

# A Doubly Nudged Elastic Band Method for Finding Transition States

Semen A. Trygubenko\* and David J. Wales

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

A modification of the nudged elastic band (NEB) method is presented that enables stable optimisations to be run using both the limited-memory quasi-Newton (L-BFGS) and slow-response quenched velocity Verlet (SQVV) minimisers. The performance of this new ‘doubly nudged’ DNEB method is analysed in conjunction with both minimisers and compared with previous NEB formulations. We find that the fastest DNEB approach (DNEB/L-BFGS) can be quicker by up to two orders of magnitude. Applications to permutational rearrangements of the seven-atom Lennard-Jones cluster (LJ<sub>7</sub>) and highly cooperative rearrangements of LJ<sub>38</sub> and LJ<sub>75</sub> are presented. We also outline an updated algorithm for constructing complicated multi-step pathways using successive DNEB runs.

---

\* E-mail: sat39@cam.ac.uk

## I. INTRODUCTION

Locating transition states on a potential energy surface (PES) provides an important tool in the study of dynamics using statistical rate theories [1, 2, 3]. Here we define a transition state according to the geometrical definition of Murrell and Laidler, i.e. as a stationary point with a single negative Hessian eigenvalue [4]. Unfortunately, it is significantly harder to locate transition states than local minima, since the system must effectively ‘balance on a knife-edge’ in one degree of freedom. Many algorithms have been suggested for this purpose, and the most efficient method may depend upon the nature of the system. For example, different considerations probably apply if second derivatives can be calculated relatively quickly, as for many empirical potentials [5]. Transformation to an alternative coordinate system may also be beneficial for systems bound by strongly directional forces [6, 7, 8, 9, 10, 11, 12, 13, 14].

Algorithms to locate transition states can generally be divided into single-ended approaches, which simply require a single starting geometry [11, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55], and double-ended methods, which are usually designed to find a transition state between two endpoints [10, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79]. The result of a single-ended search may be a transition state that is not connected to the starting point by a steepest-descent path, and such methods can be useful for building up databases of stationary points to provide a non-local picture of the potential energy surface, including thermodynamic and dynamic properties [5, 54]. In contrast, double-ended methods are usually designed to characterise a particular rearrangement, and often do not produce a tightly converged geometry for the transition state. However, the resulting structures can be further refined using single-ended strategies, particularly eigenvector-following [15, 17, 18, 19, 20, 23, 24, 34, 35, 45, 53, 55, 80], and this approach has been used in several previous studies [75, 81, 82]. Some recent overviews of the field are available [5, 83], and readers are referred to these publications for further discussion.

Our principal concern in the present work is the development of the double-ended NEB approach [71, 75, 76, 84]. The earliest double-ended methods were probably the linear and quadratic synchronous transit algorithms (LST and QST) [85], which are entirely based on

interpolation between the two endpoints. In LST the highest energy structure is located along the straight line that links the two endpoints. QST is similar in spirit, but approximates the reaction path using a parabola instead of a straight line. Neither interpolation is likely to provide a good estimate of the path except for very simple reactions, but they may nevertheless be useful to generate initial guesses for more sophisticated double-ended methods.

Another approach is to reduce the distance between reactant and product by some arbitrary value to generate an ‘intermediate’, and seek the minimum energy of this intermediate structure subject to certain constraints, such as fixed distance to an endpoint. This is the basis of the ‘Saddle’ optimisation method [86] and the ‘Line Then Plane’ [87] algorithm, which differ only in the definition of the subspace in which the intermediate is allowed to move. The latter method optimises the intermediate in the hyperplane perpendicular to the interpolation line, while ‘Saddle’ uses hyperspheres. The minimised intermediate then replaces one of the endpoints and the process is repeated.

There are also a number of methods that are based on a ‘chain-of-states’ (CS) approach, where several images of the system are somehow coupled together to create an approximation to the required path. The CS methods mainly differ in the way in which the initial guess to the path is refined. In the ‘Chain’ method [88] the geometry of the highest energy image is relaxed first using only the component of the gradient perpendicular to the line connecting its two neighbours. The process is then repeated for the next-highest energy neighbours. The optimisation is terminated when the gradient becomes tangential to the path. The ‘Locally Updated Planes’ method [89] is similar, but the images are relaxed in the hyperplane perpendicular to the reaction coordinate, rather than along the line defined by the gradient, and all the images are moved simultaneously.

The nudged elastic band (NEB) approach introduced some further refinements to these CS methods [84]. It is based on a discretised representation of the path originally proposed by Elber and Karplus [57], with modifications to eliminate corner-cutting and sliding-down problems [71], and to improve the stability and convergence properties [75]. Maragakis *et al.* applied the NEB method to various physical systems ranging from semiconductor materials to biologically relevant molecules. They report that use of powerful minimisation methods in conjunction with NEB approach was unsuccessful [79]. These problems were attributed to instabilities with respect to the extra parameters introduced by the springs. In fact,

the NEB approach has previously been used with the L-BFGS algorithm in other work, where hybrid eigenvector-following techniques were employed to produce tightly converged transition states from guesses obtained by NEB calculations [81, 82]. The main result of the present contribution is a modified ‘doubly nudged’ elastic band (DNEB) method, which is stable when combined with the L-BFGS minimiser. In comparing the DNEB approach with other methods we have also analysed quenched velocity Verlet minimisation, and determined the best point at which to remove the kinetic energy. Extensive tests show that the DNEB/L-BFGS combination provides a significant performance improvement over previous implementations. We therefore outline a new strategy to connect distant minima, which is based on successive DNEB searches to provide transition state candidates for refinement by eigenvector-following.

## II. METHODS

In the present work we used the nudged elastic band [71, 76] (NEB) and eigenvector-following [15, 17, 18, 19, 20, 23, 24, 34, 53, 55, 80] (EF) methods for locating and refining transition states. In the NEB approach the path is represented as a set of images  $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N\}$  that connect the endpoints  $\mathbf{X}_0$  and  $\mathbf{X}_{N+1}$ , where  $\mathbf{X}_i$  is a vector containing the coordinates of image  $i$  (Figure 1) [75]. In the usual framework of double-ended methodologies [90] the endpoints are stationary points on the potential energy surface (PES) (usually minima), which are known in advance. In addition to the true potential,  $V_i$ , which binds the atoms within each image, equivalent atoms in  $N$  adjacent images are interconnected by  $N + 1$  springs according to a parabolic potential,

$$\tilde{V} = \frac{1}{2}k_{spr} \sum_{i=1}^{N+1} |\mathbf{X}_i - \mathbf{X}_{i-1}|^2. \quad (1)$$

Subsequently these potentials will be referred to as the ‘true potential’ and the ‘spring potential’, respectively.

The springs are intended to hold images on the path during optimisation—otherwise they would slide down to the endpoints, or to other intermediate minima [57]. Occasionally, depending on the quality of the initial guess, we have found that some images may converge to higher index stationary points. One could imagine the whole construction as a band or rope that is stretched across the PES, which, if optimised, is capable of closely following a

curve defined in terms of successive minima, transition states, and the intervening steepest-descent paths.

In practice, the above formulation encounters difficulties connected with the coupling between the ‘true’ and ‘spring’ components of the potential. The magnitude of the springs’ interference with the true potential is system dependent and generally gives rise to corner-cutting and sliding-down problems [71]. It is convenient to discuss these difficulties in terms of the components of the true gradient,  $\mathbf{g}$ , and spring gradient,  $\tilde{\mathbf{g}}$ , parallel and perpendicular to the path. The parallel component of the gradient  $\mathbf{g}^{\parallel}$  at image  $i$  on the path is obtained by projecting out the perpendicular component  $\mathbf{g}^{\perp}$  using an estimate of the tangent to the path. The parallel and perpendicular components for image  $i$  are:

$$\mathbf{g}_i^{\parallel} = (\nabla_i V_i \cdot \hat{\boldsymbol{\tau}}_i) \hat{\boldsymbol{\tau}}_i, \quad \mathbf{g}_i^{\perp} = \nabla_i V_i - \mathbf{g}_i^{\parallel}, \quad (2)$$

where  $V_i = V(\mathbf{X}_i)$ , and the unit vector  $\hat{\boldsymbol{\tau}}_i$  is the tangent. Here and throughout this paper we denote unit vectors by a hat. The complete gradient,  $\mathbf{g}$ , has  $N \times \eta$  components for a band of  $N$  images with  $\eta$  atomic degrees of freedom each.

Corner-cutting has a significant effect when a path experiences high curvature. Here  $\tilde{\mathbf{g}}^{\perp}$  is large, which prevents the images from closely following the path because the spring force necessarily has a significant component perpendicular to the tangent. The sliding-down problem occurs due to the presence of  $\mathbf{g}^{\parallel}$ , which perturbs the distribution of images along the path, creating high-resolution regions (around the local minima) and low-resolution regions (near the transition states) [71]. Both problems significantly affect the ability of the NEB method to produce good transition state candidates. We have found that sliding-down and corner-cutting are interdependent and cannot both be remedied by adjusting the spring force constant  $k_{spr}$ ; increasing  $k_{spr}$  may prevent sliding-down but it will make corner-cutting worse.

The aforementioned problems can sometimes be eliminated by constructing the NEB gradient from the potential in the following way:  $\mathbf{g}^{\parallel}$  and  $\tilde{\mathbf{g}}^{\perp}$  are projected out, which gives the elastic band its ‘nudged’ property [76]. Removal of  $\mathbf{g}^{\parallel}$  can be thought of as bringing the path into a plane or flattening the PES [Figure 1(b)], while removal of  $\tilde{\mathbf{g}}^{\perp}$  is analogous to making the images heavier so that they favour the bottom of the valley at all times.

The choice of a method to estimate the tangent to the path is important for it affects the convergence of the NEB calculation. Originally, the tangent vector,  $\hat{\boldsymbol{\tau}}_i$ , for image  $i$

was obtained by normalising the line segment between the two adjacent images,  $i + 1$  and  $i - 1$  [71]:

$$\hat{\boldsymbol{\tau}}_i = \frac{\mathbf{X}_{i+1} - \mathbf{X}_{i-1}}{|\mathbf{X}_{i+1} - \mathbf{X}_{i-1}|}. \quad (3)$$

However, kinks can develop during optimisation of the image chain using this definition of  $\hat{\boldsymbol{\tau}}_i$ . It has been shown [75] that kinks are likely to appear in the regions where the ratio  $g_i^{\parallel}/g_j^{\perp}$  is larger than the length of the line segment,  $|\boldsymbol{\tau}|$ , used in estimating the tangent [Figure 2(a)].

