that lead downhill from the transition state are highlighted in yellow.

can occur for a cubic surface, as shown in Figure 9. (See problem 16.) However, real systems generally do not exhibit such exotic features.

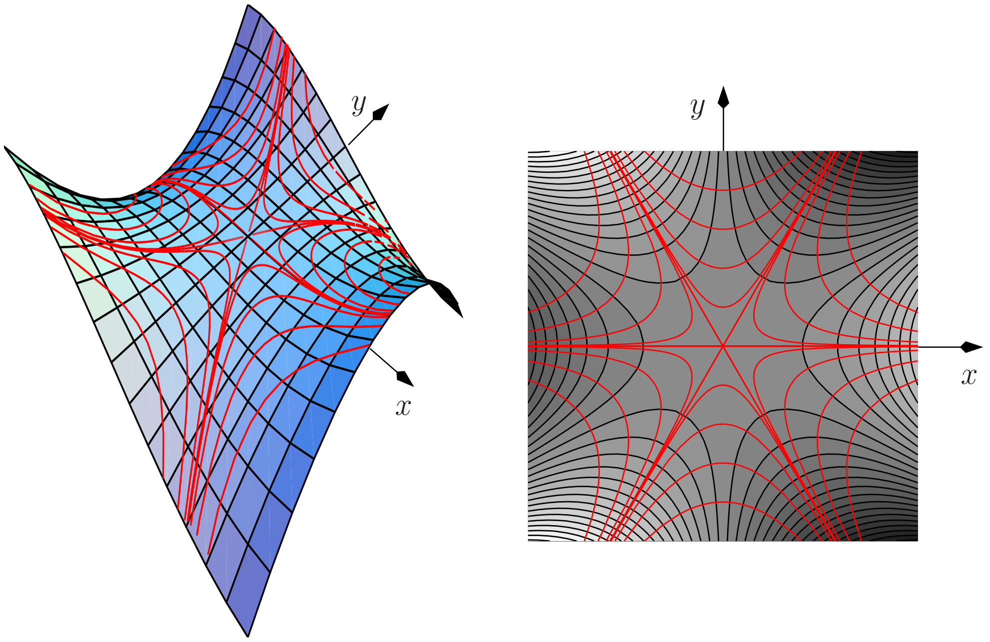


FIGURE 9: Surface and contour plots for the monkey saddle exhibited by the function $x^3 - 3xy^2$.

There is also nothing to prevent one or both the downhill steepest-descent paths starting at a transition state from terminating at a stationary point that is not a minimum. Such features are probably quite common; the addition of HF to ethene (Figure 10) is an example. A steepest-descent path links transition state 1 to the eclipsed form of fluoroethane, which is the transition state for internal rotation.

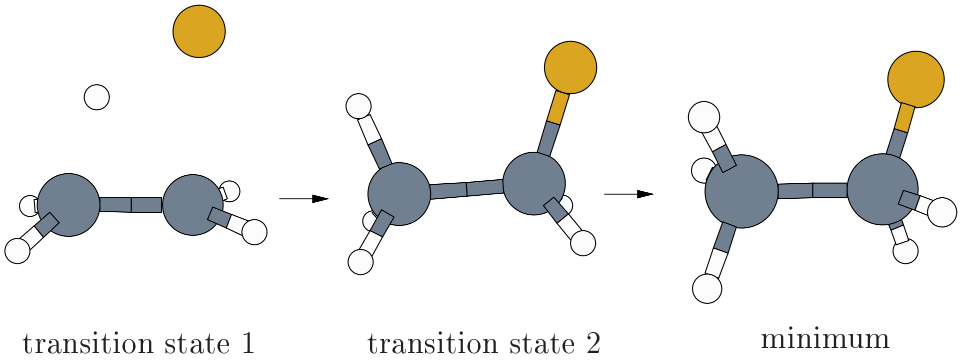


FIGURE 10: HF addition to C_2H_4.

3.4.2 Principal Directions

For any stationary point there are $2(3N-6)$ principal steepest-descent paths corresponding to positive and negative displacements along the $3N-6$ Hessian eigenvectors with non-zero eigenvalues. There is an infinite set of non-principal steepest-descent paths involving displacements along more than one principal direction. From equation (51) we see that

$$\frac{W'_\alpha(u)}{W'_\beta(u)} = \frac{W'_\alpha(0)}{W'_\beta(0)} \exp\left(-u(\varepsilon_\alpha^2 - \varepsilon_\beta^2)\right). \tag{53}$$

As $u \to \infty$ for a minimum all the components of $\mathbf{W}'$ tend to zero except for the one corresponding to the smallest curvature. Steepest-descent paths therefore approach a minimum parallel to the principal direction with the smallest curvature for which they have a non-zero projection. (See problem 15.)

3.4.3 Appearance and Disappearance of Symmetry Elements

Here we consider the effect of rotations and reflections, $\mathcal{R}$, on a nuclear configuration $\mathbf{X}^0$. The action of $\mathcal{R}$ mapping a general $3N$-dimensional vector

$\mathbf{X}$ onto $\mathbf{Y}$ may be defined by

$$Y_\alpha = C_{\alpha\beta} X_\beta \quad \text{and} \quad X_\alpha = C_{\beta\alpha} Y_\beta. \tag{54}$$

For symmetry operations of $\mathbf{X}^0$

$$\mathbf{C}\,\mathbf{X}^0 = \mathbf{X}^0, \tag{55}$$

so the symmetry operations of configuration $\mathbf{X}^0$ correspond to orthogonal matrices for which $\mathbf{X}^0$ is an eigenvector with unit eigenvalue. Note that $\mathbf{C}$ includes not only the effect of the point group operation considered as a mapping of nuclear coordinates[1], but also a possible permutation of nuclear labels. Using the chain rule gives

$$\begin{aligned} \frac{\partial V}{\partial Y_\alpha} &= \sum_{\beta=1}^{3N} C_{\alpha\beta} \frac{\partial V}{\partial X_\beta}, \quad \text{or} \quad \mathbf{G}(\mathbf{Y}) = \mathbf{C}\,\mathbf{G}(\mathbf{X}). && (56) \\ \frac{\partial^2 V}{\partial Y_\alpha \partial Y_\beta} &= \sum_{\gamma,\delta=1}^{3N} C_{\alpha\gamma} \frac{\partial^2 V}{\partial X_\gamma \partial X_\delta} C_{\beta\delta}, \quad \text{or} \quad \mathbf{H}(\mathbf{Y}) = \mathbf{C}\,\mathbf{H}(\mathbf{X})\,\mathbf{C}^T, \end{aligned}$$

If $\mathbf{C}$ corresponds to a symmetry operation at $\mathbf{X}$ then $\mathbf{X} = \mathbf{Y}$ and $\mathbf{C}\,\mathbf{G}(\mathbf{X}) = \mathbf{G}(\mathbf{X})$. Hence the gradient vector transforms as the totally symmetric irreducible representation of the prevailing point group. For a small displacement $\Delta\,\mathbf{G}(\mathbf{X})$ from $\mathbf{X}$

$$\begin{aligned} \mathbf{C}(\mathbf{X} + \Delta\,\mathbf{G}(\mathbf{X})) &= \mathbf{C}\,\mathbf{X} + \Delta\,\mathbf{C}\,\mathbf{G}(\mathbf{X}) \\ &= \mathbf{Y} + \Delta\,\mathbf{G}(\mathbf{Y}). \end{aligned} \tag{57}$$

Hence, if points $\mathbf{X}$ and $\mathbf{Y}$ are related by the operation $\mathbf{C}$ then so are the steepest-descent paths passing through $\mathbf{X}$ and $\mathbf{Y}$. Furthermore, if $\mathbf{C}$ is a symmetry operation of the point $\mathbf{X}$ then the corresponding symmetry element is conserved along the steepest-descent pathway passing through $\mathbf{X}$ because $\mathbf{C}(\mathbf{X} + \Delta\,\mathbf{G}(\mathbf{X})) = \mathbf{X} + \Delta\,\mathbf{G}(\mathbf{X})$. However, additional symmetry elements may appear when a steepest-descent path reaches a stationary point.

The above proof shows that symmetry operations are conserved for infinitesimal displacements along the gradient. Hence a stationary point must have every symmetry operation of the steepest-descent paths that converge to it. However, we cannot apply the theorem the other way round because the gradient vanishes at any stationary point. Hence, stationary points may possess additional symmetry elements that are not conserved along connecting steepest-descent paths.

---

[1] In the conventional view, point group operations are considered to act upon vibronic coordinates.

We proved in §3.4.1 that a transition state possesses precisely two downhill steepest-descent paths that correspond to plus/minus displacements along the unique Hessian eigenvector associated with negative curvature. Since there are only two such paths we deduce that a symmetry operation of the transition state must either leave both paths unchanged or interconvert them. Symmetry operations of the transition state are therefore either conserved along both downhill steepest-descent paths, or exchange them. New symmetry operations may therefore appear at a transition state if they interconvert the two downhill steepest-descent paths. A new symmetry operation of the transition state must therefore also interconvert the stationary points ('product' and 'reactant') at the two termini of the downhill steepest-descent paths. This situation can only arise if the pathways connect different permutation-inversion isomers of the same structure in a 'degenerate' rearrangement.

Let $\mathcal{R}$ be a symmetry operation of the transition state that interconverts the two downhill steepest-descent paths. Applying the operation an even number of times must leave the two paths unchanged, and any odd power of $\mathcal{R}$ must have the same effect as $\mathcal{R}$ itself. Hence all the even powers of $\mathcal{R}$ that correspond to distinct operations must be conserved along the entire path, while all odd powers of $\mathcal{R}$ correspond to operations that appear and disappear at the transition state.

In the degenerate rearrangement of an icosahedral Ar_{55} cluster via a cuboctahedron six five-fold axes are created at the two minima and three four-fold axes at the transition state. The conserved symmetry elements are those of the point group T_h (order 24), whilst the minima belong to point group I_h (order 120) and the transition state is O_h (order 48).

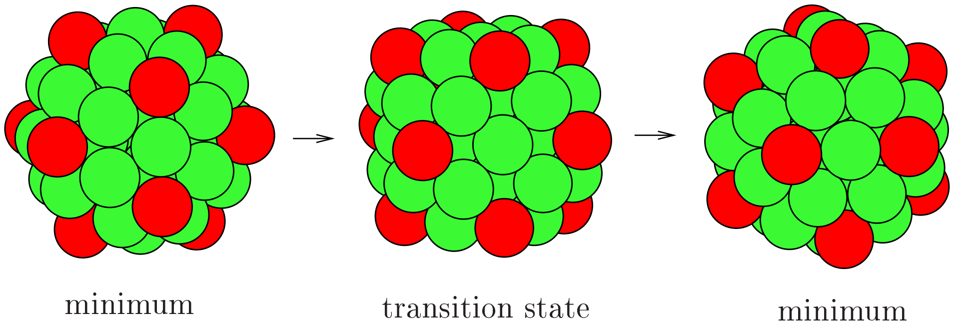


Figure 11: Symmetric degenerate rearrangement of an Ar_{55} cluster.

Clearly if a reaction pathway links different structures then no new symmetry elements can appear at the transition state. If the transition state does possess additional symmetry elements then it must have exactly twice as many symmetry operations as are conserved along the path, as illustrated for Ar_{55} in Figure 11. If g symmetry elements are conserved along the steepest descent paths then the point group of the transition state also contains precisely g such elements, which form a subgroup of $\mathcal{G}_{\rm ts}$, denoted $\mathcal{G}_{\rm ts}^{+}$. If $\mathcal{G}_{\rm ts}$ also contains an operation that interconverts the two paths, $\mathcal{R}$, then the product of $\mathcal{R}$ with any element conserved along the path must also interconvert the two sides. These operations must all be distinct and form a coset of the transition state point group denoted $\mathcal{R}\mathcal{G}_{\rm ts}^{+}$. Every element of $\mathcal{G}_{\rm ts}$ belongs to $\mathcal{G}_{\rm ts}^{+}$ or $\mathcal{R}\mathcal{G}_{\rm ts}^{+}$, and the two sets are interconverted by any member of $\mathcal{R}\mathcal{G}_{\rm ts}^{+}$. Hence the order of $\mathcal{G}_{\rm ts}$ is $2g$.

The above results can be used to discount any reaction pathway that violates the symmetry rules. (See problem 17.)

3.5 Classification of Rearrangements

We saw in §3.4.3 that the transition state for a degenerate rearrangement, which links permutational isomers of the same structure, can possess a symmetry element that maps one side of the path onto the other. This symmetry element is only present at the transition state.

Degenerate rearrangements are greatly outnumbered by non-degenerate pathways, but constitute an interesting minority. Transition state 2 in Figure 10 for the internal rotation of fluoroethane has the same point group as the minima it connects. However, the mirror plane in the transition state is not conserved along the path, and interconverts the corresponding minima. Degenerate rearrangements where the two sides of the path are related by a symmetry operation of the transition state are known as symmetric degenerate rearrangements. For HF dimer this symmetry is evident in the plot of energy versus path length shown in Figure 12.

Asymmetric degenerate rearrangements are also possible, where the two minima have the same structure but the steepest-descent paths are not related by symmetry. The energy profile is also asymmetric, and an example is shown for water trimer in Figure 13.

3.6 Branch points

The curvature along a steepest-descent path changes from negative to positive as we pass from the transition state to a minimum. Hence it must pass through zero somewhere along the path. The curvature associated with principal directions perpendicular to the steepest-descent path is normally

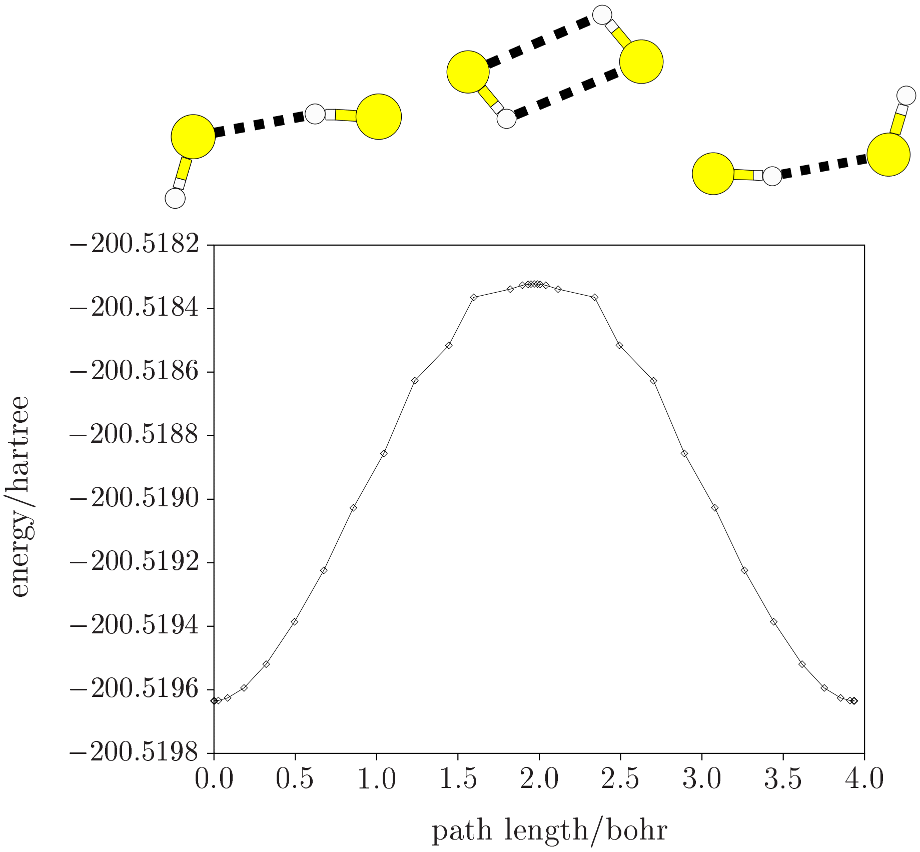


Figure 12: Energy profile for the symmetric degenerate rearrangement of HF dimer.

positive, but it can change from positive to negative if the path leads to another transition state. The point at which the curvature along an orthogonal principal direction passes through zero is called a branch point or valley-ridge inflection point (Figure 14).