Both the above ratio, the image density and  $|\hat{\boldsymbol{\tau}}|$  can vary depending on the system of interest, the particular pathway and other parameters of the NEB calculation. From equation (3) it can be seen that  $\hat{\boldsymbol{\tau}}_i$ , and, hence, the next step in the optimisation of image  $i$ , is determined by its neighbours, which are not necessarily closer to the path than image  $i$ . Therefore, a better approach in estimating the  $\hat{\boldsymbol{\tau}}_i$  would be to use only one neighbour, since then we only need this neighbour to be better converged than image  $i$ .

There are two neighbours to select from, and it is natural to use the higher-energy one for this purpose, since steepest-descent paths are easier to follow downhill than uphill:

$$\hat{\boldsymbol{\tau}}_i = \frac{(j - i)(\mathbf{X}_j - \mathbf{X}_i)}{|\mathbf{X}_j - \mathbf{X}_i|}, \quad (4)$$

where  $i$  and  $j$  are two adjacent images with energies  $E_i$  and  $E_j$ , and  $E_i < E_j$ . In this way, an image  $i$  that has one higher-energy neighbour  $j$  behaves as if it is ‘hanging’ on to it [Figure 2(b)].

The above tangent formulation requires special handling of extrema along the path, and a mechanism for switching  $\hat{\boldsymbol{\tau}}$  at such points was proposed [75]. It also fails to produce an even distribution of images in regions with high curvature [Figure 2(c)]. We presume that Henkelman and Jónsson substitute  $(\tilde{\mathbf{g}} \cdot \hat{\boldsymbol{\tau}}) \hat{\boldsymbol{\tau}}$  by  $|\tilde{\mathbf{g}}| \hat{\boldsymbol{\tau}}$  in equation (2) to obtain a spring gradient formulation that will keep the images equispaced when the tangent from equation (4) is used in the projections: [76]

$$\tilde{\mathbf{g}}_i^{\parallel} = k_{spr} (|\mathbf{X}_i - \mathbf{X}_{i-1}| - |\mathbf{X}_{i+1} - \mathbf{X}_i|) \hat{\boldsymbol{\tau}}_i. \quad (5)$$

We have previously used the NEB approach to produce candidate transition state guesses for further refinement using hybrid EF methods [81, 82], which avoid either calculating the Hessian or diagonalising it [53, 55]. Having obtained tightly converged transition states

approximate steepest-descent paths are calculated by energy minimisation, as discussed in §III F.

In the present work the NEB approach has been used in combination with two minimisers, namely the quenched velocity Verlet (QVV) and the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithms. The QVV method is based on the velocity Verlet algorithm [91] (VV) as modified by Jónsson *et al.* [71] and was originally used for NEB optimisation. VV is a symplectic integrator that enjoys widespread popularity, primarily in molecular dynamics (MD) simulations where it is used for numerical integration of Newton’s equations of motion. At each time step  $\delta t$  the coordinates and the velocities  $\mathbf{v}$  are updated from the coupled first-order differential equations in the following manner [91]:

$$\mathbf{X}(t + \delta t) = \mathbf{X}(t) + \delta t \mathbf{v}(t) - \frac{\delta t^2}{2m} \mathbf{g}(t) \quad (6)$$

$$\mathbf{v}\left(t + \frac{1}{2}\delta t\right) = \mathbf{v}(t) - \frac{\delta t}{2m} \mathbf{g}(t) \quad (7)$$

$$\mathbf{v}(t + \delta t) = \mathbf{v}\left(t + \frac{1}{2}\delta t\right) - \frac{\delta t}{2m} \mathbf{g}(t + \delta t) \quad (8)$$

The algorithm involves two stages, with a force evaluation in between. First the positions are updated according to equation (6), and the velocities at midpoint  $t + \delta t/2$  are then computed using equation (7). After the evaluation of the gradient at time  $t + \delta t$  the velocity is updated again [equation (8)] to complete the move. To obtain minimisation it is necessary to remove kinetic energy, and this can be done in several ways. If kinetic energy is removed completely every step the algorithm is equivalent to a steepest-descent minimisation, which is rather inefficient. Instead, it was proposed by Jónsson *et al.* [71] to keep only the velocity component that is antiparallel to the gradient at the current step. If the force is consistently pointing in the same direction the system accelerates, which is equivalent to increasing the time step [71]. However, a straightforward variable time step version of the above algorithm was reported to be unsuccessful [92].

L-BFGS is a version of the BFGS algorithm that limits the storage used and is hence particularly suitable for large-scale problems [93]. The difference between the L-BFGS algorithm and standard BFGS is in the Hessian matrix update. In order to store each correction to the Hessian  $2n$  storage locations are needed,  $n$  being the dimensionality of the problem [94]. L-BFGS stores a maximum of  $m$  corrections and the Hessian matrix is never formed explicitly. Every iteration  $-\mathbf{H}^{-1}\mathbf{g}$  is computed according to a recursive formula de-

scribed by Nocedal [94]. For the first  $m$  iterations L-BFGS is identical to the BFGS method, but after these the oldest correction is discarded. Since only the  $m$  most recent corrections are retained L-BFGS uses less storage. Here we employed a modified version of Nocedal’s L-BFGS implementation [95] in which the line search was removed and the maximum step size was limited for each image separately.

It is noteworthy that the objective function corresponding to the projected NEB gradient is unknown, but it is not actually required in either of the minimisation algorithms that we consider.

### III. RESULTS

The springs should distribute the images evenly along the NEB path during the optimisation, and the choice of  $k_{spr}$  must be made at the beginning of each run. It has been suggested by Jónsson and coworkers that since the action of the springs is only felt along the path the value of the spring constant is not critical as long as it is not zero [71]. If  $k_{spr}$  is set to zero then convergence is guaranteed for the first several tens of iterations only; even though  $\mathbf{g}^{\parallel}$  is projected out,  $\hat{\boldsymbol{\tau}}$  fluctuates and further optimisation will eventually result in the majority of images gradually sliding down to local minima [71].

#### A. Slow-response Quenched Velocity Verlet

In practice we find that the value of  $k_{spr}$  affects the convergence properties and the stability of the optimisation process. This result depends on the type of minimiser employed and may also depend on minimiser-specific settings. Here we analyse the convergence properties of NEB minimisations using the QVV minimiser (NEB/QVV) and their dependence on the type of velocity quenching. From previous work it is not clear when is the best time to perform quenching during the MD minimisation of the NEB [71, 75, 76]. Since the VV algorithm calculates velocities based on the gradients at both current and previous steps quenching could be applied using either of these gradients.

Specifically, it is possible to quench velocities right after advancing the system using equation (6), at the half-step in the velocity evaluation (quenching intermediate velocities at time  $t + \delta t/2$ ) using either the old or new gradient [equation (7)], or after completion of the



velocity update. In Figure 3 we present results for the stability of NEB/QVV as a function of the force constant parameter for three of these quenching approaches. We will refer to an NEB optimisation as stable for a certain combination of parameters (e.g. time integration step, number of images) if the NEB steadily converges to a well-defined path and/or stays in its proximity until the maximal number of iterations is reached or the convergence criterion is satisfied.

Figure 3 shows the results of several thousand optimisations for a 17-image band with the Müller-Brown two-dimensional potential [96] using QVV minimisation and a time step of 0.01 (consistent units) for different values of  $k_{spr}$ . This widely used surface does not present a very challenging or realistic test case, but if an algorithm does not behave well for this system it is unlikely to be useful. Each run was started from the initial guess obtained using linear interpolation and terminated when the root-mean-square (RMS) gradient became less than 0.01. We define the RMS gradient for the NEB as

$$g_{\text{RMS}}^{\perp} = \sqrt{\frac{\sum_{i=1}^N |\mathbf{g}_i^{\perp}|}{N\eta}} \quad (9)$$

where  $N$  is the number of images in the band and  $\eta$  is the number of atomic degrees of freedom available to each image.

It seems natural to remove the velocity component perpendicular to the gradient at the current point when the geometry  $\mathbf{X}(t)$ , gradient  $\mathbf{g}(t)$  and velocity  $\mathbf{v}(t)$  are available, i.e.

$$\mathbf{v}_Q(t) = \left( \mathbf{v}(t) \cdot \hat{\mathbf{g}}(t) \right) \hat{\mathbf{g}}(t), \quad (10)$$

where  $\mathbf{v}_Q(t)$  is the velocity vector after quenching. However, we found this approach to be the least stable of all—the optimisation was slow and convergence was very sensitive to the magnitude of time step. Hence we do not show any results for this type of quenching.

From Figure 3(a) we see that the best approach is to quench the velocity after the coordinate update. The optimisation is then stable for a wide range of force constant values, and the images on the resulting pathway are evenly distributed. In this quenching formulation the velocity response to a new gradient direction is retarded by one step in coordinate space: the step is still taken in the direction  $\mathbf{v}(t)$  but the corresponding velocity component is removed. To implement this slow-response QVV (SQVV) it is necessary to modify the VV algorithm described in section III A by inserting equation (10) in between the two stages described by equations (6) and (7).

The second best approach after SQVV is to quench the velocity at midstep  $t + \delta t/2$  using the new gradient. On average, this algorithm takes twice as long to converge the NEB to a given RMS gradient tolerance compared to SQVV. However, the method is stable for the same range of spring force constant values and produces a pathway in which the images are equispaced more accurately than the other formulations [see Figure 3(b)].

The least successful of the three QVV schemes considered involves quenching velocities at mid-step using the gradient from the previous iteration (stars in Figure 3). Even though the number of iterations required is roughly comparable to that obtained by quenching using the new gradient, it has the smallest range of values for the force constant where it is stable. Some current implementations of NEB [97, 98] (intended for use in combination with electronic structure codes) use this type of quenching in their QVV implementation.

We have also conducted analogous calculations for more complicated systems such as permutational rearrangements of Lennard-Jones clusters. The results are omitted for brevity, but agree with the conclusions drawn from the simpler 2D model described above. The same is true for the choice of force constant investigated in the following section.

## B. Choice of the Force Constant

We find that if the force constant is too small many more iterations are needed to converge the images to the required RMS tolerance, regardless of the type of quenching. In addition, the path exhibits a more uneven image distribution. This result occurs because at the initial stage the images may have very different gradients from the true potential along the band, because they lie far from the required path, and the true potential gradient governs the optimisation. When the true RMS force is reduced the springs start to play a more important role. But at this stage the forces are small and so is the QVV step size. The influence of the springs is actually most important during the initial optimisation stage, for it can determine the placement of images in appropriate regions. It is less computationally expensive to guide an image into the right region at the beginning of an optimisation than to restore the distribution afterwards by dragging it between two minima through a transition state region.