Such features challenge our intuitive notion of a 'minimum energy path' (MEP). The steepest-descent path, as defined above, continues over the ridge because the gradient has no component to send the path down either side of the bifurcation. Intuitively, we might feel that the MEP should follow one of the bifurcations downhill to one side or the other. However, it is not possible to define such a path in terms of the solution of a differential equation, as we could for the steepest-descent path. Hence, if we define the 'minimum energy path' as a steepest-descent path then it must go over the ridge. To produce a definition more in keeping with intuition we would have to allow

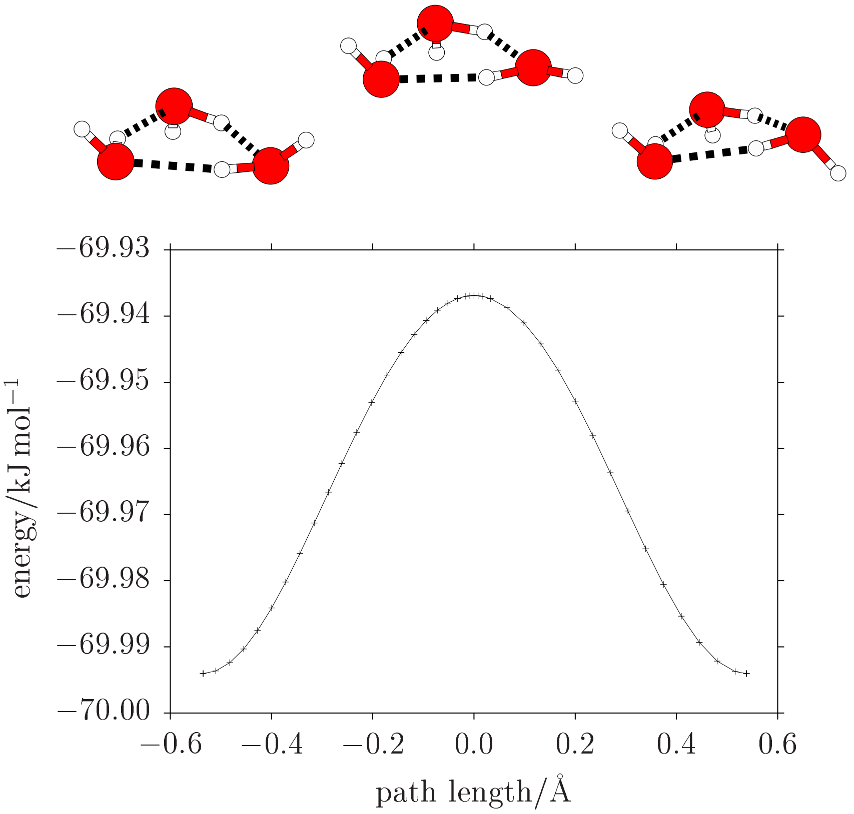


Figure 13: Energy profile for the asymmetric degenerate rearrangement of $(H_2O)_3$.

the path to bifurcate. If the system is infinitesimally perturbed from the true steepest-descent path then it will indeed follow one branch of the bifurcation downhill. Unfortunately there does not seem to be a rigorous definition of the reaction path that would actually follow such a trajectory.

A branch point must occur whenever a steepest-descent path connects two transition states: the curvature along the initial reaction coordinate must become positive, whilst the curvature correlating with the reaction coordinate of the new transition state must change from positive to negative.

A branch point must occur between the two transition states involved in the addition of HF to ethene (Figure 10). Another example is provided by rearrangements of FNH_2F^- (Figure 15). For this system two equivalent branch points exist between the C_{2v} transition state in the middle and the two transition states with C_s symmetry. C_s symmetry is preserved along the paths connecting these stationary points. The evolution of the lowest Hessian

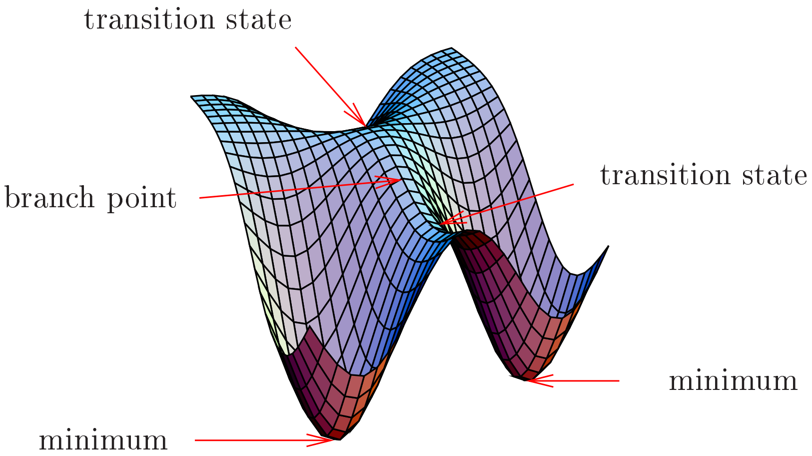


FIGURE 14: Schematic view of a branch point.

eigenvalues is shown for the $C_{2v} \to C_s$ path in Figure 16. The corresponding eigenvectors are S(ymmetric) and A(ntisymmetric) with respect to the conserved mirror plane. The motion perpendicular to the steepest-descent path at a branch point is usually associated with symmetry-breaking.

For the above path the perpendicular direction whose curvature passes through zero at the branch point corresponds to loss of the conserved mirror plane. Notice also that the C_{2v} transition state has an additional symmetry element, namely a C_2 rotation axis.

Properties of a potential energy surface such as steepest-descent paths and principal directions should be coordinate independent. To ensure this one must use covariant derivatives and include appropriate metric tensors It can also be shown that the number of positive, negative and zero eigenvalues of the Hessian and mass-weighted Hessian are the same.

4 Tunnelling

Tunnelling is the quantum mechanical situation where systems have a finite probability of being found in classically forbidden regions. The simplest example is a particle with total energy E_1 tunnelling through a barrier of height $E_2 > E_1$. The wavefunction in the barrier region is a decaying exponential and the transmission probability is largest when the energy difference and particle mass are small and the barrier is narrow. (See problem 18.) Hence tunnelling is only likely to give rise to observable effects for light atoms and

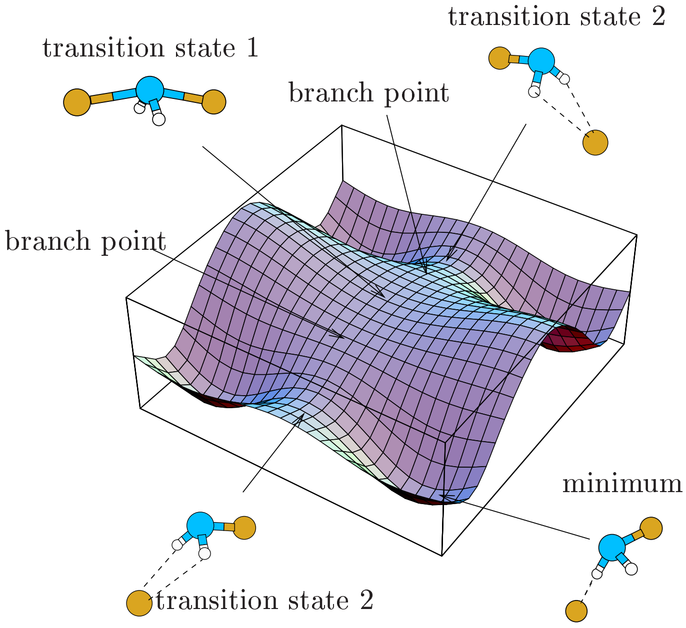


FIGURE 15: Schematic potential energy surface for FNH_2F^-.

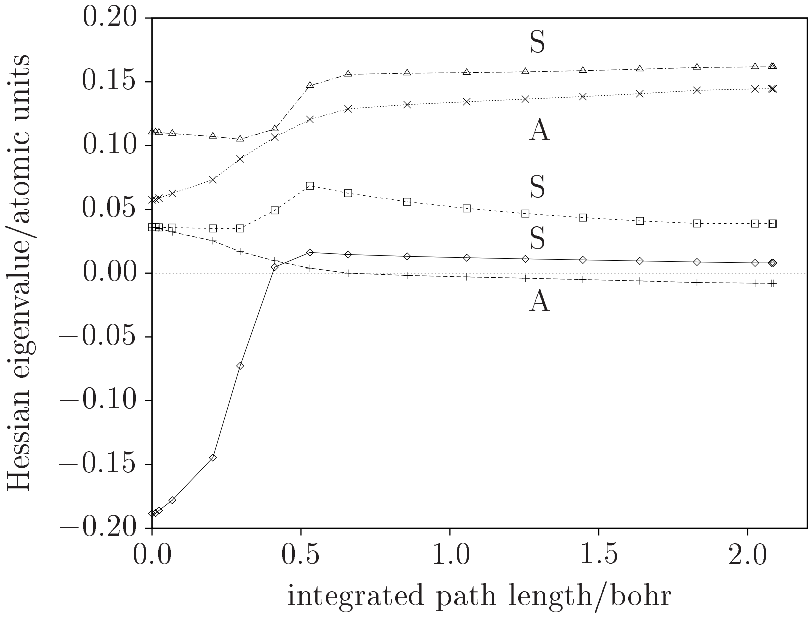


FIGURE 16: Evolution of the lowest eigenvalues for FNH_2F^- along the path connecting the C_{2v} transition state (left) to a C_s transition state (right).

short pathways with small barriers.

The classic example of tunnelling manifesting itself in spectroscopy is the inversion doubling of the ammonia molecule. The umbrella motion takes the system through the planar D_{3h} geometry from one C_{3v} minimum to a permutational isomer (Figure 17). If the two potential wells were isolated they would support identical rovibronic energy levels. However, the barrier is sufficiently small for the localised wavefunctions in the two wells to interfere. The correct wavefunctions are therefore the sum and difference of the separate wavefunctions for the two wells and are delocalized.

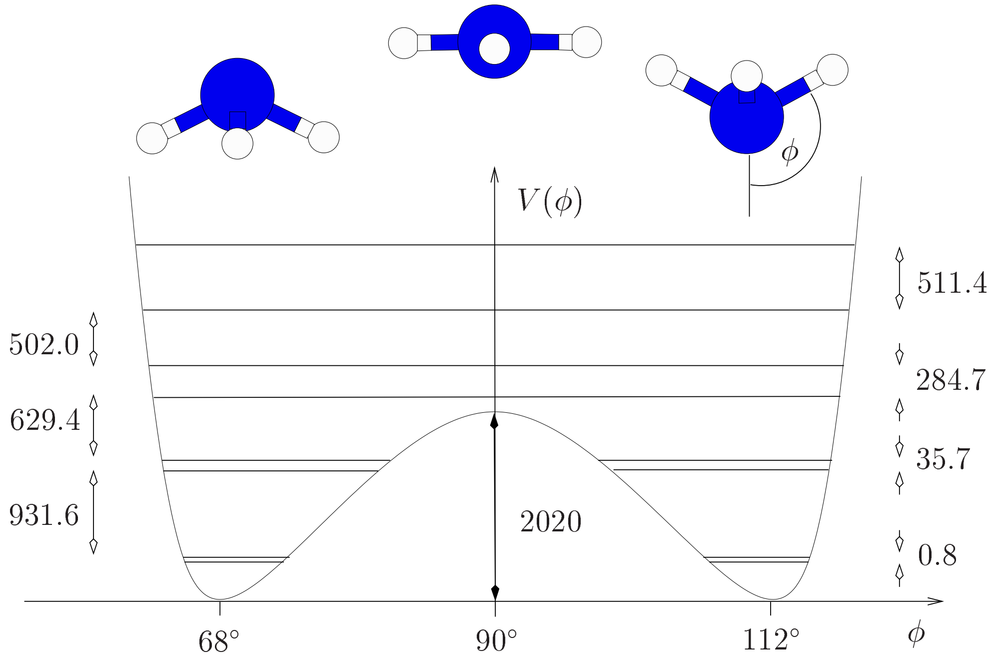


FIGURE 17: Schematic representation of the tunnelling splittings (right) and spacings (left) in the lowest four states of the ν_2 bending vibration in NH_3. The energies are in wavenumbers.

Compare the interaction of atomic orbitals to give molecular orbitals. If two hydrogen atoms are far apart then we can treat each one as isolated but as they approach the $1s$ orbitals interact. The molecular orbitals are proportional to the sum and difference of the $1s$ functions, and the energy levels become $\alpha \pm \beta$ in the Hückel approximation, where α is the isolated energy level and β is the off-diagonal Hamiltonian matrix element between the two $1s$ orbitals. We can treat the double-well problem for ammonia in just the same way, forming a delocalized linear combination of localised wavefunctions, analogous to a linear combination of atomic orbitals. If we denote the localised wavefunctions by ψ_1 and ψ_2 and the matrix elements of the Hamiltonian by $\alpha = \langle\psi_1|\hat{\mathcal{H}}|\psi_1\rangle = \langle\psi_2|\hat{\mathcal{H}}|\psi_2\rangle$ and $\beta = \langle\psi_1|\hat{\mathcal{H}}|\psi_2\rangle = \langle\psi_2|\hat{\mathcal{H}}|\psi_1\rangle$ then the resulting secular equations are entirely analogous to the hydrogen molecule problem:

$$\begin{vmatrix} \alpha - E & \beta \\ \beta & \alpha - E \end{vmatrix} = 0, \tag{58}$$

with solutions $E_\pm = \alpha \pm \beta$. The splitting is therefore 2β, where β is a tunnelling matrix element, and the optimal wavefunctions are $\psi_1 \pm \psi_2$, in agreement with symmetry arguments. The Hückel-type approximation may be justified for tunnelling problems because of the exponential decay of the wavefunctions in the barrier regions.

For NH_3 the localised levels are split into pairs—hence the term inversion doubling. The splitting of the lowest level is $0.793\,\mathrm{cm}^{-1}$, and the splitting increases for higher energy states because the barrier is narrower.

Tunnelling between minima with different structures (non-degenerate tunnelling) is generally assumed to lead to much smaller effects than tunnelling between equivalent minima. The reason is basically that only the energy levels of permutational isomers of the same structure are in resonance.

4.1 Tunnelling in $(HF)_2$

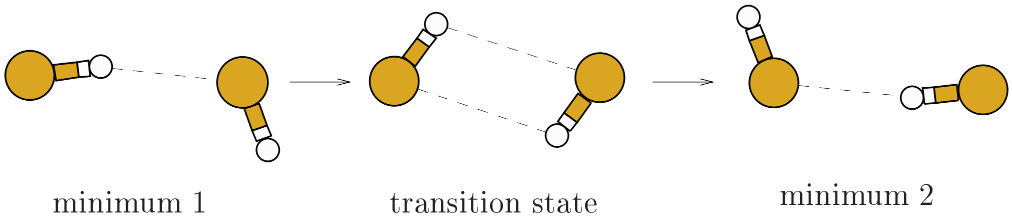


Figure 18: Degenerate rearrangement of the HF dimer.

The symmetric degenerate rearrangement of the $(HF)_2$ C_s minima via a C_{2h} transition state is shown in Figure 18. Note the appearance and disappearance of the C_2 symmetry element at the transition state. The ground vibrational state of the dimer is subject to a tunnelling splitting of $0.66\,\mathrm{cm}^{-1}$. However, this splitting decreases to 0.22 and $0.23\,\mathrm{cm}^{-1}$ for the excited vibrational states corresponding to one quantum of vibration in the non-H-bonded and H-bonded HF monomers. The dependence of the splitting of low-frequency 'floppy' modes upon excitation of high-frequency modes can often only be explained by detailed calculation.

4.2 Tunnelling in $(H_2O)_3$

New experiments in the far-infrared region have recently become possible that can probe the low frequency modes of a Van der Waals complex and resolve rotational transitions. Results for the water trimer, $(H_2O)_3$, have generated considerable theoretical activity in the last few years, and the spectra have now been assigned in some detail. The lowest energy rearrangement mechanism of the cyclic global minimum is known as the 'flip' (Figure 19). The global minimum is a 'frustrated' structure because two 'dangling' hydrogen atoms must be next to each other. The flip is an asymmetric degenerate rearrangement. Each minimum can rearrange via two distinct flips, corresponding to either of the adjacent 'dangling' hydrogens. This process is facile, as no hydrogen bonds are made or broken, and is localised at a single hydrogen atom. The mechanism leads to a low effective mass, and, in combination with the low barrier, produces large tunnelling splittings on the order of $10\,\mathrm{cm}^{-1}$.

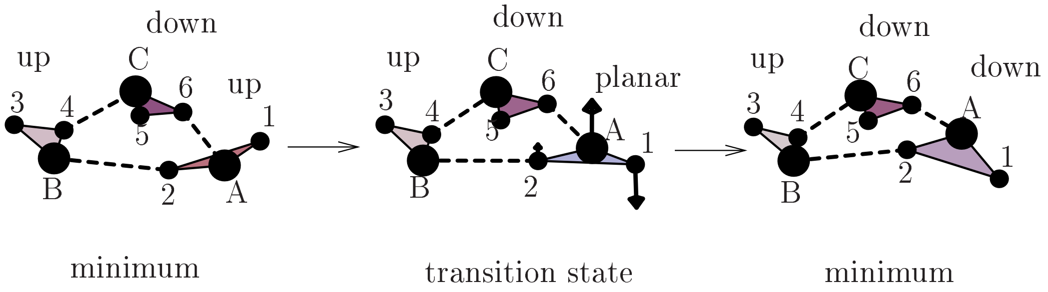


FIGURE 19: Single flip rearrangement of $(H_2O)_3$.

It is not hard to show that three successive flips connect any permutational isomer to its enantiomer, while six flips can take us through a closed set of six such isomers. The secular problem for the energy levels is analogous to the treatment of the π system of benzene in the Hückel approximation, with tunnelling levels at $\alpha \pm 2\beta$ (non-degenerate) and $\alpha \pm \beta$ (doubly degenerate), in good agreement with experiment.

5 Searching the Landscape

Finding local minima is usually straightforward. There are several efficient standard techniques available that require gradient information. Second derivatives are useful for transition state searches where we must locate a stationary point that is a local maximum in one principal direction but a local minimum in the others.

Consider again the Taylor expansion of the potential energy around a general point in nuclear configuration space, $\mathbf{X}$, truncated at second order:

$$V(\mathbf{X}+\mathbf{x}) = V(\mathbf{X}) + \mathbf{G}(\mathbf{X})^T\,\mathbf{x} + \frac{1}{2}\mathbf{x}^T\,\mathbf{H}(\mathbf{X})\,\mathbf{x}, \tag{59}$$

where $\mathbf{G}(\mathbf{X})$ and $\mathbf{H}(\mathbf{X})$ are the gradient and Hessian, and $\mathbf{x}$ is a vector of small nuclear displacements. Applying the condition $dV(\mathbf{X}+\mathbf{x})/d\mathbf{x} = \mathbf{0}$ leads us to the standard Newton-Raphson step (see problem 21):

$$\mathbf{x}_{\rm NR} = -\mathbf{H}^{-1}\mathbf{G}. \tag{60}$$

However, $\mathbf{x}_{\rm NR}$ is not useful in this form because the inverse Hessian of a free molecule is undefined due to zero eigenvalues (§3.1). To get around these difficulties we use internal coordinates or eigenvalue shifting. Since each eigenvector, $\mathbf{e}$, corresponding to overall translation or rotation is known (§3.1) the corresponding eigenvalues can be shifted arbitrarily by adding a multiple of $\hat{e}_\alpha\hat{e}_\beta$ to $H_{\alpha\beta}$ (problem 20).

Greater insight can be obtained if we express everything in terms of the coordinates, W_α, and the eigenvalues of $\mathbf{H}$, ε_α^2, which we defined previously in equation (47). In the new coordinates the Newton-Raphson step and the corresponding energy change are (problem 21)

$$x_{{\rm NR},\alpha} = -g_\alpha(\mathbf{W})/\varepsilon_\alpha^2 \quad \text{and} \quad \Delta V_{\rm NR} = -\sum_{\alpha=1}^{3N} g_\alpha(\mathbf{W})^2/2\varepsilon_\alpha^2, \tag{61}$$

where $g_\alpha(\mathbf{W}) = \partial v(\mathbf{W})/\partial W_\alpha$, and lower case letters are used for functions of the W_α coordinates. Hence contributions from terms with $\varepsilon_\alpha^2 > 0$ and $\varepsilon_\alpha^2 < 0$ lower and raise the energy, respectively. To find a transition state we may need to start at a point where precisely one of the ε_α^2 is negative.

A Newton-Raphson search can converge to a stationary point of arbitrary Hessian index. However, with only a little extra effort it is possible to obtain an algorithm that will systematically converge to a stationary point of any required index. Introducing an additional Lagrange multiplier gives increased flexibility that we can exploit to find transition states systematically. This method is known as eigenvector-following. Consider the Lagrangian:

$$L = -v(\mathbf{W}) - \sum_{\alpha=1}^{3N} \left[g_\alpha(\mathbf{W})\, x_\alpha + \frac{1}{2}\varepsilon_\alpha^2\, x_\alpha^2 - \frac{1}{2}\mu_\alpha(x_\alpha^2 - c_\alpha^2) \right]. \tag{62}$$

The step that is optimal in all directions is now

$$x_\alpha = g_\alpha(\mathbf{W})/(\mu_\alpha - \varepsilon_\alpha^2), \tag{63}$$

and the energy change corresponding to this step is

$$\Delta V = \sum_{\alpha=1}^{3N} (\mu_\alpha - \varepsilon_\alpha^2/2) g_\alpha(\mathbf{W})^2/(\mu_\alpha - \varepsilon_\alpha^2)^2. \tag{64}$$

Now we must make a choice for μ_α. We require $\mu_\alpha \to 0$ as $g_\alpha(\mathbf{W}) \to 0$, so that the Newton-Raphson step is recovered. We also require $\mu_\alpha - \varepsilon_\alpha^2/2 < 0$ for minimisation and $\mu_\alpha - \varepsilon_\alpha^2/2 > 0$ for maximisation. Good choices for μ_α have been found through a combination of theory and experiment. For example

$$\mu_\alpha = \varepsilon_\alpha^2 \pm \frac{1}{2}|\varepsilon_\alpha^2| \left(1 + \sqrt{1 + 4g_\alpha(\mathbf{W})^2/\varepsilon_\alpha^4}\right), \tag{65}$$

plus for maximisation, minus for minimisation, gives steps that obey the required conditions and are an even function of ε_α^2:

$$x_\alpha = \frac{\pm 2 g_\alpha(\mathbf{W})}{|\varepsilon_\alpha^2|(1 + \sqrt{1 + 4g_\alpha(\mathbf{W})^2/\varepsilon_\alpha^4})}. \tag{66}$$

A graphical comparison of Newton-Raphson and eigenvector-following results is shown in Figures 20 and 21. Every pixel in each of the three plots corresponds to a particular geometry in a cut through the three-dimensional configuration space of a three-atom cluster. The colour depends upon which stationary point the algorithm in question converges to and the resolution is 700 × 700. The Newton-Raphson algorithm can converge to both minima and transition states, depending upon where we start from (Figure 20). The eigenvector-following method converges only to stationary points of the specified index, i.e. minima for the left panel and transition states for the right panel of Figure 21.

Calculating second derivatives may represent a serious expense for some systems, especially in quantum mechanical treatments of the bonding and

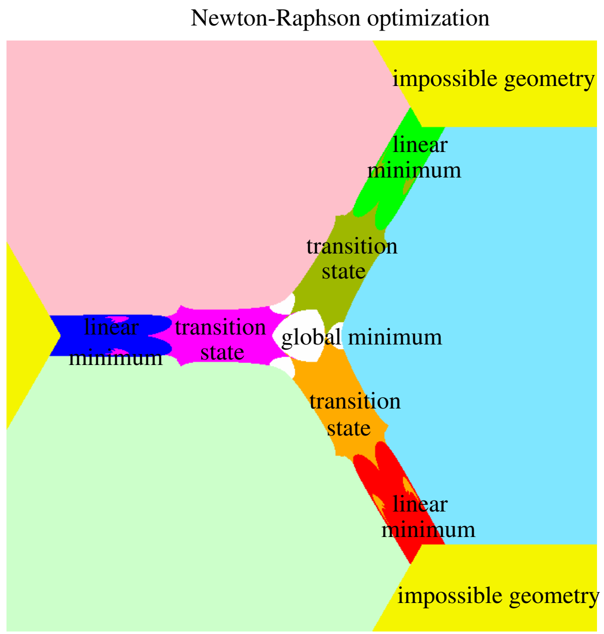
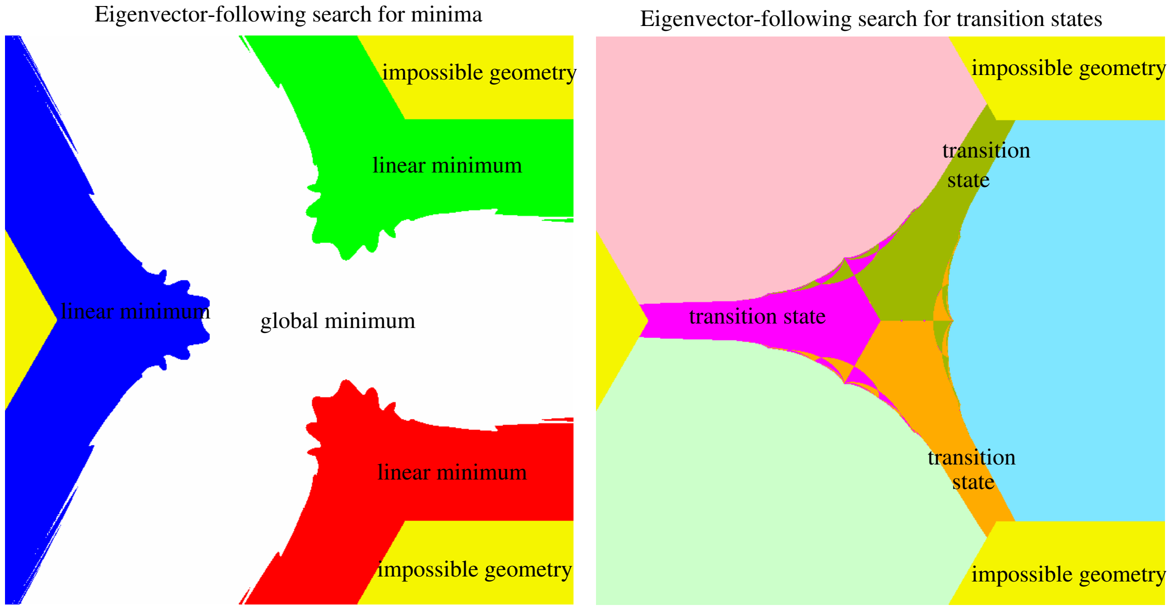

Newton-Raphson optimization

FIGURE 20: Basins of attraction for Newton-Raphson optimisation.

even for some empirical potentials. Some transition states can be found by energy minimisation if they possess symmetry elements that are not con-


Eigenvector-following search for minima

Eigenvector-following search for transition states

FIGURE 21: Basins of attraction for eigenvector-following optimisation.

served along the downhill steepest-descent paths. In this case if we constrain the system to higher symmetry the gradient vector can never have a non-zero component in a direction that would lower the point group (see problem 19.) For most transition states such tricks are not possible. An alternative approach is to use numerical techniques to update either the Hessian or inverse Hessian matrix using only gradient information. The approximate matrix thus obtained is then used in place of the real one. Techniques that use this approach are also referred to as variable metric or quasi-Newton approaches.

New hybrid eigenvector-following schemes avoid diagonalisation of the Hessian by combining the uphill step with minimisation in the tangent space. If the Hessian is unavailable then a variational method can be used to determine the uphill direction using only energies and gradients. The latter approach has been used in *ab initio* calculations for systems containing hundreds of atoms.

6 Describing the Potential Energy Landscape

6.1 Introduction

For an N-atom system the potential energy is a $3N$-dimensional function. To refer to a potential energy hypersurface we must embed the function in a $3N+1$ dimensional space where the extra dimension corresponds to the 'height' of the surface. The main problem with treating many-dimensional potential energy surfaces is that the number of stationary points grows rapidly with size. For a system composed of N_A atoms of type A, N_B atoms of type B, etc., the Hamiltonian is invariant to all permutations of equivalent nuclei and to inversion of all coordinates through a space-fixed origin. The number of permutation-inversion isomers of any given configuration could therefore be as large as

$$2 \times N_A! \times N_B! \times N_C! \times \cdots. \tag{67}$$

But if the system possesses any point group symmetry elements other than the identity it is reduced by a factor of h, where h is the order of the point group, i.e. the number of 'permutational isomers' is therefore $2N_A!N_B!N_C!\cdots/h$. The ratio of the number of permutational isomers of any two configurations of the same molecule, A and B, is therefore $n_{\mathrm{A}}/n_{\mathrm{B}} = h_{\mathrm{B}}/h_{\mathrm{A}}$.

There are no strict rules for the growth in the number of stationary points corresponding to different structures with system size. The rate of growth is certainly system dependent, but empirical observations and simple theories suggest an exponential increase in the number of local minimum structures,

$n_{\text{min}}(N)$, as a function of N (Figure 22). Multiplying by the permutational factor therefore suggests that the total number of minima for a homonuclear system is likely to increase as

$$2N!\exp(\alpha N). \tag{68}$$

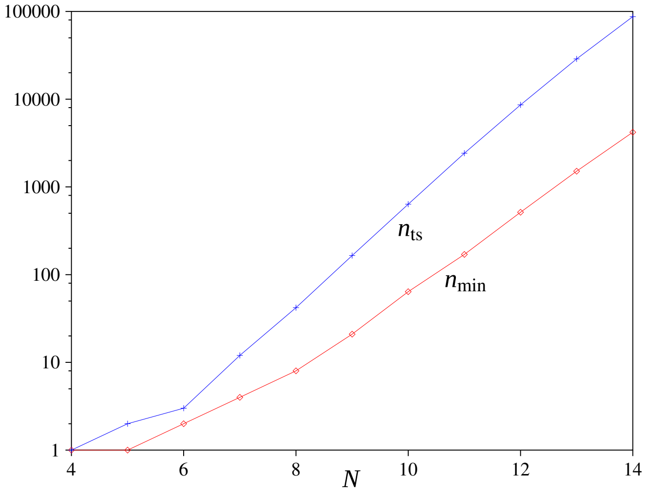


FIGURE 22: The number of minima and transition states for LJ_N clusters as a function of N.

A simple justification for the form of $n_{\text{min}}(N)$ can be given. If the system is large enough so that it can be divided into m equivalent sub-systems of N atoms each, and the stable configurations of each subsystem can be considered independent, then

$$n_{\text{min}}(mN) = n_{\text{min}}(N)^m. \tag{69}$$

The solution of this equation is

$$n_{\text{min}}(N) = \exp(\alpha N), \tag{70}$$

where α is a system-dependent constant. A similar argument can be given for the number of transition states. If the rearrangement corresponding to each transition state is localised in one sub-system, then a transition state of the complete mN-atom system occurs when one of the subsystems is at a transition state and the rest are at a minimum. This argument implies that

$$n_{\text{ts}}(mN) = m\,n_{\text{min}}(N)^{m-1}n_{\text{ts}}(N), \tag{71}$$

and the solution to this equation is

$$n_{\text{ts}}(N) = N \exp(\alpha N). \tag{72}$$

These results therefore indicate that the ratio $n_{\text{ts}}/n_{\text{min}}$ grows linearly with size, in fair agreement with the numerical results for small Lennard-Jones clusters.

6.2 Example: Morse Clusters

The Morse potential can be written in the form

$$V = \epsilon \sum_{s<t} e^{\rho(1-r_{st}/r_{\text{e}})} \left[e^{\rho(1-r_{st}/r_{\text{e}})} - 2 \right], \tag{73}$$

where r_{st} is the distance between atoms s and t. ϵ and r_{e} are the dimer well depth and equilibrium bond length. They can conveniently be set to unity and used as the units of energy and distance.

ρ is a dimensionless parameter that determines the range of the interparticle forces, with low values corresponding to long range (Figure 23). Physically meaningful values vary from $\rho \sim 3$ for sodium to around 14 for C_{60} molecules. When $\rho = 6$, the Morse potential has the same curvature as the Lennard-Jones potential at the minimum.