If  $k_{spr}$  is too big the NEB never converges to the required RMS gradient tolerance value. Instead, it stays in proximity to the path but develops oscillations: adjacent images start

to move in opposite directions. For all types of quenching we observed similar behaviour when large values of the force constant were used. This problem is related to the step in coordinate space that the optimiser is taking: for the SQVV case simply decreasing the time step remedies this problem.

### C. Comparison of SQVV and L-BFGS Minimisers for the MB Surface

We tested the NEB/L-BFGS method by minimising a 17-image NEB for the two-dimensional Müller-Brown surface [96]. Our calculations were carried out using the OPTIM program [99]. The NEB method in its previous formulation [76] and a modified L-BFGS minimiser [5] were implemented in OPTIM in a previous discrete path sampling study [81]. We used the same number of images, initial guess and termination criteria as described in section III A to make the results directly comparable.

Figure 4 shows the performance of the L-BFGS minimiser as a function of  $k_{spr}$ . We used the following additional L-BFGS specific settings. The number of corrections in the BFGS update was set to  $m = 4$  (Nocedal’s recommendation for the number of corrections is  $3 \leq m \leq 7$  [95]), the maximum step size was 0.1, and we limited the step size for each image separately, i.e.

$$|\mathbf{p}_j| \leq 0.1, \tag{11}$$

where  $\mathbf{p}_j$  is the step for image  $j$ . The diagonal elements of the inverse Hessian were initially set to 0.1.

From Figure 4 it can be seen that the performance of L-BFGS minimisation is relatively independent of the choice of force constant. All the optimisations with  $30 \leq k_{spr} \leq 10,000$  converged to the steepest-descent path, and, for most of this range, in less than 100 iterations. This method therefore gives roughly an order of magnitude improvement in speed over SQVV minimisation [see Figure 3(a)].

We found it helpful to limit the step size while optimising the NEB with the L-BFGS minimiser. The magnitude and direction of the gradient on adjacent images can vary significantly. Taking bigger steps can cause the appearance of temporary discontinuities and kinks in the NEB. The NEB still converges to the correct path, but it takes a while for these features to disappear and the algorithm does not converge any faster.

### D. Doubly Nudged Elastic Bands

The NEB/QVV approach has previously been systematically tested on systems with around 100 degrees of freedom,  $\eta$  [79]. However, in the majority of cases these test systems could be divided into a ‘core’ and a smaller part that actually changes significantly. The number of active degrees of freedom is therefore significantly smaller than the total number in these tests. For example, prototropic tautomerisation of cytosine nucleic acid base ( $\eta = 33$ ) involves motion of one hydrogen atom along a quasi-rectilinear trajectory accompanied by a much smaller distortion of the core.

We have therefore tested the performance of the NEB/SQVV and NEB/L-BFGS schemes for more complicated rearrangements of Lennard-Jones (LJ) clusters to validate the results of §III B, and to investigate the stability and performance of both approaches when there are more active degrees of freedom. Most of our test cases involve permutational isomerisation of the LJ<sub>7</sub>, LJ<sub>38</sub> and LJ<sub>75</sub> clusters. These examples include cases with widely varying separation between the endpoints, integrated path length, number of active degrees of freedom and cooperativity.

Permutational rearrangements are particularly interesting because it is relatively difficult to produce an initial guess for the NEB run. In contrast, linear interpolation between the endpoints was found to provide a useful initial guess for a number of simpler cases [71]. For example, it was successfully used to construct the NEB for rearrangements that involve one or two atoms following approximately rectilinear trajectories, and for migration of a single atom on a surface [79]. For more complex processes an alternative approach adopted in previous work is simply to supply a better initial guess ‘by hand’, e.g. construct it from the images with unrelaxed geometries containing no atom overlaps [79]. The ‘detour’ algorithm described in previous calculations that employ the ridge method could also be used to avoid ‘atom-crashing’ in the initial interpolation [62].

It has previously been suggested that it is important to eliminate overall rotation and translation (ORT) of each image during the optimisation of an NEB [71]. We have implemented this constraint in the same way as Jónsson *et al.*, by freezing one atom, restricting the motion of a second atom to a plane, and constraining the motion of a third atom to a line by zeroing the appropriate components of the NEB gradient.

We were able to obtain stable convergence in NEB/L-BFGS calculations only for simple

rearrangements, which confirms that straightforward L-BFGS optimisation of the NEB is unstable [79]. Figure 5 shows the performance of the NEB/SQVV [Figure 5(a)] and NEB/L-BFGS [Figure 5(b)] approaches for one such rearrangement. These calculations were carried out using a 7-image NEB both with (diamonds) and without (stars) removing ORT for isomerisation of an LJ<sub>7</sub> cluster (global minimum  $\rightarrow$  second-lowest minimum). The number of iterations,  $\ell$ , is proportional to the number of gradient evaluations regardless of the type of minimiser. Hence, from Figure 5 we conclude that for this system NEB/L-BFGS is faster than NEB/SQVV by approximately two orders of magnitude. However, removal of ORT leads to instability in the NEB/L-BFGS optimisation: the images do not stay in proximity to the required path for long and instead diverge from it [see inset in Figure 5(b)].

By experimentation we have found that the main source of the instabilities is the complete removal of  $\tilde{\mathbf{g}}^\perp$ . Instead, the inclusion of some portion of  $\tilde{\mathbf{g}}^\perp$  in the NEB gradient, i.e.

$$\mathbf{g}_{\text{NEB}} = \mathbf{g}^\perp + \tilde{\mathbf{g}}^\parallel + \tilde{\mathbf{g}}^*, \quad (12)$$

where  $\tilde{\mathbf{g}}^* = \xi \tilde{\mathbf{g}}^\perp$ , makes the NEB/L-BFGS calculations stable but introduces some additional corner-cutting, as well as an extra parameter,  $\xi$ . Since we use the transition state candidates from NEB as starting points for further EF calculations the corner-cutting is not a drawback as long as the transition state candidates are good enough. By adjusting  $\xi$  in the range of (0.01, 0.1) we were able to achieve satisfactory performance for the NEB/L-BFGS method in a number of cases. However, an alternative modification, described below, proved to be even more successful.

The drawback of the NEB gradient described by equation (12) stems from the interference of  $\mathbf{g}^\perp$  and  $\xi \tilde{\mathbf{g}}^\perp$ , and becomes particularly noticeable when the projection of  $\xi \tilde{\mathbf{g}}^\perp$  on  $\mathbf{g}^\perp$  and  $\mathbf{g}^\perp$  itself are of comparable magnitude. This problem is analogous to the interference of  $\mathbf{g}$  and  $\tilde{\mathbf{g}}$  in the original elastic band method, which was previously solved by ‘nudging’ [76]. We have therefore constructed the gradient of a new ‘doubly’ nudged elastic band (DNEB) using

$$\tilde{\mathbf{g}}^* = \tilde{\mathbf{g}}^\perp - (\tilde{\mathbf{g}}^\perp \cdot \hat{\mathbf{g}}^\perp) \hat{\mathbf{g}}^\perp. \quad (13)$$

In this formulation some corner-cutting may still occur because the images tend to move cooperatively during optimisation; the spring gradient  $\tilde{\mathbf{g}}_{\text{DNEB}}^\perp$  acting on one image can still indirectly interfere with the true gradients of its neighbours. In our calculations this drawback was not an issue, since we are not interested in estimating properties of the path directly

from its discrete representation. Instead we construct it from steepest-descent paths calculated after converging the transition states tightly using the EF approach. We have found DNEB perfectly adequate for this purpose.

We have also tested a number of approaches that might be useful if one wants to produce a full pathway involving a number of transition states for a complicated rearrangement in just one NEB run. One of these, for instance, is a gradual removal of the  $\tilde{\mathbf{g}}^*$  component from the NEB gradient once some convergence criterion is achieved. This removal works remarkably well, particularly in situations with high energy initial guesses, which occur frequently if the guessing is fully automated. This adjustment can be thought of as making the band less elastic in the beginning in order to resolve the highest-energy transition state regions first.

### **E. Comparison of the DNEB/L-BFGS and DNEB/SQVV Methods for Permutational Isomerisations of LJ<sub>7</sub>**

It is sometimes hard to make a direct comparison of different double-ended methods for a particular rearrangement because the calculations may converge to different paths. Another problem concerns the choice of a consistent termination criterion: the RMS force usually converges to some finite system-dependent value, which in turn may depend on the number of images and other parameters. A low-energy chain of NEB images does not necessarily mean that a good pathway has been obtained, since it may arise because more images are associated with regions around local minima, rather than the higher energy transition state regions. Here we present the results of DNEB/L-BFGS, DNEB/SQVV and, where possible, NEB/SQVV calculations for all the distinct permutational rearrangements of the global minimum for the LJ<sub>7</sub> cluster (see Figure 6 for the endpoints and nomenclature).

It is possible to draw a firm conclusion as to how well the NEB represents the pathway when the corresponding stationary points and steepest-descent paths are already known. We therefore base our criterion for the effectiveness of an NEB calculation on whether we obtain good estimates of all the transition states. By considering several systems of increasing complexity we hope to obtain comparisons that are not specific to a particular pathway.

Connections between two minima are defined by calculating an approximation to the two steepest-descent paths that lead downhill from each transition state, and two transition

states are considered connected if they are linked to the same minimum via a steepest-descent path. We will say that minima are ‘connected’ if there exists a path consisting of one or more transition states and intermediate minima linking them. Permutational isomers of the same minimum are distinguished in these calculations. We refer to the chain of images produced by the NEB calculation as ‘connected’ if going downhill from each transition state using steepest-descent minimisation yields a set of minima that contains the endpoints linked together.

For NEB/SQVV calculations we used the NEB formulation defined in Ref. 76. DNEB is different from the above method because it includes an additional component in the NEB gradient, as described by equations (12) and (13). In addition, for the following DNEB calculations we did not remove overall rotation and translation (ORT), because we believe it is unnecessary when our gradient modification is used. To converge transition state candidates tightly we employed EF optimisation, limiting the maximum number of EF iterations to five with an RMS force convergence tolerance of  $10^{-5}$ . (Standard reduced units for the Lennard-Jones potential are used throughout this paper.) Initial guesses for all the following calculations were obtained by linear interpolation between the endpoints. To prevent ‘atom-crashing’ from causing overflow in the initial guess we simply perturbed such images slightly using random atomic displacements of order  $10^{-2}$  reduced units.