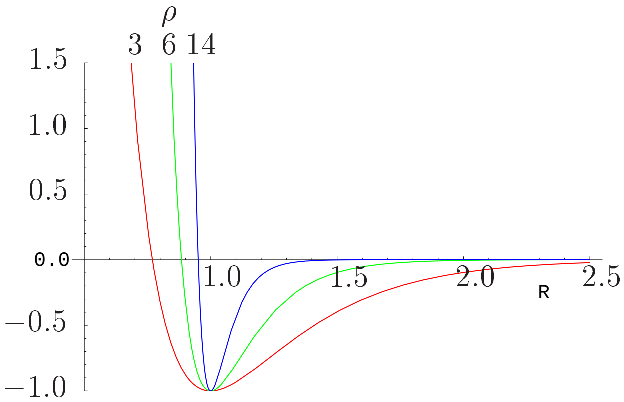


Figure 23: The Morse potential for $\rho = 3$, $\rho = 6$ and $\rho = 14$.

13-atom Morse clusters, denoted M_{13}, provide a benchmark system because they are small enough for exhaustive searches of minima and transition states to be possible (Table 1). By changing the range of the potential we can tune the surface from smooth (few minima, small ρ) to rough (many minima, large ρ) (Figure 24).

Table 1: Numbers of minima, $n_{\min}$, and transition states, n_{ts}, for M_{13} clusters at four values of the range parameter ρ.

ρ	3	4	6	10
$n_{\min}$	7	159	1 477	10 814
n_{ts}	47	1 366	26 431	$> 550\,000$
$n_{\mathrm{ts}}/n_{\min}$	6.7	8.6	17.9	> 50.8

Some global properties of the M_{13} energy landscape are listed in Table 2 as a function of the range parameter, ρ. The number of minima not lying on downhill sequences terminating at the global minimum increases with ρ and the gap between the two lowest-lying minima decreases.

The wells become narrower and the paths over the transition states become steeper as ρ increases. The average uphill and downhill barriers decrease and increase with ρ, respectively, and the mean energy difference between connected minima also decreases with ρ. The barriers for relaxation to the global minimum therefore increase with ρ, and the surface becomes flatter, yet rougher.

As the extent of configuration space is similar irrespective of ρ, the average path length to the global minimum does not vary much. However, as the number of minima increases rapidly with ρ, the minima are therefore closer together, and the number of steps on the shortest path to the global minimum increases. The pathways become rougher.

These trends are illustrated by the pathways depicted in Figure 25. For $\rho = 4$ the landscape of M_{13} is a perfect funnel with all minima lying on downhill sequences terminating at the global minimum. As the landscape

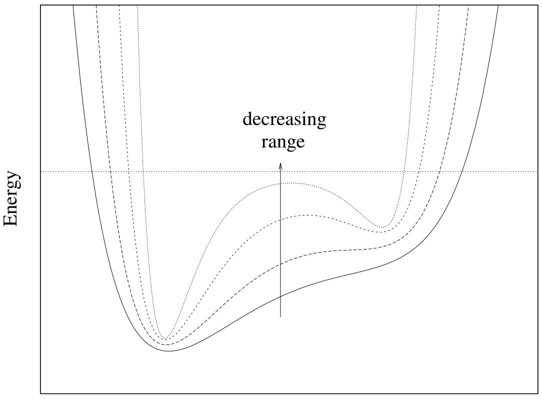


Figure 24: The ‘swallowing up’ of higher energy minima as the range increases.

TABLE 2: Global analysis of M_{13} energy landscapes as a function of ρ. $\bar{\nu}$ is the geometric mean normal mode frequency of a minimum, and $\nu_{\rm im}$ is the imaginary frequency of a transition state. $b^{\rm up}$ and $b^{\rm down}$ are the uphill and downhill barriers associated with a transition state. $\Delta E^{\rm con}$ and S are the energy difference and path length between connected minima, respectively. $n^{\rm gm}$ and $S^{\rm gm}$ are the number of steps and the path length for the shortest path from a minimum to the global minimum.

ρ	4	6	10	14
$\Delta E_{\rm gap}$	3.024	2.864	2.245	0.468
$\langle\bar{\nu}\rangle$	1.187	1.625	2.615	3.660
$\langle\|\nu^{\rm im}\|\rangle$	0.396	0.473	0.637	0.628
$\langle b^{\rm up}\rangle$	3.666	2.070	1.470	1.536
$\langle b^{\rm down}\rangle$	0.461	0.543	0.583	0.784
$\langle\Delta E^{\rm con}\rangle$	3.205	1.526	0.887	0.752
$\langle S\rangle$	2.471	1.735	1.030	0.971
$\langle n^{\rm gm}\rangle$	1.354	2.447	3.744	3.885
$\langle S^{\rm gm}\rangle$	2.772	3.534	3.573	3.357

becomes rougher with increasing ρ, relaxation to the global minimum will become slower.

The general increase in energy with ρ, decreasing energy gap, increasing barrier heights, shorter pathways and decreasing energy bias towards the global minimum are all visible.

6.3 Disconnectivity Graphs

One helpful way of visualising a complex energy landscape is to use topological mapping to construct a disconnectivity graph. The analysis begins by mapping every point in configuration space on to the local minimum reached by following the steepest-descent path. Thus, configuration space is represented by the discrete set of minima, each of which has an associated 'catchment basin' of configuration space.

At a given total energy, E, the minima can be grouped into disjoint sets, called basins, whose members are mutually accessible at that energy. In other words, each pair of minima in a basin are connected directly or through other minima by a path whose energy never exceeds E, but would require more energy to reach a minimum in another basin. At low energy there is just one basin—that containing the global minimum. At successively higher energies, more basins come into play as new minima are reached. At still

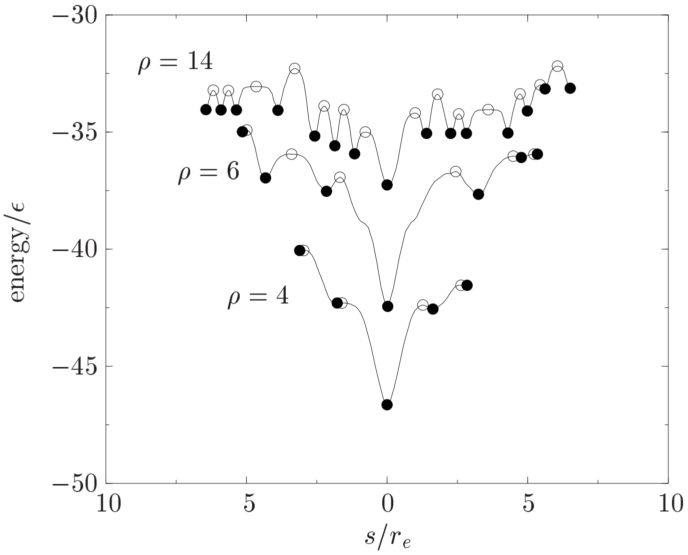


FIGURE 25: Typical pathways to the global minimum for M_{13} as a function of ρ. Minima are indicated by filled circles and transition states by open circles.

higher energies, the basins coalesce as higher barriers are overcome, until finally there is just one basin containing all the minima (provided there are no infinite barriers).

The significance of dividing the landscape in this way is that it provides a picture of the barriers between different parts of configuration space, and hence the dynamics.

The disconnectivity graph is constructed by performing the basin analysis at a series of energies, plotted on a vertical scale. At each energy, a basin is represented by a node, with lines joining nodes in one level to their daughter nodes in the level below. The choice of the energy levels is important; too wide a spacing and no topological information is left, whilst too close a spacing produces a vertex for every transition state and hides the longer range structure of the landscape. The horizontal position of the nodes is arbitrary, and can be chosen for clarity.

In the resulting graph, all branches terminate at local minima, while all minima connected directly or indirectly to a node are mutually accessible at the corresponding energy. Graphs such as these, which are connected but contain no cycles, are known as 'trees' for reasons which should be clear from Figure 26. In part (a) we see a gently sloping funnel with high barriers resembling a 'weeping willow', in part (b) a steeper funnel with lower barriers resembles a 'palm tree', and in part (c) the 'rough' landscape produces a 'banyan tree'.

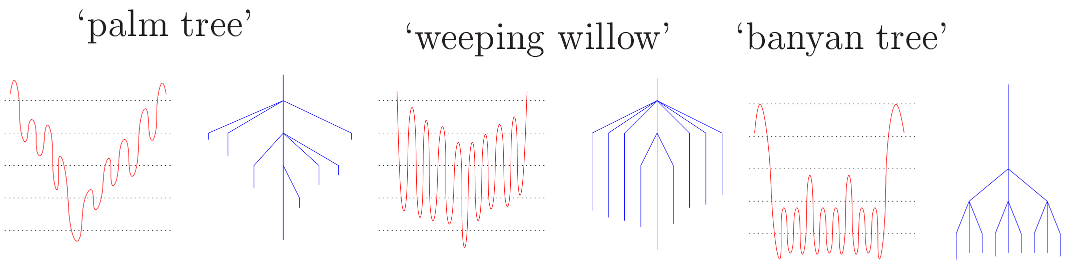


Figure 26: Disconnectivity graphs. Dotted lines indicate energies where the basin analysis was performed.

When landscapes possess more than one funnel, inter-funnel motion is likely to occur on a slower time scale than inter-well flow within a funnel, so funnels constitute the next level in a hierarchy of structure above minima.

The disconnectivity graphs for M_{13} with $\rho = 4$ and 6 are plotted on the same scale in Figure 27 with a linear energy spacing of one well depth. Both ‘palm’ trees are typical of a funnel-like landscape: as the energy is lowered, minima are cut off a few at a time with no secondary funnels, which would appear as side branches.

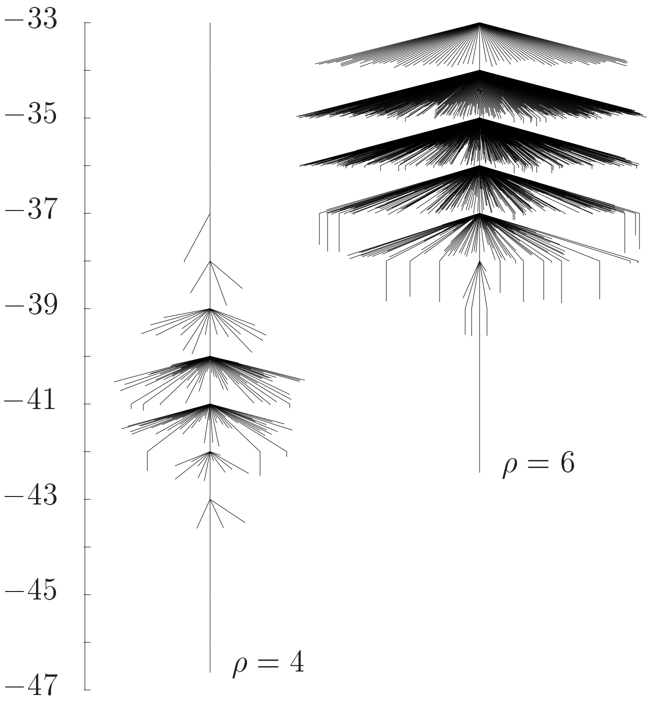


Figure 27: Disconnectivity trees for M_{13} with $\rho = 4$ and $\rho = 6$ plotted on the same energy scale (in units of the pair well depth).

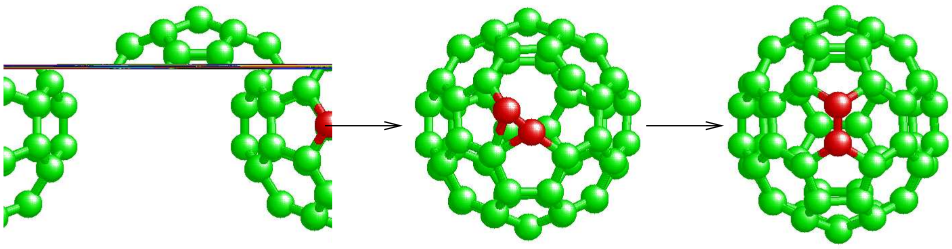


Figure 28: SW rearrangement of C_{60}.

6.3.1 Annealing of C_{60}

A number of studies have proposed mechanistic schemes to explain the formation of buckminsterfullerene (BF), but simulations of fullerene nucleation usually produce defective structures. For potentials that support realistic barriers the time scale for fullerene annealing lies beyond the reach of traditional simulation techniques. The calculations below employed a semi-empirical, quantum-mechanical tight-binding potential. The disconnectivity graph for C_{60} has been constructed using the most likely annealing mechanism, assuming that a C_{60} molecule with 12 pentagonal and 20 hexagonal

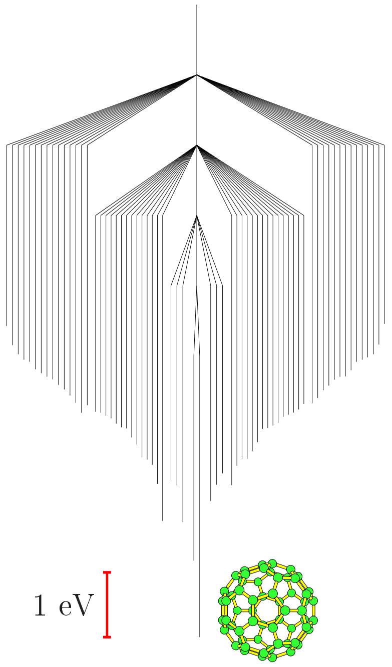


Figure 29: Disconnectivity graph for C_{60}.

faces has already been reached.

Excluding enantiomers there are 1812 different isomers for C_{60} with 12 pentagonal and 20 hexagonal faces. The 'pyracylene' or 'Stone-Wales' (SW) mechanism provides a means for pentagonal and hexagonal rings to migrate over the surface of the cluster (Figure 28). The largest subset of the 1812 C_{60} isomers that are mutually accessible via SW rearrangements contains 1710 minima, one of which is BF. In fact, the minima can be grouped into 'stacks' according to how many SW processes are required to convert them into BF. The first SW stack contains BF itself and the second stack contains the adjacent minimum illustrated in Fig. 28. The resulting tree is shown in Figure 29 and clearly corresponds to the 'weeping willow' pattern, i.e. a gentle funnel with relatively high barriers. The long branches correspond to high barriers, explaining why high temperatures are needed to produce a

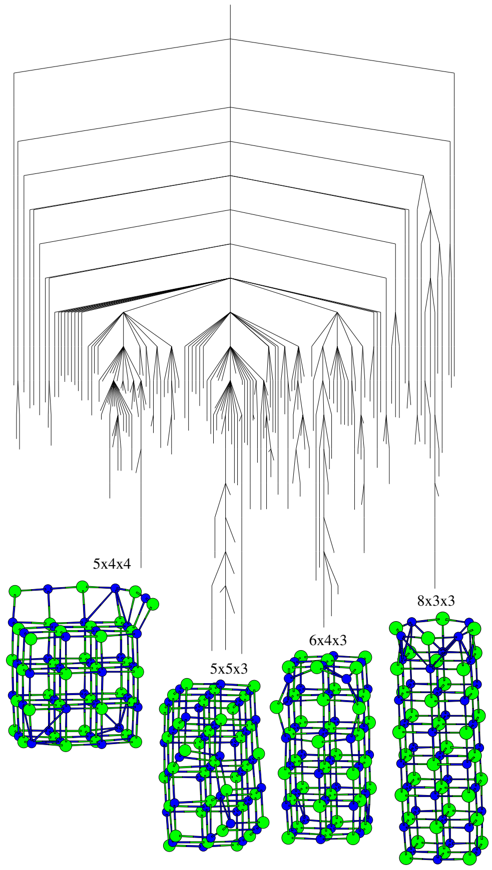


FIGURE 30: Disconnectivity graph showing multiple-funnel character for $(NaCl)_{35}Cl^{-}$.

good yield of C_{60}.

6.3.2 $(\mathrm{NaCl})_{35}\mathrm{Cl}^-$

The potential energy surface of a binary salt cluster, $(\mathrm{NaCl})_{35}\mathrm{Cl}^-$, has recently been analysed in detail. There are several families of low energy rocksalt cuboids. The corresponding disconnectivity graph is shown in Figure 30. Only branches ending at the lowest 100 minima are included for clarity. The tree is roughly divided into four families based on $5\times5\times3$, $6\times4\times3$, $8\times3\times3$ and $5\times4\times4$ cuboids. The different cuboidal rocksalt morphologies lie at the bottom of separate funnels. Relaxation within each funnel is efficient, but relaxation between funnels occurs on a much longer timescale.

In drift tube 'ion chromatography' experiments three peaks are seen in the arrival time spectrum. These may be assigned to the $5\times5\times3$, $6\times4\times3$ and $8\times3\times3$ cuboids. The clusters relax into the $5\times5\times3$ funnel on the experimental timescale. The $(\mathrm{NaCl})_{35}\mathrm{Cl}^-$ clusters relax rapidly down any one funnel, but may not reach the global minimum except over much longer time scales.

The mechanisms that lead to changes in morphology are all highly cooperative processes in which two parts of the cluster 'glide' past each other on a $\{110\}$ plane in a $\langle 1\bar{1}0\rangle$ direction (Figure 31).

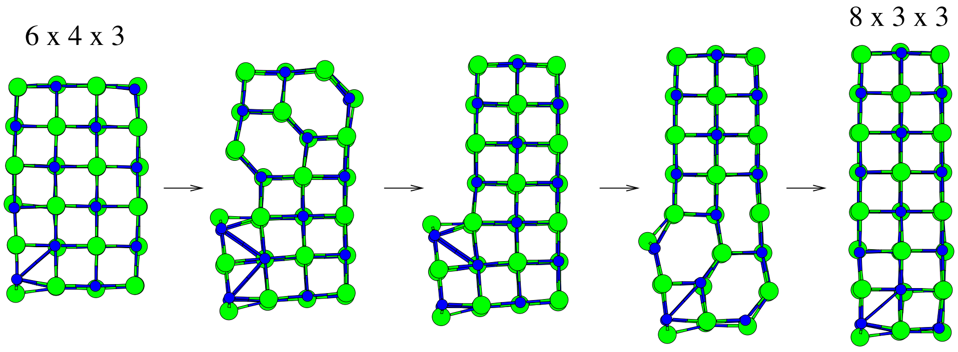


FIGURE 31: Glide mechanism for $(\mathrm{NaCl})_{35}\mathrm{Cl}^-$.

6.3.3 P46: A Model Protein

This model protein consists of 46 'beads' linked by stiff springs. There are three types of bead: B=hydrophobic, L=hydrophilic, N=neutral. The global minimum of $B_9N_3(LB)_4N_3B_9N_3(LB)_5L$ is a four-stranded β-barrel (Figure 32).

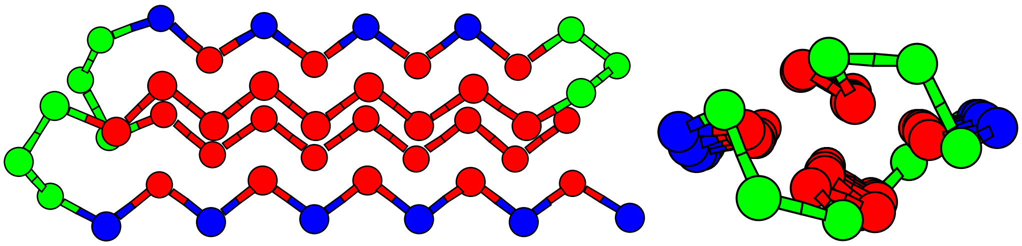


FIGURE 32: P46 global minimum

As can be seen from its disconnectivity graph, the model exhibits frustration in terms of competing low energy β-barrel structures. The corresponding 'Gō' model only retains attractive interactions for the contacts defined by the global minimum, the others are turned off; i.e. the Gō model is minimally frustrated. The disconnectivity graph for Gō-P46 is much more funnel-like.

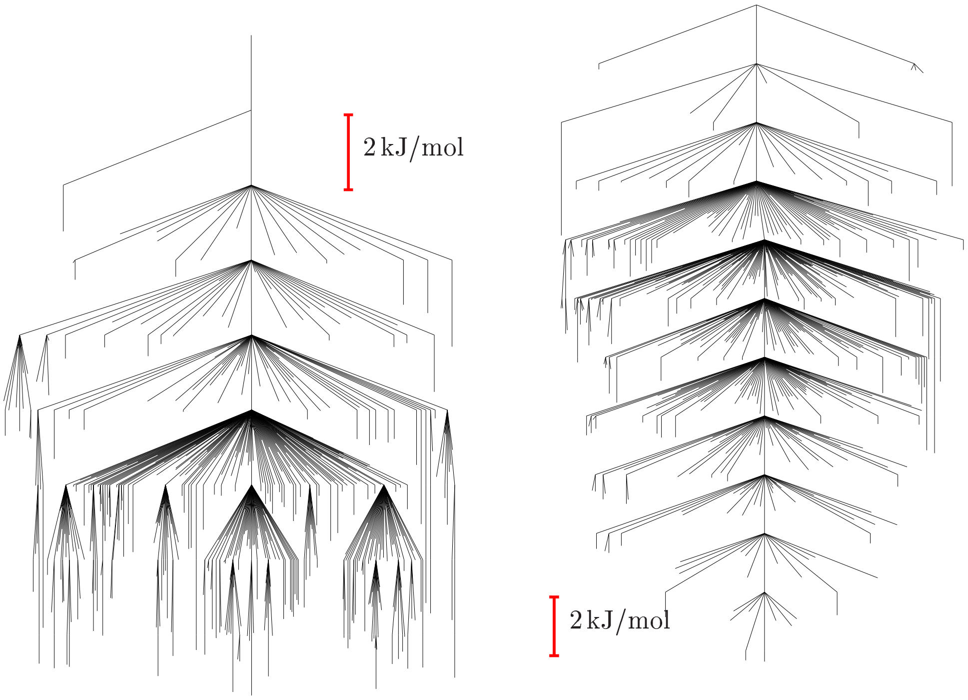


FIGURE 33: Disconnectivity graphs for P46 (left) and its Gō model (right)

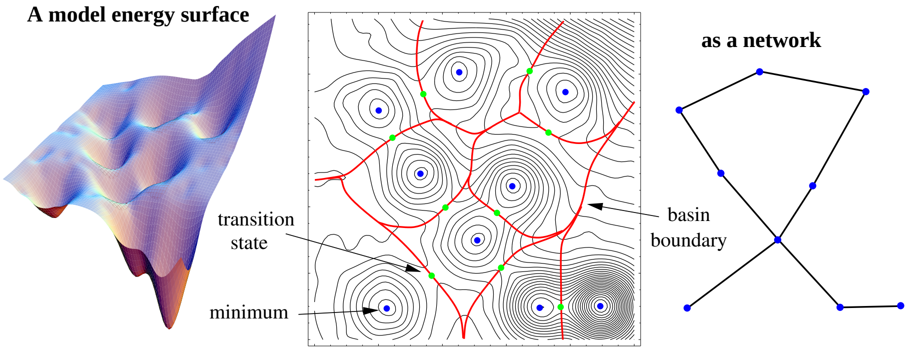


FIGURE 34: A model potential energy surface and its associated network.

6.4 Energy landscapes as networks

The dynamics of a system can be visualized as a hopping between minima on the potential energy surface along transition state valleys. i.e. on a network where the minima are the nodes and edges connect those minima with a common transition state, as illustrated in Figure 34. The topology of this network can provide information on the global organization and structure of the energy landscape.

6.4.1 Complex Networks

The study of complex networks has recently boomed. Many different types have been analysed:

- Communication networks: the internet, world-wide-web, peer-to-peer networks.
- Transport networks: airplane routes, road networks.
- Social networks: friendship networks, actor and scientific collaboration, sexual networks, networks of company directors.
- Biochemical networks: protein-protein interactions, metabolic networks, neural networks.
- Ecological networks: food webs.

Most of these real-world networks have two key features:

Clustering : The neighbours of a node are also likely to be connected. This property is typical of a lattice.

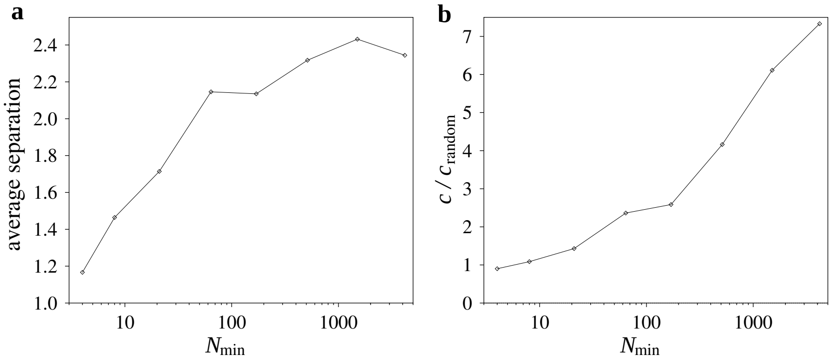


Figure 35: Properties of the energy landscape of Lennard-Jones clusters for $N = 7 - 14$.

Small world effect : The average separation between nodes scales logarithmically with the system size. E.g. any two web pages are separated by 18 clicks; any two people in the world can be connected by 'six' intermediate acquaintances. This property is typical of a random graph.

Many also have a third property:

Scale-free topology: The distribution for the number of connections to a node decays with a power-law. The network is extremely heterogeneous. Most nodes are only connected to a few others, but there are very small fraction of highly-connected hubs (E.g. Yahoo for the WWW). These hubs mediate the short paths between nodes.

6.4.2 Scale-free landscapes

It is possible to find virtually all the minima and transition states on a potential energy landscape for small Lennard-Jones clusters up to $N = 14$. Analyzing the topology of this network reveals that the energy landscape is a small-world network. The average separation scales sublogarithmically with network size (or sublinearly with system size), and the clustering coefficient, c, is significantly higher than for a random network (Figure 35). (c is the probability that the neighbours of a node are also connected.)

This result is good news for global optimization. The number of steps to the global minima from a random minimum scales no more than linearly with the number of atoms. Of course finding this pathway may not be so easy. If the topography of the PES is also favourable, close to the ideal scaling is possible. Using the basin-hopping global optimization algorithm (§10) the global minimum of LJ_{55} can be located after less than 150 minimizations from a random starting point, even though there are $\sim 10^{21}$ minima.

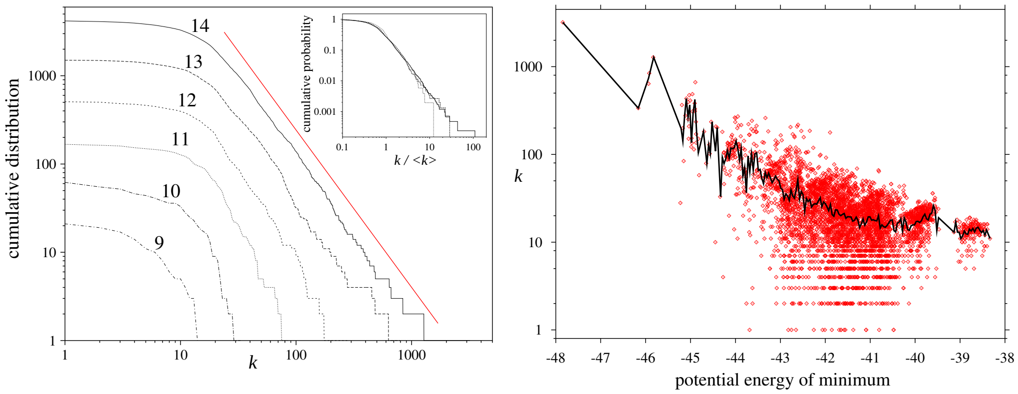


FIGURE 36: The distributions for the number of minima with more than k connections for Lennard-Jones clusters for $N = 9 - 14$ (left). The dependence of the connectivity on the potential energy of the minima (right).

The network also has a scale-free topology, where $p(k) = k^{-2.78}$ for large k (Figure 36). The most-connected hubs correspond to the lowest-energy minima. The corresponding basins of attraction have the largest area and are connected to many much smaller basins around their edges. The topology reflects the topography of the PES. Many pathways converge into the minima at the bottom of the landscapes, as postulated in the funnel hypothesis for protein folding (§1.1).

7 Global Thermodynamics

For a classical system the calculation of the canonical partition function, $Q(T)$, or the microcanonical density of states, $\Omega(E)$, involve integrals over the whole of phase space: e.g

$$Q(T) = \frac{1}{h^{\kappa}} \int \exp(-\beta H) dp^{\kappa} dr^{\kappa} \tag{74}$$

and

$$\Omega(E) = \frac{1}{h^{\kappa}} \frac{d}{dE} \int_{H \leq E} dp^{\kappa} dr^{\kappa}, \tag{75}$$

where $\beta = 1/kT$ and κ is the number of degrees of freedom. For a complex PES these integrals are only possible by numerical methods (typically Monte Carlo).

An alternative 'landscape' approach is to divide configuration space into basins of attraction surrounding each minimum. The partition function can then be exactly decomposed into the sum

$$Q(T) = \sum_i Q_i(T) \tag{76}$$

where $Q_i(T)$ is the partition function for the basin surrounding minimum i and the sum is over all the minima on the PES. This expression is much more tractable as approximate expressions for $Q_i(T)$ are readily available.

The classical partition function for a one-dimensional harmonic oscillator is $1/\beta h\nu$ (Problem 22). The partition function for a system with $\kappa = 3N - 6$ vibrational degrees of freedom is just a product of the partition functions for the individual modes. After including the appropriate Boltzmann weight,

$$Q_i(T) = \frac{e^{-\beta E_i^0}}{(\beta h \bar{\nu}_i)^\kappa} \tag{77}$$

where E_i^0 is the energy of minimum i, $\bar{\nu}_i$ is the geometric mean normal mode frequency. The system has fixed centre of mass and fixed orientation with respect to space fixed axes.

The total partition function in this harmonic superposition approximation is

$$Q(T) = \sum_s \frac{n_s e^{-\beta E_s^0}}{(\beta\, h\, \overline{\nu}_s)^\kappa}, \tag{78}$$

where the sum is over all the structurally distinct minima on the surface and n_s, the number of permutational isomers of minimum s, is given by $n_s = 2N!/h_s$ where h_s is the order of the point group of s (§6.1). As

$$Q(T) = \int \Omega(E) e^{-\beta E} dE \tag{79}$$

$\Omega(E)$ can be calculated by an inverse Laplace transform (problem 23). This gives

$$\Omega(E) = \sum_{E_s^0 < E} \frac{n_s \, (E - E_s^0)^{\kappa - 1}}{\Gamma(\kappa) (h \bar{\nu}_s)^\kappa}. \tag{80}$$

From equations (78) and (80) many canonical and microcanonical thermodynamical properties can be calculated by applying standard thermodynamical formulae (Problem 24). More accurate anharmonic expressions for Q_i can also be used, as well as quantum versions, although the algebra becomes more involved.

The main practical problem with equations (78) and (80) is that for all but the smallest systems it becomes impossible to obtain all the minima on the PES. However, for thermodynamic properties at low temperatures the number of thermodynamically relevant minima may be much smaller, e.g. the relative free energies of two structural forms of a cluster. Methods are also available to obtain the energy distribution of minima, thus extending the method's usage to phase transitions, such as melting. This will be illustrated for a bulk liquid.

As the minima in the liquid form a quasicontinuous distribution, the sum in the superposition approximation can be approximated by an integral. Furthermore, it is reasonable to assume that the vibrational frequecies only depend on energy, allowing the configurational and vibrational contributions to the entropy to be separated. This gives:

$$Q(T) = \int \frac{\Omega_{\min}(E^0)\exp(-\beta E^0)}{(\beta h\bar{\nu}(E^0))^{\kappa}} dE^0, \tag{81}$$

where $\Omega(E^0)$ is the density of minima with energy E^0. By performing a simulation at a temperature T and then minimizing configurations at regular intervals along this trajectory, one can obtain the probability distribution:

$$p_{\min}(E^0, T) \propto \frac{\Omega_{\min}(E^0)\exp(-\beta E^0)}{(\beta h\bar{\nu}(E^0))^{\kappa}}. \tag{82}$$

Therefore, $\Omega_{\min}(E^0)$ can be obtained within a multiplicative factor from this equation, and the absolute values if one also knows the absolute free energy at one temperature. It has been found the density of minima for a liquid is a Gaussian (Figure 37).

8 Global Dynamics

Molecular dynamics is the usual way to study the dynamics of a system by computation. However, the time scale of the simulations is constrained by the short time step required to integrate the fast vibrational motions accurately. By contrast, the master equation approach, which focusses on the inter-minimum dynamics, has no time scale constraints and provides a landscape perspective on the dynamics.