In each case we first minimised the Euclidean distance between the endpoints with respect to overall rotation and translation using the method described in Ref. 100. SQVV minimisation was performed with a time step of 0.01 and a maximum step size per degree of freedom of 0.01. This limit on the step size prevents the band from becoming ‘discontinuous’ initially and plays an important role only during the first 100 or so iterations. The limit was necessary because for the cases when the endpoints are permutational isomers linear interpolation usually yields bands with large gradients, and it is better to refrain from taking excessive steps at this stage. We did not try to select low energy initial guesses for each rearrangement individually, since one of our primary concerns was to automate this process. For the same reason, all the L-BFGS optimisations were started from guesses preoptimised using SQVV until the RMS force dropped below 2.0.

Table I shows the minimum number of images and gradient calls required to produce a connected pathway using the DNEB/L-BFGS and DNEB/SQVV methods. These calculations were run assuming no prior knowledge of the path. Normally there is no initial

information available on the integrated path length or the number of intermediate minima between the endpoints, and it takes some experimentation to select an appropriate number of images. Our strategy is therefore to gradually increase the number of images to make the problem as computationally inexpensive as possible. Hence we increment the number of images and maximum number of NEB iterations in each calculation until a connected path is produced, in the sense defined above. The permitted image range was  $2 \leq N \leq 20$  and the maximum number of NEB iterations ranged from  $1 \leq l \leq 3,000N$ . We were unable to obtain connected pathways for any of the four LJ<sub>7</sub> rearrangements using the NEB/SQVV approach.

Table II presents the results of analogous calculations where we keep the number of images fixed to 50. Unlike the performance comparison where the number of images is kept to a minimum (Table I), these results should provide insight into the performance of the DNEB approach when there are sufficient images to resolve all the transition states. All the optimisations for a particular rearrangement converged to the same or an enantiomeric pathway unless stated otherwise. The energy profiles that correspond to these rearrangements are shown in Figure 7.

From Table I and II we conclude that in all cases the DNEB/L-BFGS approach is more than an order of magnitude faster than DNEB/SQVV. It is also noteworthy that the DNEB/SQVV approach is faster than NEB/SQVV because overall rotation and translation are not removed. Allowing the images to rotate or translate freely can lead to numerical problems, namely a vanishing norm for the tangent vector, when the image density is very large or the spring force constant is too small. However, when overall rotation and translation are not allowed there is less scope for improving a bad initial guess, because the images are more constrained. This constraint usually means that more images are needed or a better initial guess is required. Our experience is that such constraints usually slow down convergence, depending on which degrees of freedom are frozen: if these are active degrees of freedom (see above) the whole cluster must move instead, which is usually a slow, concerted multi-step process.



### F. A revised connection algorithm

In previous work we have used the NEB approach to supply transition state guesses for further EF refinement [81, 82]. Double-ended searches are needed in these discrete path sampling runs to produce alternative minimum–transition state–minimum... sequences from an initial path. The end minima that must be linked in such calculations may be separated by relatively large distances, and a detailed algorithm was described for building up a connected path using successive transition state searches. The performance of the DNEB/L-BFGS approach is sufficiently good that we have changed this connection strategy in our OPTIM program. In particular, the DNEB/L-BFGS method can often provide good candidates for more than one transition state at a time, and may even produce all the necessary transition states on a long path. However, it is still generally necessary to consider multiple searches between different minima in order to connect a pair of endpoints. In particular, we would like to use the minimum number of NEB images possible for reasons of efficiency, but automate the procedure so that it eventually succeeds or gives up after an appropriate effort for any pair of minima that may arise in a discrete path sampling run. These calculations may involve the construction of many thousands of discrete paths. As in previous work we converge the NEB transition state candidates using eigenvector-following techniques and then use L-BFGS energy minimisation to calculate approximate steepest-descent paths. These paths usually lead to local minima, which we also converge tightly. The combination of NEB and hybrid eigenvector-following techniques [53, 55] is similar to using NEB with a ‘climbing image’ as described in Ref. 75

The initial parameters assigned to each DNEB run are the number of images and the number of iterations, which we specify by image and iteration densities. The iteration density is the maximum number of iterations per image, while the image density is the maximum number of images per unit distance. The distance in question is the Euclidean separation of the endpoints, which provides a crude estimation of the integrated path length. This approach is based on the idea that knowing the integrated path length, which means knowing the answer before we start, we could have initiated each DNEB run with the same number of images per unit of distance along the path. In general it is also impossible to provide a lower bound on the number of images necessary to fully resolve the path, since this would require prior knowledge of the number of intervening stationary points. Our

experience suggests that a good strategy is to employ as small an image and iteration density as possible at the start of a run, and only increase these parameters for connections that fail.

All NEB images,  $i$ , for which  $E_i > E_{i\pm 1}$  are considered for further EF refinement. The resulting distinct transition states are stored in a database and the corresponding energy minimised paths were used to identify the minima that they connect. New minima are also stored in a database, while for known minima new connections are recorded. Consecutive DNEB runs aim to build up a connected path by progressively filling in connections between the endpoints or intermediate minima to which they are connected. This is an advantageous strategy because the linear interpolation guesses usually become better as the separation decreases, and therefore fewer optimisation steps are required. Working with sections of a long path one at a time is beneficial because it allows the algorithm to increase the resolution only where it is needed. Our experience is that this approach is generally significantly faster than trying to characterise the whole of a complex path with a single chain of images.

When an overall path is built up using successive DNEB searches we must select the two endpoints for each new search from the database of known minima. It is possible to base this choice on the order in which the transition states were found, which is basically the strategy used in our previous work [81, 82]. However, when combined with the new DNEB approach a better strategy is to connect minima based upon their Euclidean separation. For this purpose it is convenient to classify all the minima into those already connected to the starting endpoint (the S set), the final endpoint (the F set), and the remaining minima, which are not connected to either endpoint (the U set). The endpoints for the next DNEB search are then chosen as the two that are separated by the shortest distance, where one belongs to S or F, and the other belongs to a different set. The distance between these endpoints is then minimised with respect to overall rotation and translation, and an initial guess for the image positions is obtained using linear interpolation. Further details of the implementation of this algorithm and the OPTIM program are available from authors upon request.

As test cases for this algorithm we have considered various degenerate rearrangements of LJ<sub>7</sub>, LJ<sub>13</sub>, LJ<sub>38</sub> and LJ<sub>75</sub>. (A degenerate rearrangement is one that links permutational isomers of the same structure [5, 101].) The PES's of LJ<sub>38</sub> and LJ<sub>75</sub> have been analysed in a number of previous studies [102, 103, 104, 105], and are known to exhibit a double-

funnel morphology: for both clusters the two lowest-energy minima are structurally distinct and well separated in configuration space. This makes them useful benchmarks for the above connection algorithm. Figure 8 depicts the energy profiles obtained using the revised connection algorithm for rearrangements between the two lowest minima of each cluster. In each case we have considered two distinct paths that link different permutational isomers of the minima in question, and these were chosen to be the permutations that give the shortest Euclidean distances. These paths will be identified using the distance between the two endpoints; for example, in the case of LJ<sub>38</sub> we have paths LJ<sub>38</sub> 3.274  $\sigma$  and LJ<sub>38</sub> 3.956  $\sigma$ , where  $2^{1/6}\sigma$  is the pair equilibrium separation for the LJ potential.

For each calculation we used the following settings: the initial image density was set to 10 and the iteration density to 30. If a connection failed for a particular pair of minima then up to two more attempts were allowed before moving to the pair with the next smallest separation. For the second and third attempts the number of images was increased by 50% each time. The maximum number of EF optimisation steps was set to 30 with RMS force convergence criterion of  $10^{-5}$ . In Figure 8 every panel is labelled with the separation between the endpoints, the number of transition states in the final pathway, the number of DNEB runs required, and the total number of gradient calls.

Individual pathways involving a single transition state have been characterised using indices such as

$$\tilde{N} = \frac{(\sum_t |\mathbf{X}_t(S) - \mathbf{X}_t(F)|^2)^2}{\sum_t |\mathbf{X}_t(S) - \mathbf{X}_t(F)|^4}, \quad (14)$$

which is a measure of the number of atoms that participate in the rearrangement. Here  $\mathbf{X}_t(S)$  and  $\mathbf{X}_t(F)$  are the position vectors of atom  $t$  in the starting and finishing geometries, respectively. The largest values are marked in Figure 8 next to the corresponding transition state. It is noteworthy that the pathways LJ<sub>38</sub> 3.956 $\sigma$  and LJ<sub>75</sub> 4.071 $\sigma$  both involve some highly cooperative steps, and the average value of  $\tilde{N}$  is more than 12 for both of them.

We have found that it is usually easier to locate good transition state candidates for a multi-step path if the stationary points are separated by roughly equal distances, in terms of the integrated path length. Furthermore, it seems that more effort is needed to characterise a multi-step path when transition states involving very different path lengths are present. In such cases it is particularly beneficial to build up a complete path in stages. To further characterise this effect we introduce a path length asymmetry index  $\pi$  defined as  $\pi = |s_+ -$

$s_-/(s_+ + s_-)$ , where  $s_+$  and  $s_-$  are the two integrated path lengths corresponding to the two downhill steepest-descent paths from a given transition state. For example, in rearrangement LJ<sub>38</sub>  $3.956\sigma$ , five steps out of nine have  $\pi > 0.5$ .

Barrier asymmetry also plays a role in the accuracy of the tangent estimate, the image density required to resolve particular regions of the path, and in our selection process for transition state candidates, which is based on the condition  $E_i > E_{i\pm 1}$ . To characterise this property we define a barrier asymmetry index,  $\beta$ , as  $\beta = |E_+ - E_-|/\max(E_+, E_-)$ , where  $E_+$  and  $E_-$  are the barriers corresponding to the forward and reverse reactions, respectively. The test cases in Figure 8 include a variety of situations, with barrier asymmetry index  $\beta$  ranging from 0.004 to 1.000. The maximum values of  $\pi$  and  $\beta$  are shown next to the corresponding transition states in this Figure.

We note that the total number of gradient evaluations required to produce the above paths could be reduced significantly by optimising the DNEB parameters or the connection strategy in each case. However, our objective was to find parameters that give reasonable results for a range of test cases, without further intervention.

#### IV. CONCLUSIONS

The most important result of this work is probably the doubly nudged elastic band (DNEB) formulation, in which a portion of the spring gradient perpendicular to the path is retained. With this modification we found that L-BFGS minimisation of the images is stable, thus providing a significant improvement in efficiency. Constraints such as elimination of overall rotation and translation are not required, and the DNEB/L-BFGS method has proved to be reliable for relatively complicated cooperative rearrangements in a number of clusters.