A 'master equation' can be used to describe the dynamics on a multi-dimensional PES, giving the time evolution of the probability distribution, $\mathbf{P}(t) = (P_1(t), P_2(t), \ldots)^T$, where P_{a} is the probability of the system being in state a. The master equation has the form

$$\frac{dP_{\mathrm{a}}(t)}{dt} = \sum_{\mathrm{b}\neq\mathrm{a}} [k_{\mathrm{ab}} P_{\mathrm{b}}(t) - k_{\mathrm{ba}} P_{\mathrm{a}}(t)], \tag{83}$$

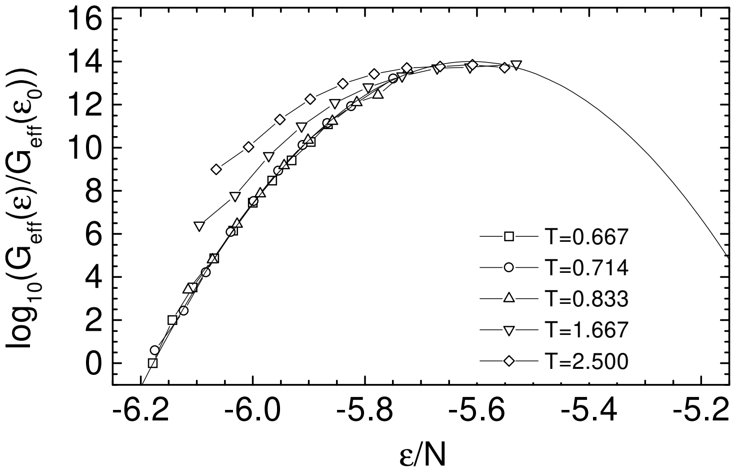


FIGURE 37: The density of minima for a binary Lennard-Jones liquid, obtained from probability distributions at different temperatures.

where k_{ab} is the rate constant for transitions from minimum b to minimum a.

Equation (83) can be solved analytically. First we define the transition matrix, $\mathbf{W}$, with elements

$$W_{\text{ab}} = k_{\text{ab}} - \delta_{\text{ab}} \sum_{\text{c}=1}^{n_{\min}} k_{\text{ca}}, \tag{84}$$

where $n_{\min}$ is the number of local minima in the sample and the diagonal components W_{aa} contain minus the total rate constant out of minimum a. The master equation (83) can now be written in matrix form:

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{W}\,\mathbf{P}(t). \tag{85}$$

The system has a unique equilibrium state if $\mathbf{W}$ cannot be decomposed into block form, for which a necessary condition is that each minimum must be connected to at least one of the others in the database. In this case the vector of equilibrium occupation probabilities, $\mathbf{P}^{\text{eq}}$, must satisfy

$$\frac{d\mathbf{P}^{\text{eq}}}{dt} = 0, \tag{86}$$

and so $\mathbf{W}$ has a zero eigenvalue for which the corresponding eigenvector is

$\mathbf{P}^{\mathrm{eq}}$. Substituting the equilibrium probabilities into equation (83) produces

$$\sum_{\mathrm{b}\neq\mathrm{a}} [k_{\mathrm{ab}} P_{\mathrm{b}}^{\mathrm{eq}} - k_{\mathrm{ba}} P_{\mathrm{a}}^{\mathrm{eq}}] = 0, \tag{87}$$

which is satisfied if the rate constants and equilibrium probabilities obey the detailed balance condition:

$$k_{\mathrm{ab}} P_{\mathrm{b}}^{\mathrm{eq}} = k_{\mathrm{ba}} P_{\mathrm{a}}^{\mathrm{eq}}. \tag{88}$$

Detailed balance can also be used to construct a symmetric transition matrix $\widetilde{\mathbf{W}}$:

$$\begin{aligned} \widetilde{W}_{\mathrm{ab}} &= \sqrt{\frac{P_{\mathrm{b}}^{\mathrm{eq}}}{P_{\mathrm{a}}^{\mathrm{eq}}}}\, W_{\mathrm{ab}} = \sqrt{\frac{P_{\mathrm{b}}^{\mathrm{eq}}}{P_{\mathrm{a}}^{\mathrm{eq}}}} \left(k_{\mathrm{ab}} - \delta_{\mathrm{ab}} \sum_{\mathrm{c}} k_{\mathrm{ca}} \right) \\ &= \sqrt{\frac{P_{\mathrm{a}}^{\mathrm{eq}}}{P_{\mathrm{b}}^{\mathrm{eq}}}} \left(k_{\mathrm{ba}} - \delta_{\mathrm{ba}} \sum_{\mathrm{c}} k_{\mathrm{ca}} \right) = \widetilde{W}_{\mathrm{ba}}. \end{aligned} \tag{89}$$

Writing $\widetilde{P}_{\mathrm{a}} = P_{\mathrm{a}} / \sqrt{P_{\mathrm{a}}^{\mathrm{eq}}}$ the master equation becomes

$$\frac{d\widetilde{\mathbf{P}}(t)}{dt} = \widetilde{\mathbf{W}}\, \widetilde{\mathbf{P}}(t). \tag{90}$$

Since $\widetilde{\mathbf{W}}$ is symmetric we can define its eigenvalues, λ_a, and eigenvectors using an orthogonal transformation matrix $\mathbf{u}$ (§2.1.1):

$$\sum_{\mathrm{b}} \widetilde{W}_{\mathrm{ab}} u_{\mathrm{bc}} = \lambda_{\mathrm{c}}\, u_{\mathrm{ac}}. \tag{91}$$

Transforming to new coordinates, p_{a}, where

$$p_{\mathrm{a}} = \sum_{\mathrm{b}} u_{\mathrm{ba}}\, \widetilde{P}_{\mathrm{b}}, \qquad \text{and} \qquad \widetilde{P}_{\mathrm{a}} = \sum_{\mathrm{b}} u_{\mathrm{ab}}\, p_{\mathrm{b}}, \tag{92}$$

equation (90) now becomes

$$\sum_{\mathrm{c}} u_{\mathrm{ac}}\, \dot{p}_{\mathrm{c}}(t) = \sum_{\mathrm{c,b}} \widetilde{W}_{\mathrm{ab}}\, u_{\mathrm{bc}}\, p_{\mathrm{c}}(t) = \sum_{\mathrm{c}} \lambda_{\mathrm{c}}\, u_{\mathrm{ac}}\, p_{\mathrm{c}}(t), \tag{93}$$

with solution $p_a(t) = p_a(0) \exp(\lambda_a t)$. Transforming back to the P_{a} variables:

$$\begin{aligned} \widetilde{P}_{\mathrm{a}}(t) &= \sum_{\mathrm{b}} u_{\mathrm{ab}}\, p_{\mathrm{b}}(0)\, \mathrm{e}^{\lambda_{\mathrm{b}} t} \\ &= \sum_{\mathrm{b,c}} u_{\mathrm{ab}}\, u_{\mathrm{cb}}\, \widetilde{P}_{\mathrm{c}}(0)\, \mathrm{e}^{\lambda_{\mathrm{b}} t} \\ \text{so} \quad P_{\mathrm{a}}(t) &= \sqrt{P_{\mathrm{a}}^{\mathrm{eq}}} \sum_{\mathrm{b}} u_{\mathrm{ab}}\, \mathrm{e}^{\lambda_{\mathrm{b}} t} \left[\sum_{\mathrm{c}} u_{\mathrm{cb}} \frac{P_{\mathrm{c}}(0)}{\sqrt{P_{\mathrm{c}}^{\mathrm{eq}}}} \right]. \end{aligned} \tag{94}$$

Apart from the zero eigenvalue, all the λ_{b} are negative. The eigenvalues and eigenvectors are normally labelled in decreasing order according to the value of λ_{b}, so that $\lambda_1 = 0$, $\lambda_{\mathrm{b}} < 0$ for $2 \leq \mathrm{b} \leq n_{\mathrm{min}}$, and λ_2 is the largest nonzero eigenvalue with the smallest magnitude corresponding to the slowest relaxation process. As $t \to \infty$, only the $\mathrm{b} = 1$ term in equation (94) remains because the contributions of the other modes decay exponentially, and $\mathbf{P}(t) \to \mathbf{P}^{\mathrm{eq}}$, as expected.

8.1 Statistical Rate Theories

In applying the master equation approach we must use consistent theories for thermodynamic properties, such as P^{eq}, and the rate constants, so that detailed balance is satisfied. The transition state theory expression for the unimolecular canonical rate constant $k_{\mathrm{a}}^{\dagger}$, out of minimum a through transition state † is

$$k_{\mathrm{a}}^{\dagger}(T) = \frac{kT}{h}\frac{Q^{\dagger}}{Q_{\mathrm{a}}}\mathrm{e}^{-\Delta V/kT}, \tag{95}$$

where the transition state partition function $Q^{\dagger}$ does not include the unique mode with imaginary frequency, and ΔV is the potential energy difference between the transition state and minimum a. Using the harmonic expressions for these partition function (equation (77)) then gives

$$k_{\mathrm{a}}^{\dagger}(T) = \frac{\bar{\nu}_{\mathrm{a}}^{\kappa}}{\bar{\nu}^{\dagger(\kappa-1)}}\mathrm{e}^{-\Delta V/kT}. \tag{96}$$

8.2 Annealing of C_{60}

Using the same pathways that were employed to construct the disconnectivity graph in §6.3.1 we can apply the master equation approach to study the annealing of C_{60}. A high temperature is required to produce appreciable population of the buckminsterfullerene minimum on the experimentally observed time scales, in good agreement with the high temperature annealing that is needed to produce significant yields of this structure. These results reflect the 'weeping willow' energy landscape revealed in Figure 29, namely the presence of a funnel leading to the global minimum but with relatively large barriers.

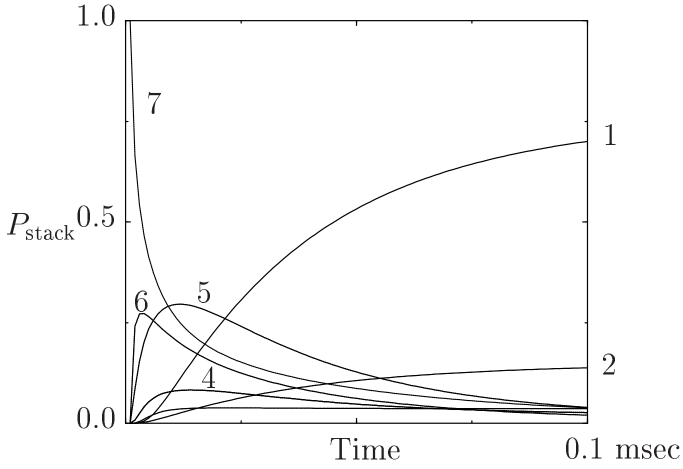


FIGURE 38: Evolution of the occupation probabilities of the seven lowest SW stacks of minima at $E = 3728\,\text{kJmol}^{-1}$.

9 Further examples

Clusters bound by the Lennard-Jones potential:

$$V = 4\epsilon \sum_{s<t} \left[\left(\frac{\sigma}{r_{st}} \right)^{12} - \left(\frac{\sigma}{r_{st}} \right)^{6} \right]. \tag{97}$$

have become a commonly-studied model system to understand the fundamental properties of clusters. They also provide a reasonable model for rare gas clusters, such as Ar.

9.1 The $\mathbf{LJ_{13}}$ Cluster

The LJ_{13} cluster provides an example of a landscape with a well-defined global minimum and a disconnectivity graph that we will categorise as a 'palm tree'. For this size the lowest energy structure is a Mackay icosahedron with full I_h point group symmetry. The next three lowest energy isomers correspond to defect structures where a single atom is removed from the icosahedral shell and placed in one of the three distinct capping sites. There is a relatively large energy gap of $2.85\,\epsilon$ between the icosahedron and the next-lowest structure, because of the favourable geometrical packing in the icosahedron.

Two disconnectivity graphs for this system are shown in Figures 39 and 40. In the first graph all the local minima are included, but permutational isomers are not distinguished. In the second graph permutational isomers of one 'tagged' atom are represented separately, and so two distinct isomers of

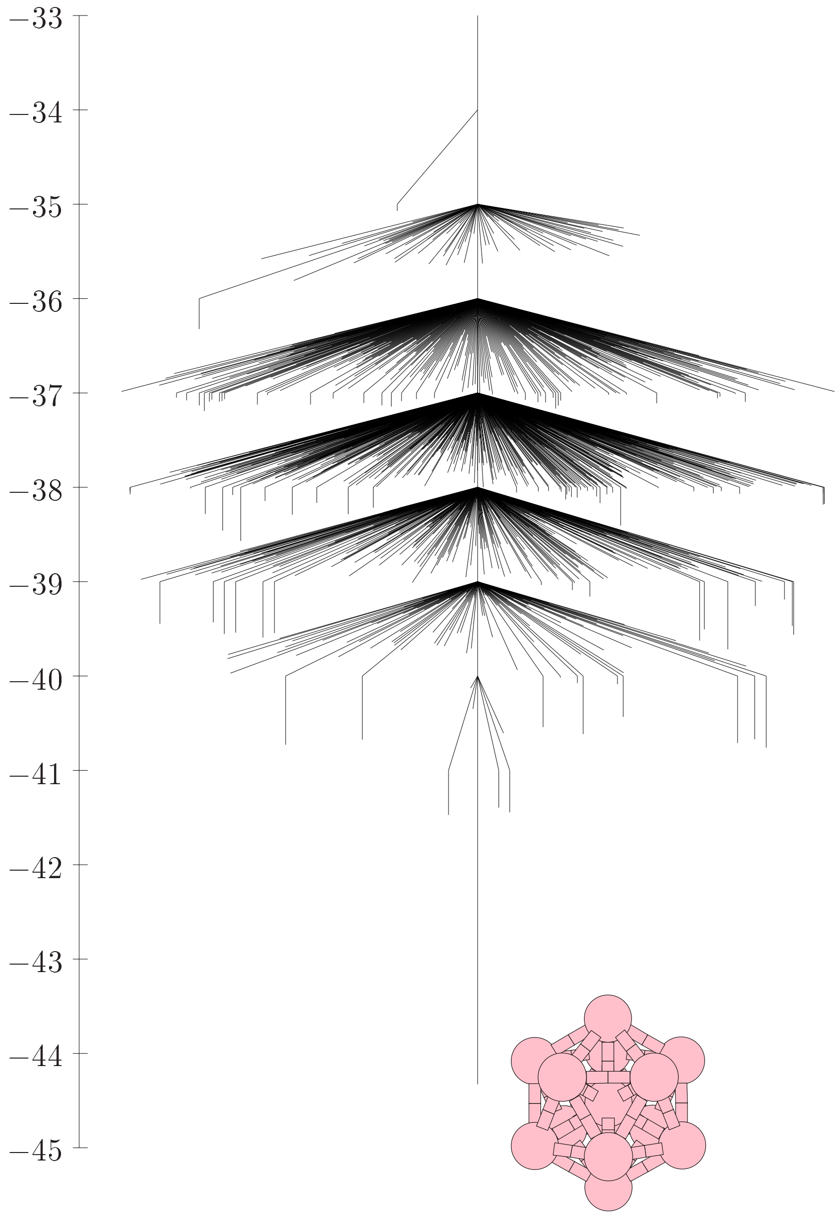


Figure 39: Disconnectivity graph for the LJ_{13} cluster. In this graph all permutation-inversion isomers of each structure are grouped together.

the global minimum are present where the tagged atom occupies either the centre or a surface site. Relaxation to any one of the $2 \times 13!/120$ distinct permutational isomers of the global minimum is efficient over a wide range of temperature or total energy. However, interconversion between different permutational isomers of the icosahedron, with interchange of an atom between