In comparing the performance of the L-BFGS and quenched velocity Verlet (QVV) methods for optimising chains of images we have also investigated a number of alternative QVV schemes. We found that the best approach is to quench the velocity after the coordinate update, so that the velocity response to the new gradient lags one step behind the coordinate updates. However, this slow-response QVV (SQVV) method does not appear to be competitive with L-BFGS.

Finally, we have revised our previous scheme for constructing connections between distant

minima using multiple transition state searches. Previously we have used an NEB/L-BFGS framework for this purpose, with eigenvector-following refinement of transition state candidates and characterisation of the connected minima using energy minimised approximations to the steepest-descent paths [81, 82]. When the DNEB/L-BFGS approach is used we have found that it is better to spend more effort in the DNEB phase of the calculation, since a number of good transition state guesses can often be obtained even when the number of images is relatively small. In favourable cases a complete path linking the required endpoints may be obtained in one cycle. Of course, this was always the objective of the NEB approach [71, 75, 76, 84], but we have not been able to achieve such results reliably for complex paths without the current modifications. When a number of transition states are involved we still find it more efficient to build up the overall path in stages, choosing endpoints that become progressively closer in space. This procedure has been entirely automated within the OPTIM program, which can routinely locate complete paths for highly cooperative multi-step rearrangements, such as those connecting different morphologies of the LJ<sub>38</sub> and LJ<sub>75</sub> clusters.

## V. ACKNOWLEDGEMENTS

S.A.T. is a Cambridge Commonwealth Trust/Cambridge Overseas Trust scholar. Most of the calculations were performed using computational facilities funded by the Isaac Newton Trust.

- 
- [1] H. Eyring, *Chem. Rev.* **17**, 65 (1935).
  - [2] M. G. Evans and M. Polanyi, *Trans. Faraday Soc.* **31**, 875 (1935).
  - [3] K. J. Laidler, *Chemical Kinetics* (Harper & Row, New York, 1987).
  - [4] J. N. Murrell and K. J. Laidler, *Trans. Faraday Soc.* **64**, 371 (1968).
  - [5] D. J. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses* (Cambridge University Press, 2003).
  - [6] P. Pulay, G. Fogorasi, F. Pang, and J. E. Boggs, *J. Am. Chem. Soc.* **101**, 2550 (1979).
  - [7] G. Fogorasi, X. Zhou, P. W. Taylor, and P. Pulay, *J. Am. Chem. Soc.* **114**, 8191 (1992).
  - [8] P. Pulay and G. Fogorasi, *J. Chem. Phys.* **96**, 2856 (1996).

- [9] J. Baker, A. Kessi, and B. Delley, *J. Chem. Phys.* **105**, 192 (1996).
- [10] C. Peng, P. Y. Ayala, H. B. Schlegel, and M. J. Frisch, *J. Comp. Chem.* **17**, 49 (1996).
- [11] J. Baker, D. Kinghorn, and P. Pulay, *J. Chem. Phys.* **110**, 4986 (1999).
- [12] S. R. Billeter, A. J. Turner, and W. Thiel, *Phys. Chem. Chem. Phys.* **2**, 2177 (2000).
- [13] B. Paizs, J. Baker, S. Suhai, and P. Pulay, *J. Chem. Phys.* **113**, 6566 (2000).
- [14] V. Bakken and T. Helgaker, *J. Chem. Phys.* **117**, 9160 (2002).
- [15] G. M. Crippen and H. A. Scheraga, *Arch. Biochem. Biophys.* **144**, 462 (1971).
- [16] J. W. McIver and A. Komornicki, *J. Am. Chem. Soc.* **94**, 2625 (1972).
- [17] J. Pancíř, *Coll. Czech. Chem. Comm.* **40**, 1112 (1974).
- [18] R. L. Hilderbrandt, *Comput. Chem* **1**, 179 (1977).
- [19] C. J. Cerjan and W. H. Miller, *J. Chem. Phys.* **75**, 2800 (1981).
- [20] J. Simons, P. Jørgenson, H. Taylor, and J. Ozment, *J. Phys. Chem.* **87**, 2745 (1983).
- [21] D. O'Neal, H. Taylor, and J. Simons, *J. Phys. Chem.* **88**, 1510 (1984).
- [22] S. Bell and J. S. Crighton, *J. Chem. Phys.* **80**, 2464 (1984).
- [23] A. Banerjee, N. Adams, J. Simons, and R. Shepard, *J. Phys. Chem.* **89**, 52 (1985).
- [24] J. Baker, *J. Comp. Chem.* **7**, 385 (1986).
- [25] D. T. Nguyen and D. A. Case, *J. Phys. Chem.* **89**, 4020 (1985).
- [26] J. Baker, *J. Comp. Chem.* **8**, 563 (1987).
- [27] M. C. Smith, *Int. J. Quant. Chem.* **37**, 773 (1990).
- [28] J. Nichols, H. Taylor, P. Schmidt, and J. Simons, *J. Chem. Phys.* **92**, 340 (1990).
- [29] J. Baker and W. H. Hehre, *J. Comp. Chem.* **12**, 606 (1991).
- [30] J. Baker, *J. Comp. Chem.* **13**, 240 (1992).
- [31] C. J. Tsai and K. D. Jordan, *J. Phys. Chem.* **97**, 11227 (1993).
- [32] D. J. Wales, *J. Chem. Soc., Faraday Trans.* **89**, 1305 (1993).
- [33] S. F. Chekmarev, *Chem. Phys. Lett.* **227**, 354 (1994).
- [34] D. J. Wales, *J. Chem. Phys.* **101**, 3750 (1994).
- [35] D. J. Wales and J. Uppenbrink, *Phys. Rev. B* **50**, 12342 (1994).
- [36] J.-Q. Sun and K. Ruedenberg, *J. Chem. Phys.* **101**, 2157 (1994).
- [37] J.-Q. Sun and K. Ruedenberg, *J. Chem. Phys.* **101**, 2168 (1994).
- [38] F. Jensen, *J. Chem. Phys.* **102**, 6706 (1995).
- [39] J. M. Bofill and M. Comajuan, *J. Comp. Chem.* **16**, 1326 (1995).

- [40] Q. Zhao and J. B. Nicholas, *J. Chem. Phys.* **104**, 767 (1996).
- [41] W. Quapp, *Chem. Phys. Lett.* **253**, 286 (1996).
- [42] J. Baker and F. Chan, *J. Comp. Chem.* **17**, 888 (1996).
- [43] T. R. Walsh and D. J. Wales, *J. Chem. Soc., Faraday Trans.* **92**, 2505 (1996).
- [44] P. Y. Ayala and H. B. Schlegel, *J. Chem. Phys.* **107**, 375 (1997).
- [45] L. J. Munro and D. J. Wales, *Faraday Discuss.* **106**, 409 (1997).
- [46] A. Ulitsky and D. Shalloway, *J. Chem. Phys.* **106**, 10099 (1997).
- [47] B. Paizs, G. Fogarasi, and P. Pulay, *J. Chem. Phys.* **109**, 6571 (1998).
- [48] O. Farkas and H. B. Schlegel, *J. Chem. Phys.* **109**, 7100 (1998).
- [49] N. Mousseau and G. T. Barkema, *Phys. Rev. E* **57**, 2419 (1998).
- [50] H. Goto, *Chem. Phys. Lett.* **292**, 254 (1998).
- [51] W. Quapp, M. Hirsch, O. Imig, and D. Heidrich, *J. Comp. Chem.* **19**, 1087 (1998).
- [52] O. Farkas and H. B. Schlegel, *J. Chem. Phys.* **111**, 10806 (1999).
- [53] L. J. Munro and D. J. Wales, *Phys. Rev. B* **59**, 3969 (1999).
- [54] D. J. Wales, J. P. K. Doye, M. A. Miller, P. N. Mortenson, and T. R. Walsh, *Adv. Chem. Phys.* **115**, 1 (2000).
- [55] Y. Kumeda, L. J. Munro, and D. J. Wales, *Chem. Phys. Lett.* **341**, 185 (2001).
- [56] L. R. Pratt, *J. Chem. Phys.* **85**, 5045 (1986).
- [57] R. Elber and M. Karplus, *Chem. Phys. Lett.* **139**, 375 (1987).
- [58] R. S. Berry, H. L. Davis, and T. L. Beck, *Chem. Phys. Lett.* **147**, 13 (1988).
- [59] R. Czerminski and R. Elber, *J. Chem. Phys.* **92**, 5580 (1990).
- [60] S. Fischer and M. Karplus, *Chem. Phys. Lett.* **194**, 252 (1992).
- [61] L. L. Stachó and M. I. Bán, *Theor. Chim. Acta* **83**, 433 (1992).
- [62] I. V. Ionova and E. A. Carter, *J. Chem. Phys.* **98**, 6377 (1993).
- [63] L. L. Stachó and M. I. Bán, *Theor. Chim. Acta* **84**, 535 (1993).
- [64] G. Dömötör, M. I. Bán, and L. L. Stachó, *J. Comput. Chem.* **14**, 1491 (1993).
- [65] C. Peng and H. B. Schlegel, *Israeli J. Chem.* **33**, 449 (1993).
- [66] A. Matro, D. L. Freeman, and J. D. Doll, *J. Chem. Phys.* **101**, 10458 (1994).
- [67] O. S. Smart, *Chem. Phys. Lett.* **222**, 503 (1994).
- [68] M. I. Bán, G. Dömötör, and L. L. Stachó, *J. Mol. Struct. (Theochem)* **311**, 29 (1994).
- [69] G. Mills and H. Jónsson, *Phys. Rev. Lett.* **72**, 1124 (1994).