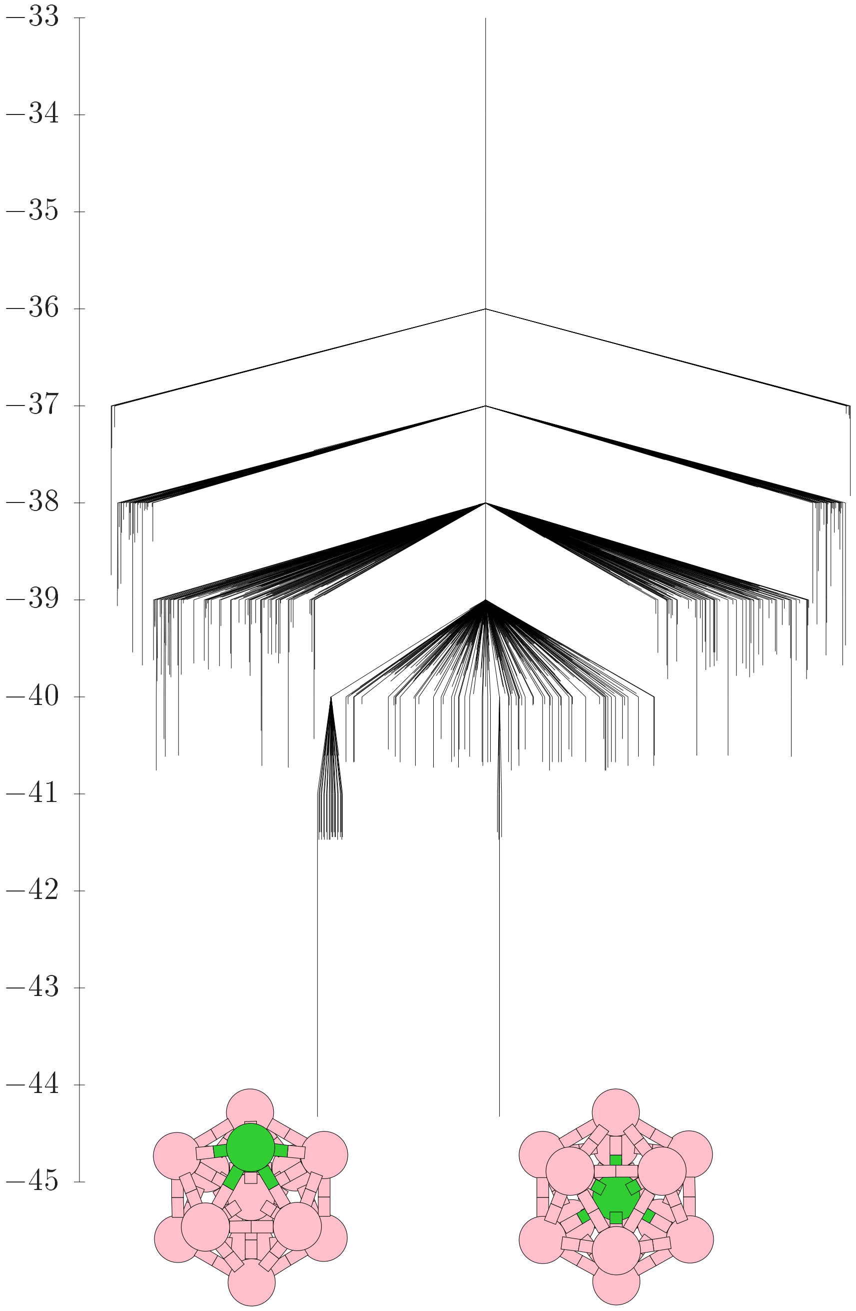


FIGURE 40: Disconnectivity graph for LJ_{13} where permutation-inversion isomers of the shaded atom are distinguished. There are two different isomers of the global minimum where the tagged atom occupies either the centre or a surface site. Permutation-inversion isomers of the other twelve atoms are grouped together.

the centre and surface sites, is slower, because of the high barriers involved

(Figure 40). In fact, relaxation to the equilibrium distribution is about two orders of magnitude slower if we start from a distribution consisting entirely of isomers with the tagged atom in the centre site (Figure 41b).

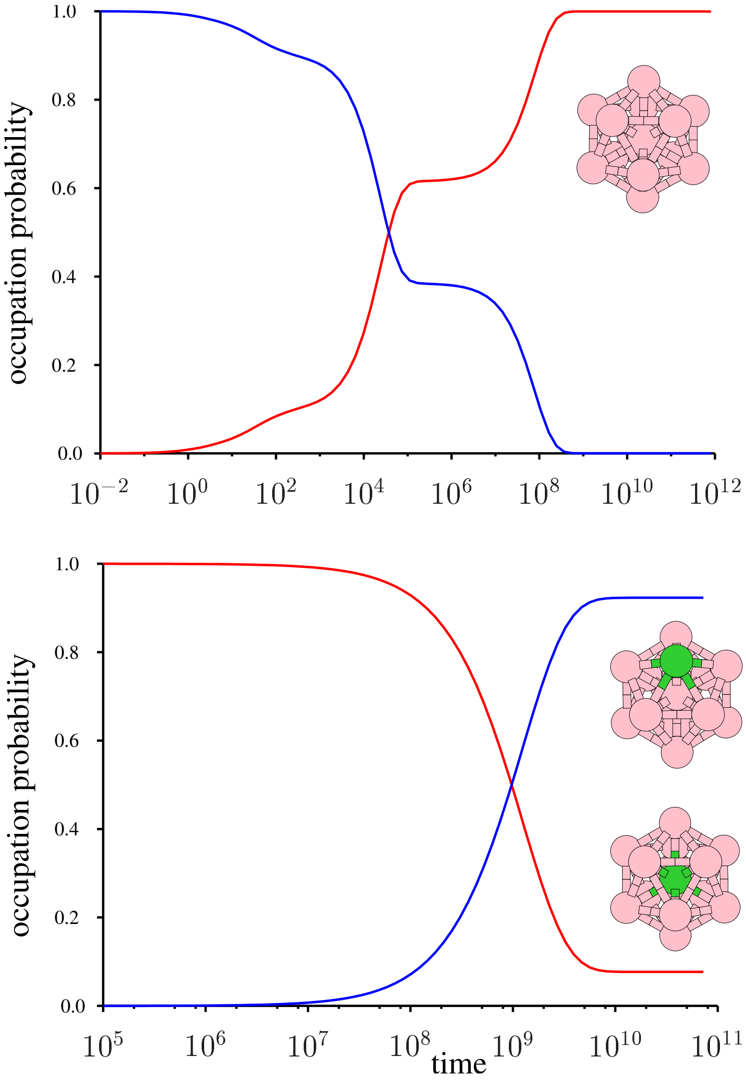


FIGURE 41: Relaxation of non-equilibrium occupation probability distributions for LJ_{13} calculated using the master equation approach described in §8. (a) The initial distribution, corresponding to equilibrium at high temperature. The occupation probabilities of all the icosahedral isomers are added together to produce the red line, while the blue line represents the occupation probability of all the other minima. (b) An initial distribution where only icosahedral isomers with the tagged atom in the centre are occupied.

9.2 The Double Funnel LJ_{38} Surface

In contrast to most Lennard-Jones clusters the global minimum is not icosahedral, but is a face-centred-cubic truncated octahedron. As these two regions of configuration space are well-separated, both in distance and by a large energy barrier, the energy landscape is like a double-funnel (Figure 42).

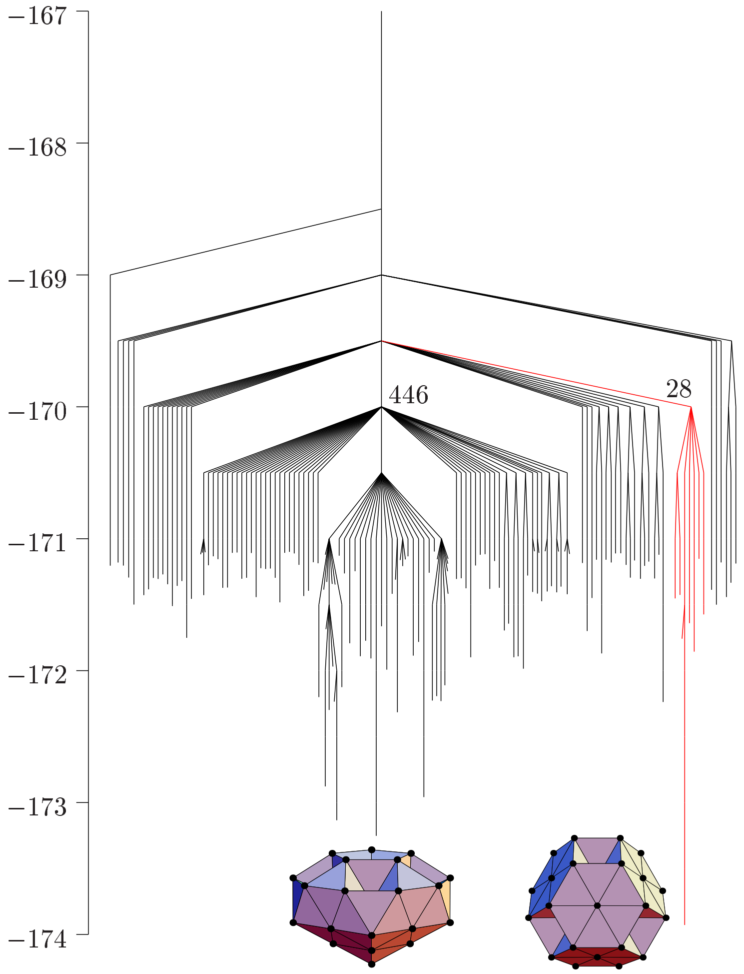


FIGURE 42: Disconnectivity graph for LJ_{38}. Only branches that lead to the lowest 150 minima are included.

As there are many more low-energy icosahedral minima, the icosahedral funnel is much wider. This configurational entropy difference, combined with the greater vibrational entropy of the icosahedral structures, drives a low-temperature solid-solid transition, at the point where the icosahedral structures become the lowest in free energy. This transition occurs just below that for melting and gives rise to a small peak in the heat capacity (Figure 43).

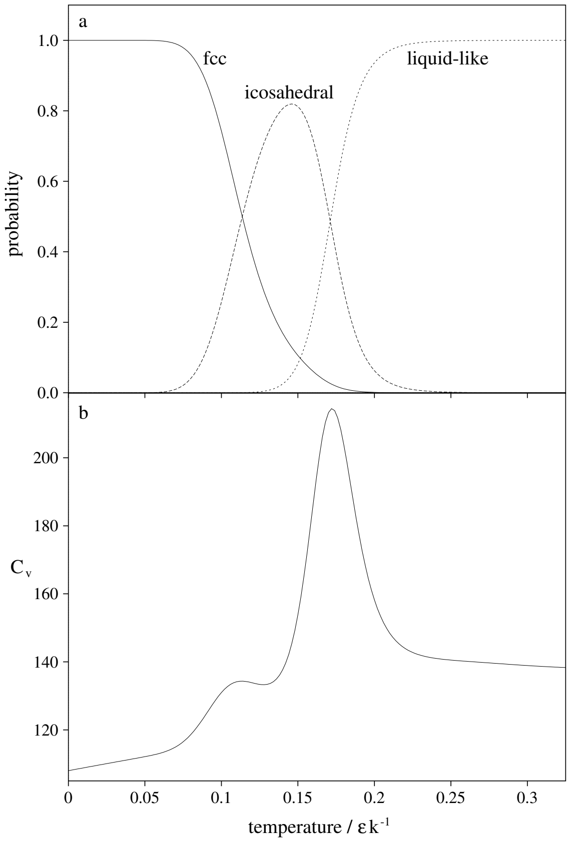


FIGURE 43: The thermodynamic properties of LJ_{38} obtained using the superposition approximation (§7).

Relaxation to the global minimum is hampered by the presence of a competing morphology. The system preferentially enters the wider icosahedral funnel. Interfunnel passage is extremely slow and so equilibrium between the funnels is only obtained on much longer time scales (Figure 44). Similarly, if one cools a liquid cluster, it is thermodynamically favourable to enter the icosahedral funnel, where it then becomes trapped, even when the fcc truncated octahedron becomes lowest in free energy ($T_f < T_g$ (§1.1)).

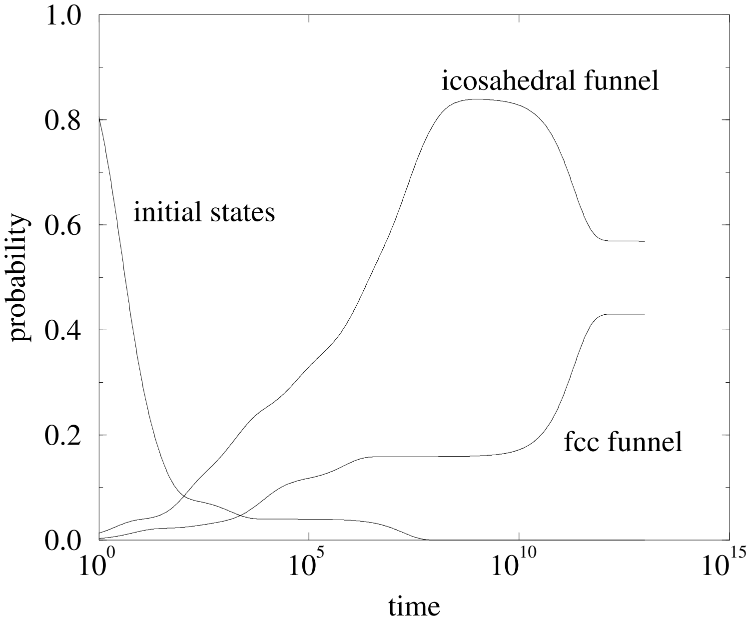


FIGURE 44: Relaxation to the global minimum of LJ_{38} from a high temperature distribution, obtained using the master equation approach (§8).

10 Global Optimization

The global optimization problem is a subject of intense current interest. Applications of obvious economic importance include travelling salesman-type problems and the design of microprocessor circuitry. In the domain of atoms and molecules, one often wishes to discover the lowest-energy isomer or crystal structure for a system with a given composition. Attempting to predict the three-dimensional structure of a protein solely from its amino acid sequence by computer simulation is a major research effort.

A random conformational search for the global minimum of a typical protein would take an astronomical time (problem 1). However, it is now generally agreed that the search is not random, but instead the PES is biased towards the global minimum, resulting in efficient relaxation. Hence the amino acid sequences of naturally occurring proteins have presumably evolved to fold rapidly into a unique native structure. In treating any non-trivial global optimization problem, the principal difficulty arises from the exponentially large number of minima on the PES (§6.1).

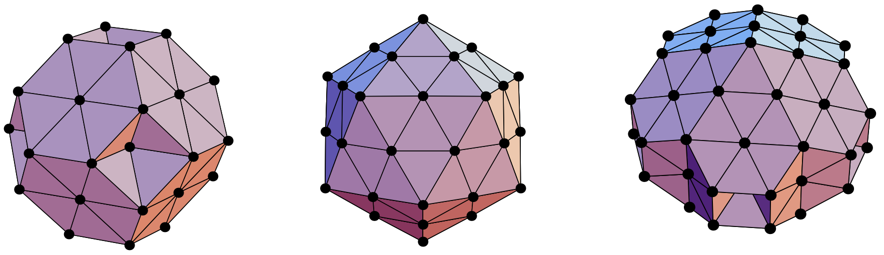


FIGURE 45: From left to right: global minima of the LJ potential for 38 atoms (truncated octahedron), 55 atoms (Mackay icosahedron), and 75 atoms (Marks decahedron).

For the cluster of 55 atoms interacting by a Lennard-Jones (LJ) potential (Figure 45) the number of minima (excluding permutational isomers) is at least 10^{10}, but the global minimum is relatively easy to locate because of the focusing properties of the energy landscape. Most global minima for LJ clusters containing fewer than 100 atoms are based upon icosahedral packing. The exceptions (LJ_{38}, LJ_{75-77} and LJ_{98}) serve as particularly interesting test cases, because the corresponding energy landscapes consist of two families of structures. Two examples, the truncated octahedron for LJ_{38} and the Marks decahedron for LJ_{75}, are shown in Figure 45. The disconnectivity graph for LJ_{38} was illustrated in §9.2. At these sizes, the lowest-energy minimum based upon icosahedral packing acts as a trap, and is widely separated from the true global minimum. Actually, it is quite easy to find the global minima of these clusters by seeding the starting geometry with a core of the right type.

Simulated annealing provided the first generally applicable technique for global optimization. Here the state of the system is followed by simulation as the temperature is decreased slowly from a high value, in the hope that it will eventually come to rest at the global potential energy minimum. Unfortunately, the free energy global minimum may change at a temperature where energy barriers are too high for the system to escape from a local minimum (§9.2).

The difficulty of global optimization is that any algorithm needs to perform both local and global search. One needs to know the lowest-energy point within a particular basin of attraction to know whether it is the global minimum, but one also needs to search all relevant regions of configuration space. In simulated annealing this is achieved through the temperature. At high temperature many regions of configuration space are sampled, whereas at low temperature the simulation should converge to the nearest local minimum. A much more efficient way to achieve both these elements is to couple the search method to a local minimization algorithm.

10.1 Basin-Hopping Global Optimization

There is a simple transformation of the energy landscape that does not change the global minimum, or the relative energies of any local minima:

$$\tilde{E}(\mathbf{X}) = \min\{E(\mathbf{X})\} \tag{98}$$

where 'min' signifies that an energy minimization is carried out starting from $\mathbf{X}$. The transformed energy, $\tilde{E}(\mathbf{X})$, at any point, $\mathbf{X}$, becomes the energy of the structure obtained by minimization. Each local minimum is, therefore, surrounded by a catchment basin of constant energy consisting of all the neighbouring geometries from which that particular minimum is obtained. The catchment basin transformation removes all the transition state regions from the surface and accelerates the dynamics because the system can pass between basins all along their boundary. Atoms can even pass through each other without encountering prohibitive energy barriers.

On this landscape the system can hop easily between basins, whereas on the original landscapes the system can spend a considerable time undergoing vibrational motion within a basin before a transition happens.

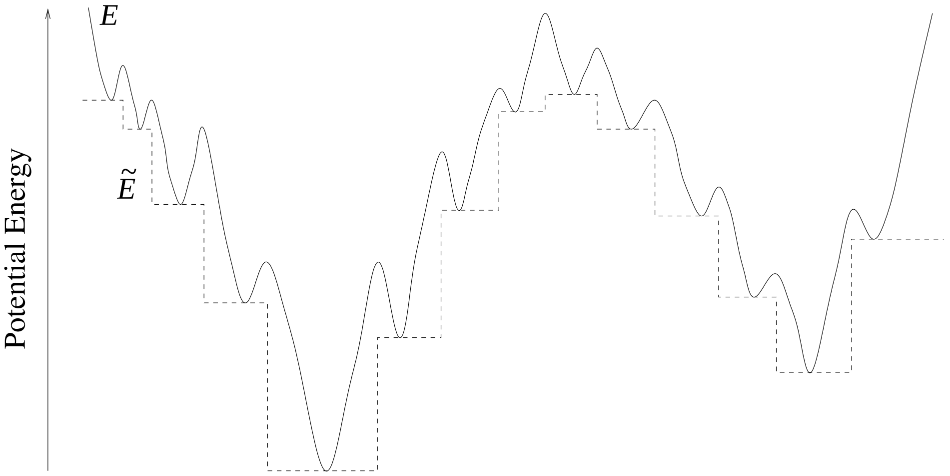


FIGURE 46: Illustration of the $\tilde{E}(\mathbf{X})$ energy landscape transformation.

The basin-hopping approach therefore transforms the energy landscape to a discrete set of energy levels corresponding to the energies of local minima (Figure 46), and must be combined with a search strategy. In the 'Monte

Carlo plus energy minimization' (MCM) procedure steps are proposed by perturbing the current coordinates and carrying out a minimization from the resulting geometry. A step is accepted if the energy of the new minimum, $E_{\rm new}$, is lower than the starting point, $E_{\rm old}$. If $E_{\rm new} > E_{\rm old}$ then the step is accepted if $\exp[(E_{\rm old} - E_{\rm new})/kT]$ is greater than a random number drawn from the interval [0,1]. The temperature, T, becomes an adjustable parameter.

Other methods, such as genetic algorithms, are also used to search the transformed landscape.

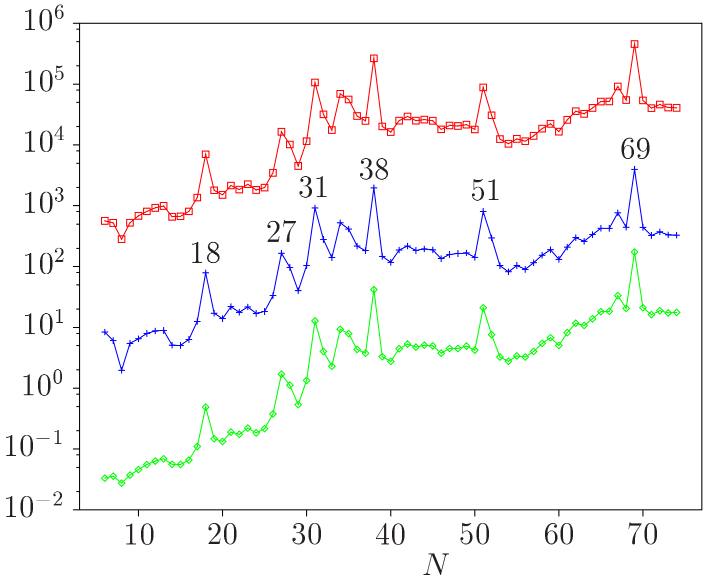


FIGURE 47: Mean cpu time in seconds (green), number of basin-hopping steps (blue), and energy/gradient calls (red) required to find the global minimum with the basin-hopping MCM procedure for LJ_N clusters up to $N = 74$. The cpu time (seconds) corresponds to a 336 MHz Sun UltraSparc II processor.

The mean number of basin-hopping steps and cpu time required to find the global minimum at a fixed temperature are shown in Figure 47 for LJ_n up to $n = 74$. The results are averages over 100 different random starting points in each case.

Problems

The easiest problems are labelled with a single $*$, the hardest with $* * *$.

1. Naturally occurring proteins are unbranched chains consisting of L-amino acids linked together by amide or peptide bonds between the carboxyl group of one residue and the amino group of the next:

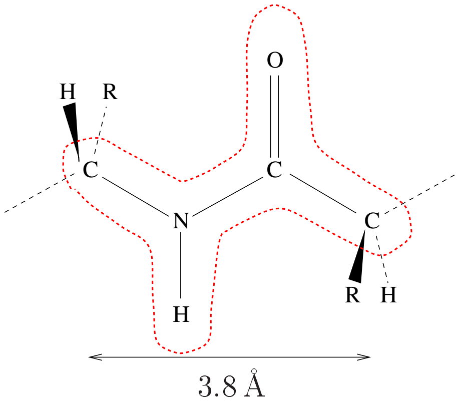


FIGURE 48: Schematic view of a peptide bond.

   The six atoms in each unit surrounded by dashed lines in Figure 48 are coplanar.

   Estimate the number of stable configurations for a protein containing r amino acids if each one has d locally stable configurations. If these local minima could be explored at a rate corresponding to a typical vibration, how long would it take to search them all if $r = 100$ and $d = 2$? Compare this with the observed time scale for protein folding, which is typically less than a second. $*$

2. Show that the general relaxation time defined in equation (2) is just τ if $p = p(0)e^{-t/\tau}$. $*$

3. Show how the VTF equation (1) form is recovered from the Adam-Gibbs equation (3) if $C_{p,\text{liquid}} - C_{p,\text{crystal}} \propto 1/T$. $**$

4. Derive handout equation (8) by substituting the product function $\psi(\mathbf{x}, \mathbf{X}) = \psi_\text{e}(\mathbf{x}; \mathbf{X})\psi_\text{n}(\mathbf{X})$ into equation (4). $*$

5. The components of the Born-Oppenheimer product wavefunction in question 4 satisfy

$$\begin{aligned} \left(\hat{\mathcal{H}} - \hat{T}_{\rm n}\right) \psi_{\rm e}(\mathbf{x};\mathbf{X}) &= E_{\rm e}(\mathbf{X})\psi_{\rm e}(\mathbf{x};\mathbf{X}) & (99) \\ \left(\hat{T}_{\rm n} + E_{\rm e}(\mathbf{X})\right) \psi_{\rm n}(\mathbf{X}) &= E_{\rm total}\psi_{\rm n}(\mathbf{X}). & (100) \end{aligned}$$

   First apply the variational principle to the electronic problem and the Hamiltonian $\hat{\mathcal{H}} - \hat{T}_{\rm n}$ by integrating the exact wavefunction only over the coordinates $\mathbf{x}$. Then apply the variational principle again to the operator $\hat{T}_{\rm n} + E_{\rm e}(\mathbf{X})$ integrating only over $\mathbf{X}$. Now multiply through by the normalising factors and integrate over all space to prove that $E_{\rm exact} \geq E_{\rm total}$. $**$

6. For a hydrogen atom the normalised ground state Born-Oppenheimer electronic wavefunction is $\sqrt{\frac{1}{\pi}}e^{-r}$. What is the energy of this function in atomic units? If the term corresponding to the nuclear kinetic energy is treated as a perturbation evaluate the first order correction to the ground state energy. The exact result in this case can be obtained by replacing the electron mass by the reduced mass of the electron plus proton in the expression for the unperturbed energy (think about the dimensions of energy). How does this compare with the perturbation result? $**$

7. By considering the solutions of a secular determinant for two interacting states determine the conditions under which two potential energy surfaces can cross, both for a diatomic and a polyatomic molecule. $***$

8. Show that the matrix $\mathbf{C}$ described in §2.1.1 is orthogonal for a transformation between orthonormal sets of basis vectors. Hence show that such transformations preserve both length and the angles between vectors. $**$

9. Verify the chain rule result in handout equation (21) and the transformation of the gradient components in equation (19). $*$

10. Show that $\mathbf{A}^T\,\overline{\mathbf{H}}\,\mathbf{A}$ is a diagonal matrix starting from equation (25) and using the orthogonality of $\mathbf{A}$. $*$

11. Find $\partial/\partial q_\alpha$ in terms of the $\partial/\partial Q_\beta$'s using the chain rule. $*$

12. Consider a linear triatomic molecule ABA aligned along the z axis with atomic masses M_A and M_B. If the atoms are constrained to move along the z axis subject to the potential

$$V = \frac{1}{2}k(z_{A1} - z_B + r_e)^2 + \frac{1}{2}k(z_{A2} - z_B - r_e)^2, \qquad (101)$$

where r_e is the equilibrium separation, derive, (a) the normal mode frequencies and (b) the normal mode coordinates both in terms of z_{A1}, z_{A2} and z_B and in terms of the mass weighted coordinates $q_{A1} = z_{A1}\sqrt{M_A}$, $q_{A2} = z_{A2}\sqrt{M_A}$ $q_B = z_B\sqrt{M_B}$. What are the relative z displacements of the three atoms in each normal mode? $***$

13. Why are there only five zero Hessian eigenvalues for a linear molecule? How many non-zero Hessian eigenvalues are there for an interacting system containing N_a atoms, N_l rigid linear molecules and N_{nl} rigid non-linear molecules? (Linear and non-linear rigid molecules have 5 and 6 degrees of freedom, respectively.) $**$

14. Using $X_\alpha = \sum_{\beta=1}^{3N} B_{\alpha\beta} W_\beta$ and equation (48) transform equation (43) into equation (49). $**$

15. How do steepest-descent paths approach a transition state in terms of the principal directions? $*$

16. For the two-dimensional potential surface

$$V(x, y) = -\alpha x^2 - \beta y^3, \qquad \alpha,\ \beta > 0, \tag{102}$$

show that an infinite number of steepest-descent paths originate from the origin, $x = y = 0$, even though this point is a true transition state according to Murrell and Laidler's definition. (Start from the definition of a steepest-descent line and then transform to u before integrating.) Why is this? $***$

17. What can you deduce from symmetry arguments about the stationary points connected by a transition state that is (a) linear, (b) planar, (c) optically inactive, (d) optically active? Assume that the products and reactants are unrelated by symmetry. Prove that the point group of the transition state cannot be larger than the largest subgroup common to the reactants and products in this case. $**$

18. To derive the transmission probability for a rectangular barrier consider a potential of the form

$$V(x) = \begin{cases} 0, & x < 0, \\ V, & 0 < x < a, \\ 0, & a < x. \end{cases}$$

Let the total energy be E and change to the more convenient variables $E' = 2mE/\hbar^2$ and $V' = 2mV/\hbar^2$, and define $k_1 = \sqrt{E'}$ and $k_2 = \sqrt{V' - E'}$.

Assuming $E < V$ show that the wavefunction must have the form

$$\psi(x) = \begin{cases} Ae^{ik_1x} + Be^{-ik_1x}, & x < 0, \\ Ce^{k_2x} + De^{-k_2x}, & 0 < x < a, \\ Ge^{ik_1x} + Fe^{-ik_1x}, & a < x. \end{cases}$$

By matching the wavefunctions and their first derivatives at $x = 0$ and $x = a$ and assuming that the incident wave approaches from $-x$ show that

$$\begin{aligned} A + B &= C + D, \\ A - B &= \frac{k_2}{ik_1}(C - D), \\ C &= \frac{G}{2}\left(\frac{k_2 + ik_1}{k_2}\right) e^{(ik_1 - k_2)a}, \\ D &= \frac{G}{2}\left(\frac{k_2 - ik_1}{k_2}\right) e^{(ik_1 + k_2)a}. \end{aligned}$$

Hence deduce an expression for A in terms of G and show that the transmission coefficient, $T = |G/A|^2$, is

$$T = \left(\frac{V^2}{4E(V-E)}\sinh^2\left[a\sqrt{2m(V-E)}/\hbar\right] + 1\right)^{-1}. \tag{103}$$

(It is probably easiest to first consider $|A/G|^2$ and use the identities $(e^x + e^{-x})/2 = \cosh x = 2\sinh^2(x/2) + 1$.) If $E > V$ show that $T = 1$ when $a\sqrt{E' - V'} = n\pi$ where n is an integer.

Sketch T versus E/V for $0.9 < E/V < 1.4$ for a proton impinging on a barrier of height 1 eV and width 1.5 Å. $***$

19. Prove that the gradient vector transforms as the totally symmetric irreducible representation of the prevailing point group. What implications does this result have for geometry optimisation? $**$

20. If vector $\mathbf{e}$ is a normalised eigenvector of matrix $\mathbf{H}$ with eigenvalue μ show that it is also an eigenvector of the matrix with elements $H_{\alpha\beta} + \lambda e_\alpha e_\beta$ and determine the eigenvalue. $*$

21. Derive the expression for the Newton-Raphson step, $\mathbf{h}_{\mathrm{NR}} = -\mathbf{H}^{-1}\mathbf{g}$. Verify the results for the change of coordinates in equation (61). Derive the corresponding expressions for the eigenvector-following step. $**$

22. Show that the classical partition function for a 1D harmonic oscillator is $1/\beta h\nu$ by (i) applying the classical limit to the quantum partition function, and (ii) performing the phase space integral:

$$q = \frac{1}{h}\int\int \exp(-\beta H)dp\,dq, \tag{104}$$

where the Hamiltonian is

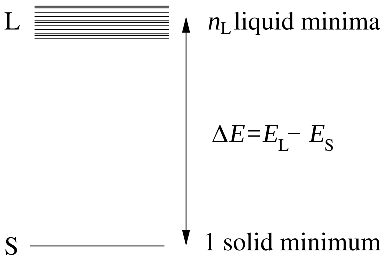
$$H = \frac{p^2}{2\mu} + \frac{k}{2}x^2. \tag{105}$$

$*$

23. Perform the integral in equation (79) (a Laplace transform) to confirm the consistency of the expressions for the canonical partition function and the density of states in the harmonic superposition approximation. $**$

24. Using the superposition approximation for $Z(T)$ in equation (78) calculate the average value of the internal energy and the heat capacity in the canonical ensemble. $**$

25. (i) Write down the partition functions of the liquid and solid using the harmonic superposition approximation in a two-state model of melting: where



$\bar{\nu}_S$ and $\bar{\nu}_L$ are the geometric mean vibrational frequencies of the solid and liquid minima, respectively. All the liquid minima have C_1 symmetry, and the order of the point group of the solid is h_S. Hence, calculate the melting temperature. Do changing the energy gap, and the configurational and vibrational entropy of the liquid have the expected effects on the melting point?

(ii) Using this as a model of the melting transition in LJ$_{13}$, calculate the melting temperature when the symmetry of the icosahedral global minimum is (a) included, and (b) ignored. Use $\Delta E = 2.85\epsilon$, $n_L = 10$ and $\bar{\nu}_S/\bar{\nu}_L = 1.07$. $**$

26. Assuming a Gaussian distribution of the liquid minima show that the average energy of the occupied minima has the following temperature dependence:

$$\langle E^0 \rangle = E_G - \frac{\sigma^2}{kT}, \tag{106}$$

where E_G and σ^2 are the mid-point and the variance of the Gaussian. You should assume that the vibrational frequencies are independent of the energy of the minima.

When this model of the liquid is used to investigate the equilibrium with the solid (as in Q. 25) as well as the melting transition, one can also find a further low temperature transition. To what does this correspond?**