- [70] I. V. Ionova and E. A. Carter, *J. Chem. Phys.* **103**, 5437 (1995).
- [71] H. Jónsson, G. Mills, and W. Jacobsen, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne, G. Ciccotti, and D. F. Coker (World Scientific, 1998), p. 385.
- [72] L. L. Stachó, G. Dömötör, and M. I. Bán, *Chem. Phys. Lett.* **311**, 328 (1999).
- [73] R. Elber and M. Karplus, *Chem. Phys. Lett.* **311**, 335 (1999).
- [74] G. Henkelman and H. Jónsson, *J. Chem. Phys.* **111**, 7010 (1999).
- [75] G. Henkelman, B. P. Uberuaga, and H. Jónsson, *J. Chem. Phys.* **113**, 9901 (2000).
- [76] G. Henkelman and H. Jónsson, *J. Chem. Phys.* **113**, 9978 (2000).
- [77] G. Henkelman and H. Jónsson, *J. Chem. Phys.* **115**, 9657 (2001).
- [78] R. A. Miron and K. A. Fichthorn, *J. Chem. Phys.* **115**, 8742 (2001).
- [79] P. Maragakis, S. A. Andreev, Y. Brumer, D. R. Reichman, and E. Kaxiras, *J. Chem. Phys.* **117**, 4651 (2002).
- [80] D. J. Wales and T. R. Walsh, *J. Chem. Phys.* **105**, 6957 (1996).
- [81] D. J. Wales, *Mol. Phys.* **100**, 3285 (2002).
- [82] D. A. Evans and D. J. Wales, *J. Chem. Phys.* **119**, 9947 (2003).
- [83] H. B. Schlegel, *J. Comp. Chem.* **24**, 1514 (2003).
- [84] G. Henkelman, G. Johannesson, and H. Jónsson, in *Progress in Theoretical Chemistry and Physics*, edited by S. D. Schwartz (Kluwer Academic Publishers, 2000), p. 269.
- [85] T. A. Halgren and W. N. Lipscomb, *Chem. Phys. Lett.* **49**, 225 (1977).
- [86] M. J. S. Dewar, E. F. Healy, and J. J. P. Stewart, *J. Chem. Soc., Faraday Trans. 2* **80**, 227 (1984).
- [87] C. Cardenas-Lailhacar and M. C. Zerner, *Int. J. Quant. Chem.* **55**, 429 (1995).
- [88] D. A. Liotard, *Int. J. Quant. Chem.* **44**, 723 (1992).
- [89] C. Choi and R. Elber, *J. Chem. Phys.* **94**, 751 (1991).
- [90] F. Jensen, *Introduction to Computational Chemistry* (John Wiley & Sons, 1999).
- [91] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1989).
- [92] R. Crehuet and M. J. Field, *J. Chem. Phys.* **118**, 9563 (2003).
- [93] D. C. Liu and J. Nocedal, *Math. Prog.* **45**, 503 (1989).
- [94] J. Nocedal, *Math. Comp.* **35**, 773 (1980).



- [95] J. Nocedal, *Software for Large-scale Unconstrained Optimization*, <http://www.ece.northwestern.edu/~nocedal/lbfgs.html> (2000).
- [96] K. Müller and L. D. Brown, *Theor. Chim. Acta* **53**, 75 (1979).
- [97] D. R. Alfonso and K. D. Jordan, *J. Comp. Chem.* **24**, 990 (2003).
- [98] D. R. Alfonso, *Driver for Performing Minimum Energy Path Optimizations Using the Nudged Elastic Band Algorithm*, <http://www.pitt.edu/~alfonso/NEB/neb.html> (2002).
- [99] D. J. Wales, *OPTIM: A Program for Optimizing Geometries and Calculating Reaction Pathways*, <http://www-wales.ch.cam.ac.uk/software.html> (2003).
- [100] Y. M. Rhee, *J. Chem. Phys.* **113**, 6021 (2000).
- [101] R. E. Leone and P. v. R. Schleyer, *Angew. Chem. Int. Ed. Engl.* **9**, 860 (1970).
- [102] J. P. K. Doye, M. A. Miller, and D. J. Wales, *J. Chem. Phys.* **111**, 8417 (1999).
- [103] M. A. Miller, J. P. K. Doye, and D. J. Wales, *Phys. Rev. E* **60**, 3701 (1999).
- [104] J. P. Neirotti, F. Calvo, D. L. Freeman, and J. D. Doll, *J. Chem. Phys.* **112**, 10340 (2000).
- [105] F. Calvo, J. P. Neirotti, D. L. Freeman, and J. D. Doll, *J. Chem. Phys.* **112**, 10350 (2000).

## VI. TABLES

TABLE I: The minimal number of images and total number of gradient calls (in parentheses) are shown for degenerate rearrangements of LJ<sub>7</sub>. The image range was  $2 \leq N \leq 20$  and the iteration range was  $1 < \ell \leq 3,000N$ . Each SQVV calculation was started from the guess produced using linear interpolation, while guesses for L-BFGS runs were preoptimised using DNEB/SQVV until the RMS force dropped below 2.0. Every iteration the images that satisfy  $E_i > E_{i\pm 1}$  were optimised further using eigenvector-following. The transition state candidates that converged to a true transition state within five iterations were used to generate the connected minima using energy minimisation. If this procedure yielded a connected pathway the calculation was terminated and the rest of the parameter range was not explored. Otherwise, the number of images was incremented and the procedure repeated. The number of gradient calls is a product of the number of images and the total number of iterations. For the L-BFGS calculations the number of iterations includes the SQVV preoptimisation steps (100 on average) and the actual number of L-BFGS steps. Dashes signify cases where we were unable to obtain a connected pathway.

Method	1-2	2-3	3-4	4-5
DNEB/L-BFGS	5(1720)	18(30276)	11(2486)	18(8010)
DNEB/SQVV	16(21648)	–	10(14310)	–

TABLE II: The minimal number of iterations needed to produce connected pathways for four degenerate rearrangements of LJ<sub>7</sub> using a 50-image NEB. The strategy of this calculation is identical to the one described in the caption to Table I, except that the number of images was fixed.

Method	1-2	2-3	3-4	4-5
DNEB/L-BFGS	131 <sup>a</sup>	493	171	326
DNEB/SQVV	1130	15178	2777	23405 <sup>b</sup>
NEB/SQVV	11088 <sup>b</sup>	–	30627	–

<sup>a</sup> The number of iterations is the sum of the SQVV preoptimisation steps (100 on average) and the actual number of iterations needed by L-BFGS minimiser. <sup>b</sup> This value is not directly comparable since DNEB converged to a different path that contains more intermediate minima.

Dashes signify cases where we were unable to obtain a connected pathway.

## VII. FIGURE CAPTIONS

1. Graphical representation of the nudged elastic band approach. (a) The optimised nudged elastic band for a two-dimensional model surface. The band contains 21 images and connects two minima  $\mathbf{X}_0$  and  $\mathbf{X}_{23}$ . Image  $\mathbf{X}_9$  has the highest energy and might therefore be used to estimate transition state properties or as a starting guess for further refinement. (b) ‘Nudging’: the NEB depicted in (a) is projected onto the  $xy$  plane and feels only the perpendicular component of the true gradient from the effective potential  $V^\perp$ .
  
2. Details of recent NEB implementations. (a) Conditions under which kinks appear during optimisation of the NEB using the tangent estimated from the line segment  $\hat{\boldsymbol{\tau}}$  connecting images  $i+1$  and  $i-1$ . Displacement of image  $i$  from the path (dash-dotted line) creates forces  $\mathbf{F}_{i-1}^\perp = -\mathbf{g}_{i-1}^\perp$  and  $\mathbf{F}_i^\perp = -\mathbf{g}_i^\perp$ . While  $\mathbf{F}_i^\perp$  is a restoring force that originates from  $V^\perp$ ,  $\mathbf{F}_{i-1}^\perp$  is destabilising and originates from  $V^\parallel$  (and is non-zero due to the fact that the tangent at image  $i-1$  has changed after displacement of image  $i$ ). For the case of small displacements the potential may be resolved into two contributions,  $V^\perp = (k^\perp/2)x^2$  and  $V^\parallel = -k^\parallel y$ , and kinks will not appear if  $k^\parallel/k^\perp < |\hat{\boldsymbol{\tau}}|$ . (b) Tangent estimate using the higher energy neighbour: image  $i+1$  is ‘hanging’ on to image  $i$ . The separation  $d$  is controlled by the lower-lying images ( $> i+1$ ) but not  $V$ . (c) An NEB that follows the curved region of the path: since the spring force  $\mathbf{F}_1$  acting on image  $i$  is compensated by projection  $\mathbf{F}_2$ , the distribution of images becomes uneven. (d) Corner-cutting displayed on a cross-section of the curved part of the path depicted in (c): the image is displaced from the path due to the presence of  $\mathbf{F}_{spr}^\perp$ .
  
3. (a) Number of iterations,  $\ell$ , and (b) average deviation from the average image separation,  $\varsigma$ , as a function of the spring force constant,  $k_{spr}$ , obtained using a 17-image NEB on the Müller-Brown surface [96]. Minimisation was performed using QVV with time step 0.01 and RMS force termination criterion 0.01. The number of iterations is shown for velocity quenching after the coordinate update (diamonds), after the gradient evaluation (squares) and at the half-step through the velocities update (stars).
  
4. (a) Number of iterations,  $\ell$ , and (b) average deviation from average image separation,  $\varsigma$ , as a function of the spring force constant,  $k_{spr}$ , obtained using a 17-image NEB

for the Müller-Brown surface [96]. Minimisation was performed using L-BFGS with number of corrections  $m = 4$ , maximum step size 0.1 and RMS force termination criterion 0.01.

5. RMS gradient  $g_{\text{RMS}}^\perp$  as a function of iteration number  $\ell$ . A 7-image NEB was used to model an isomerisation path in the LJ<sub>7</sub> cluster (global minimum  $\rightarrow$  second-lowest minimum). Minimisation was performed using the SQVV (a) and L-BFGS (b) methods. Results are shown for minimisations with and without removing overall rotation and translation (diamonds and stars, respectively). The inset in (a) depicts the average deviation from the average image separation,  $\varsigma$ , as a function of iteration number for minimisations using SQVV, while the inset in (b) shows  $g_{\text{RMS}}^\perp$  recorded for 1000 iterations of L-BFGS minimisations. These calculations were all continued for a fixed number of iterations, regardless of convergence.
6. Structures of the most stable isomers for (a) LJ<sub>7</sub>, (b) LJ<sub>38</sub> and (c) LJ<sub>75</sub> clusters, which were used as endpoints in the NEB calculations. The first endpoint was the global minimum in each case. For LJ<sub>38</sub> and LJ<sub>75</sub> the second endpoint was chosen to be second-lowest minimum shown on the right in parts (b) and (c), respectively, while a permutational isomer of the global minimum was used as the second endpoint in all the LJ<sub>7</sub> calculations. The notation 1–2 denotes an LJ<sub>7</sub> rearrangement where the second endpoint is structure (a) with atoms 1 and 2 swapped. The structures and numbering employed for LJ<sub>38</sub> and LJ<sub>75</sub> are defined at <http://www-wales.ch.cam.ac.uk/~sat39/DNEBtests/>.
7. The energy,  $E$ , as a function of the integrated path length,  $s$ , for four degenerate rearrangements of LJ<sub>7</sub>. These profiles were constructed using energy minimisation to characterise the paths connected to transition states obtained by EF refinement of candidate structures obtained from DNEB calculations [34].
8. The energy,  $E$ , as a function of integrated path length,  $s$ , for pathways linking the two lowest minima of LJ<sub>38</sub> and LJ<sub>75</sub>. Calculations were initiated between two different sets of permutational isomers of these minima. For each profile the number of transition states,  $N_t$ , number of DNEB runs,  $N_c$ , and the total number of gradient calls,  $N_g$ , are shown. Maximum values of  $\tilde{N}$ ,  $\beta$  and  $\pi$  are marked next to the corresponding transition states. The endpoints were illustrated in Figure 6.

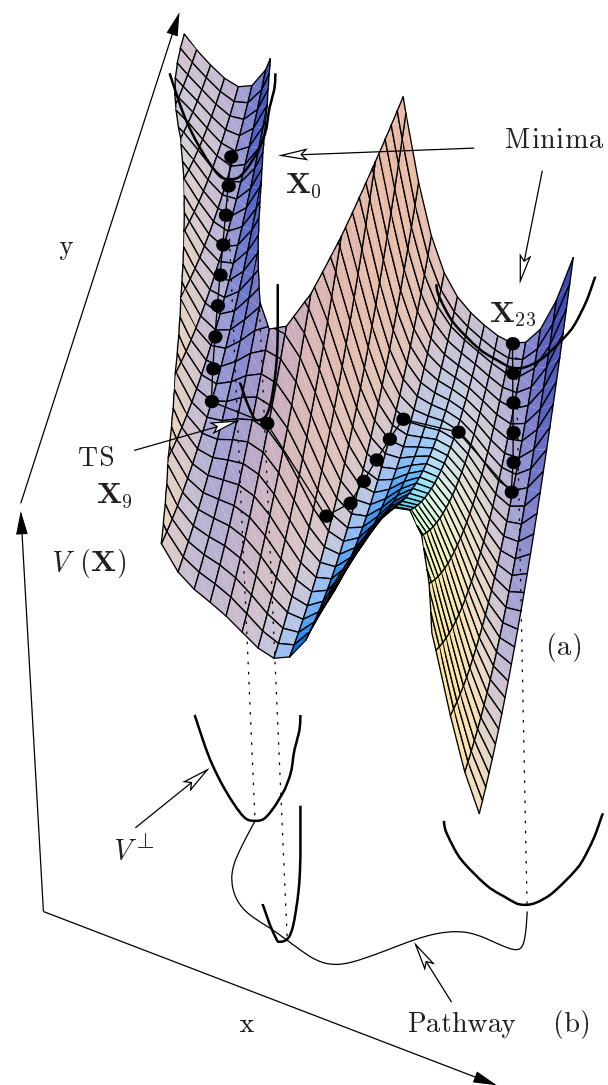


FIG. 1:

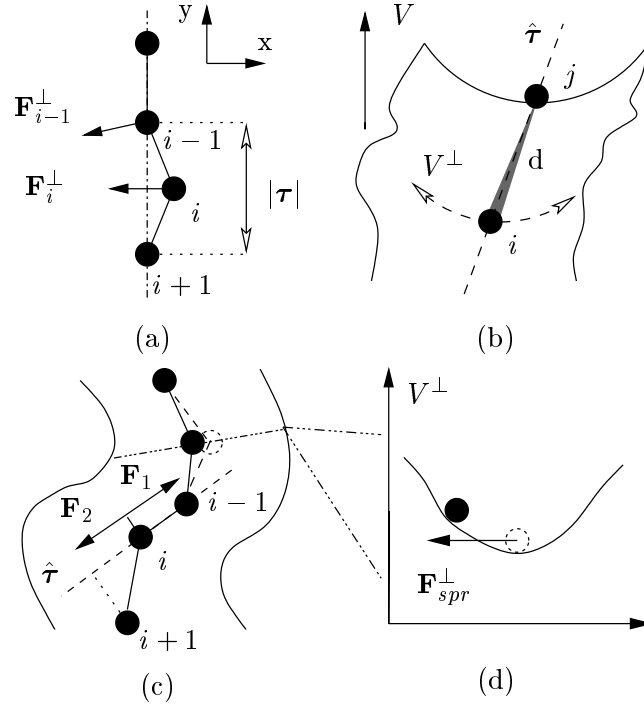


FIG. 2:

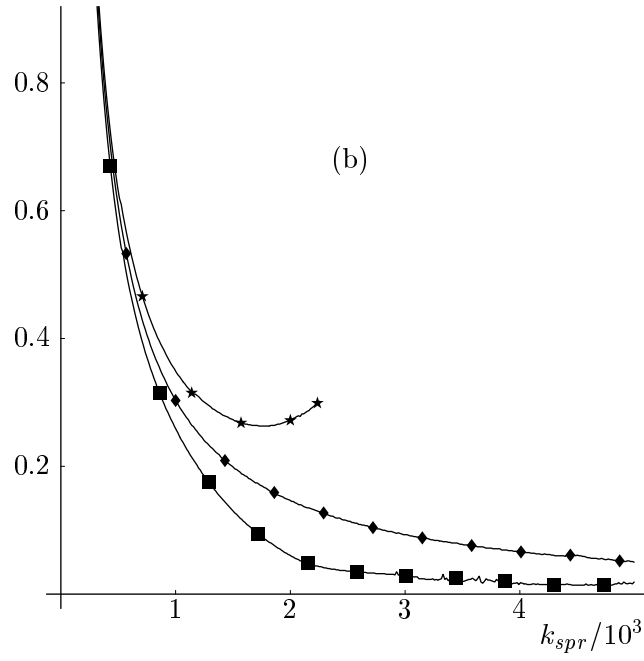
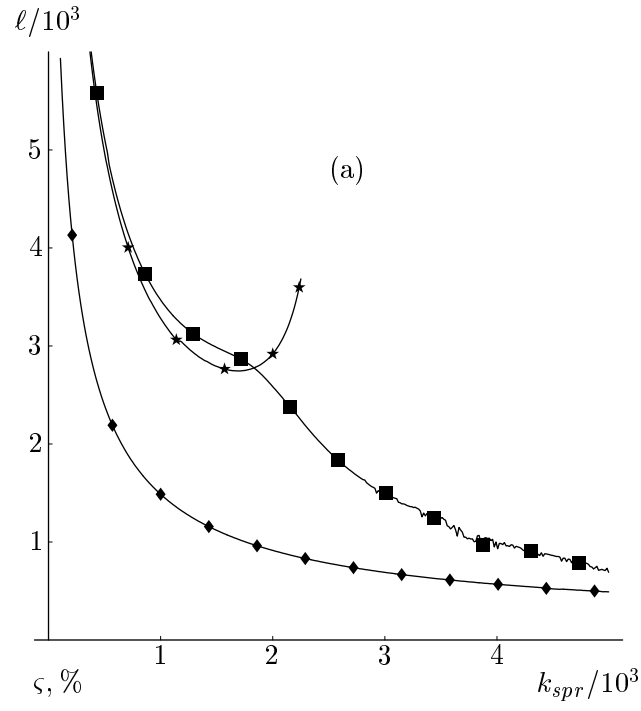


FIG. 3:



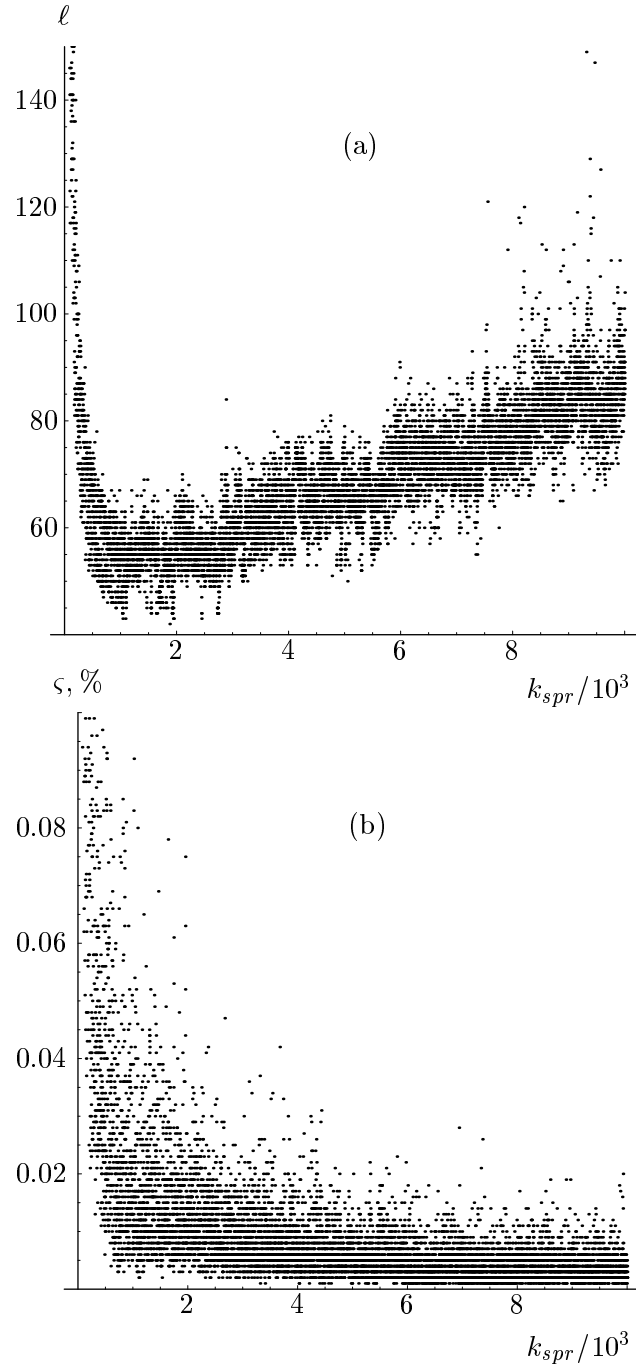


FIG. 4:

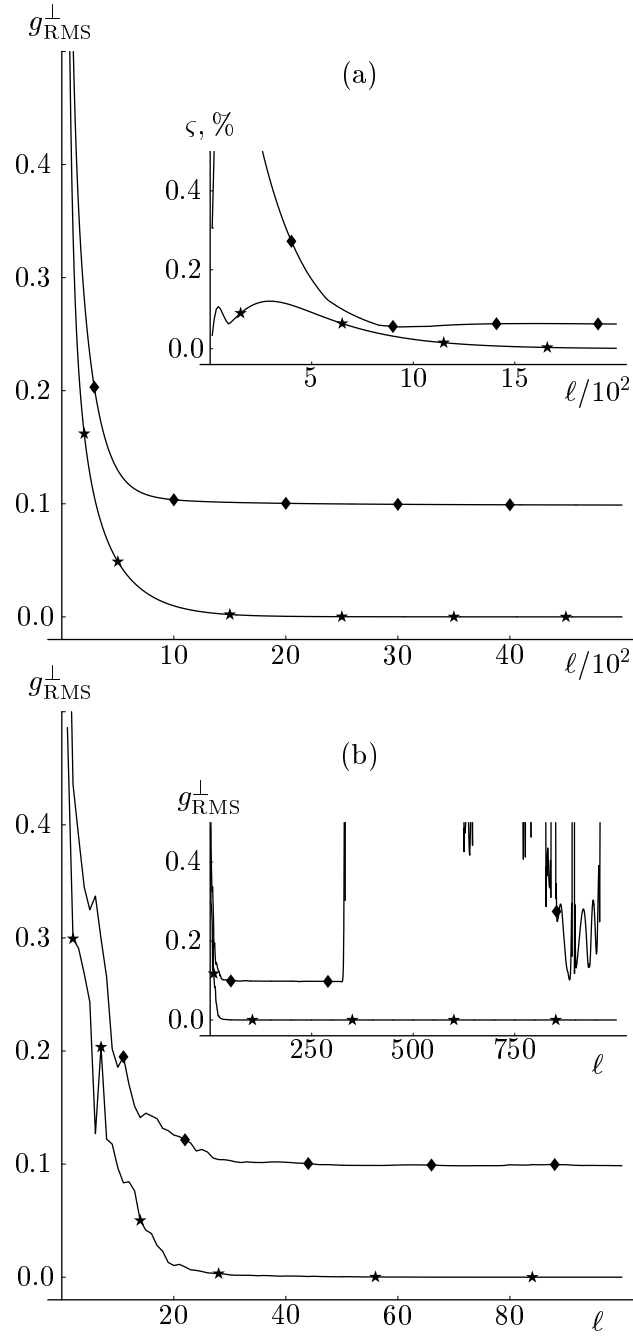
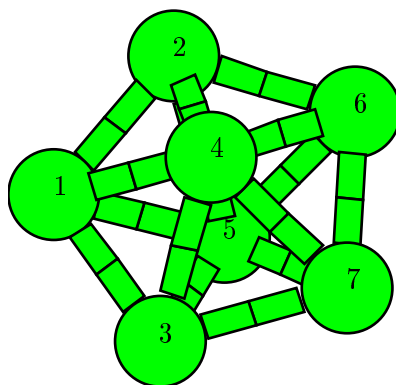
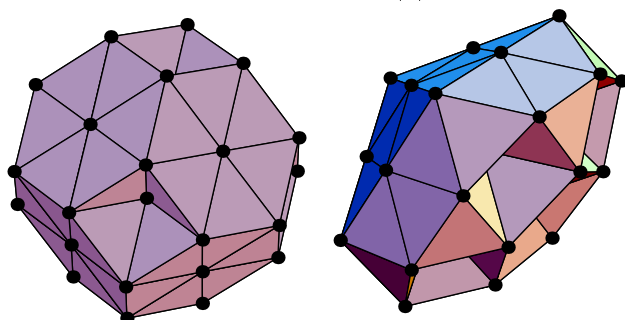


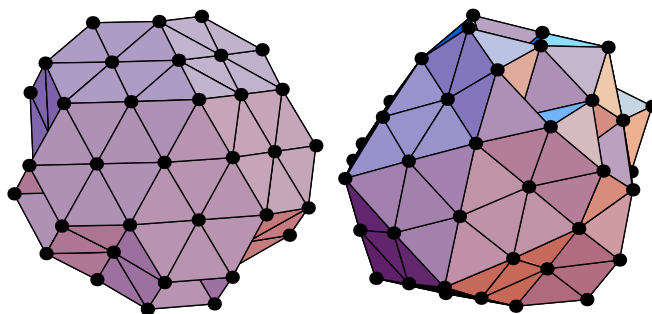
FIG. 5:



(a)



(b)



(c)

FIG. 6:

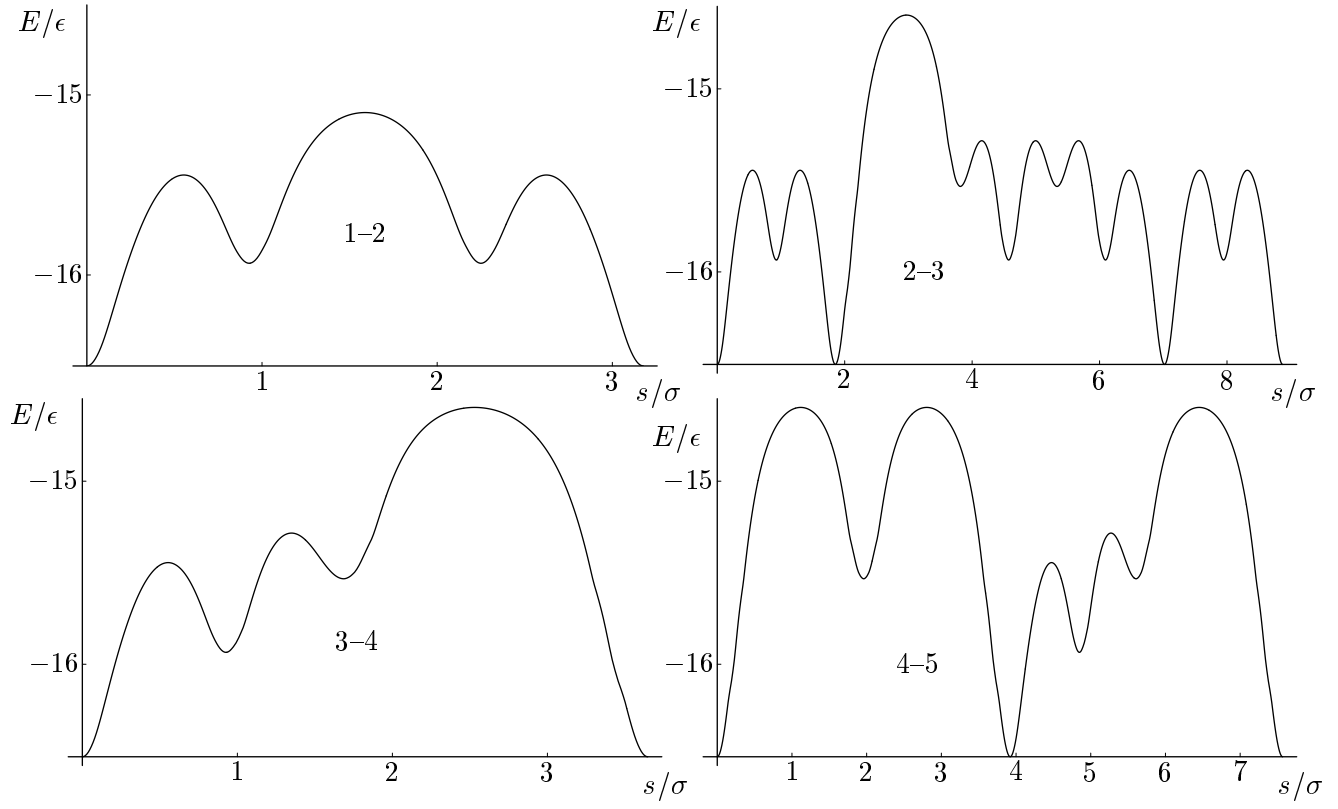


FIG. 7:

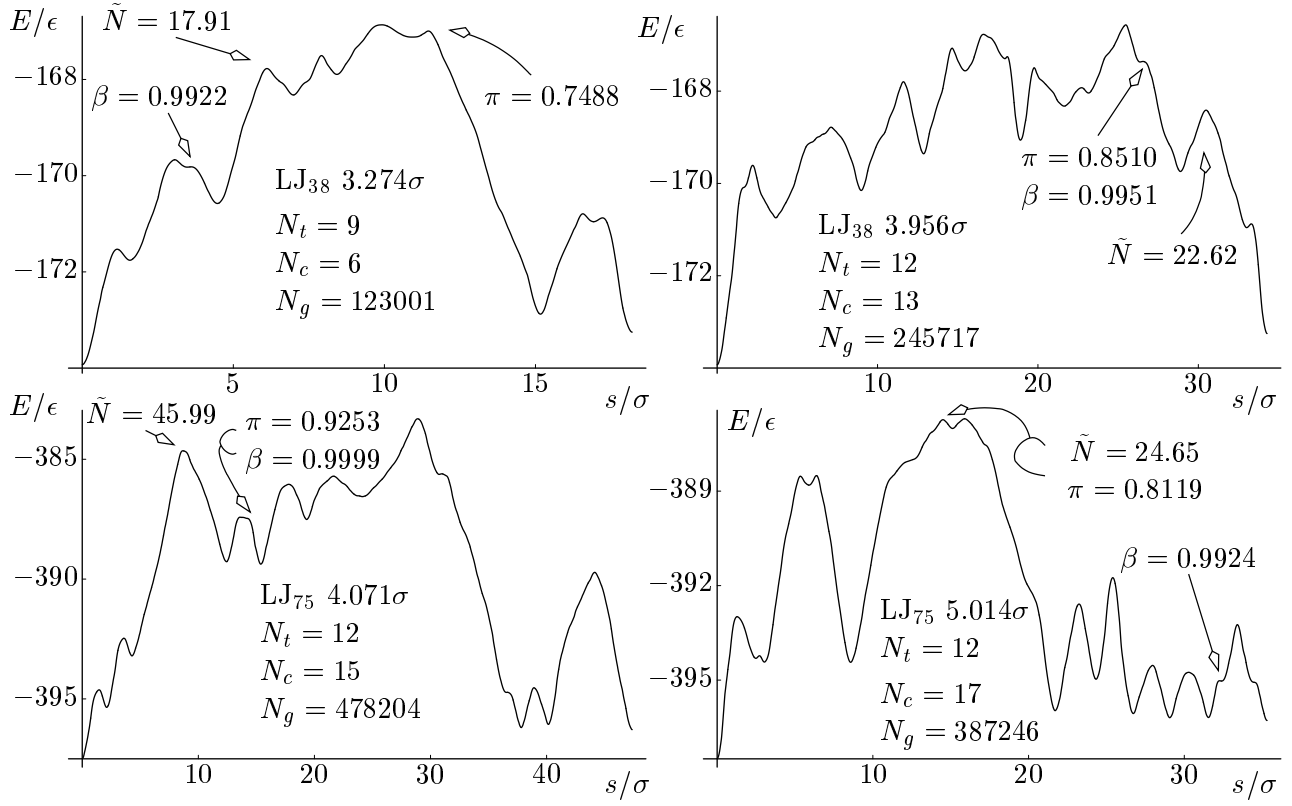


FIG. 8: