CHAPTER 16

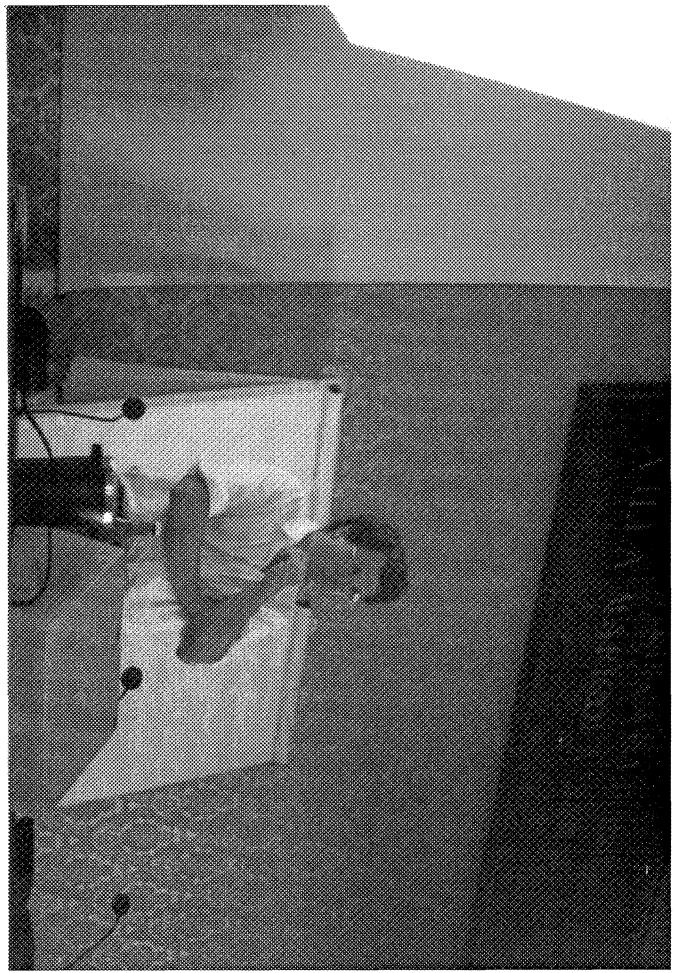
Nudged elastic band method for finding minimum energy paths of transitions

HANNES JÓNSSON[a,b,*], GREG MILLS[a], AND KARSTEN W. JACOBSEN[b]

[a] *Department of Chemistry, Box 351700, University of Washington, Seattle, WA 98195-1700*

[b] *CAMP, Department of Physics, Technical University of Denmark, DK-2800 Lyngby, Denmark*

[*] *To whom correspondance should be addressed, e-mail hannes@u.washington.edu*



Contents

This page is intentionally left blank

1. – Introduction

A common and important problem in theoretical chemistry and in condensed matter physics is the identification of a lowest energy path for a rearrangement of a group of atoms from one stable configuration to another. Such a path is often referred to as the 'minimum energy path' (MEP). It is frequently used to define a 'reaction coordinate' [1] for transitions, such as chemical reactions, changes in conformation of molecules, or diffusion processes in solids. The potential energy maximum along the MEP is the saddle point energy which gives the activation energy barrier, a quantity of central importance for estimating the transition rate within harmonic transition state theory [2].

Many different methods have been presented for finding reaction paths and saddle points (see reference [3] for a recent review). Some of the schemes start at the local minimum on the potential energy surface representing the initial state and then trace stepwise, in a sequential manner, a path of slowest ascent [4, 5, 6]. Such slowest ascent paths, however, do not necessarily lead to saddle points. Figure 1 shows an example of such a system (the details of the model, Model II, are given in the Appendix). A commonly used method involves calculating normal modes of a local harmonic approximations of the potential energy surface and then following each of the modes until a saddle point is found[7, 8]. Each step in this procedure requires evaluation and diagonalization of the second derivative matrix and is therefore limited to rather small systems and descriptions of atomic interactions where second derivatives are readily available (excluding, for example, plane wave based density functional theory calculations).

Other methods make use of a two point boundary condition, i.e. both the initial and final configurations for the transition are given. These would normally be two local minima on the multidimensional potential energy surface which may have been obtained by finite temperature molecular dynamics or by Monte Carlo simulated annealing methods. We will focus here on such problems. We will also focus the discussion on methods that only require first derivatives of the potential energy. The simplest of these is the 'drag' method or the 'reaction coordinate' method, where some subset of the coordinates in the system is used to define a progress variable, often through linear interpolation between the initial and final configurations. This one degree of freedom is then varied stepwise from the initial to the final value and at each step a minimization is carried out over the remaining degrees of freedom (a total of $(3N-1)$ degrees of freedom if the system consists of N atoms in 3 dimensions). While this method can work well in simple cases, there are many instances were it fails. The path generated may be discontinuous and the procedure may depend on the direction of the drag (hysteresis effects). In particular, some atomic coordinates may 'slip' near the saddle point region and the saddle point configuration will then be missed [9, 10]. An example of this is given in figure 1. Even though the initial straight line interpolation goes very close to the saddle point, the relaxation of the remaining degree of freedom gives a result that is quite far from the MEP. For a range of values of the drag coordinate, there are, in fact, two minima. If the drag coordinate is incremented gradually using the previous minimum as an initial

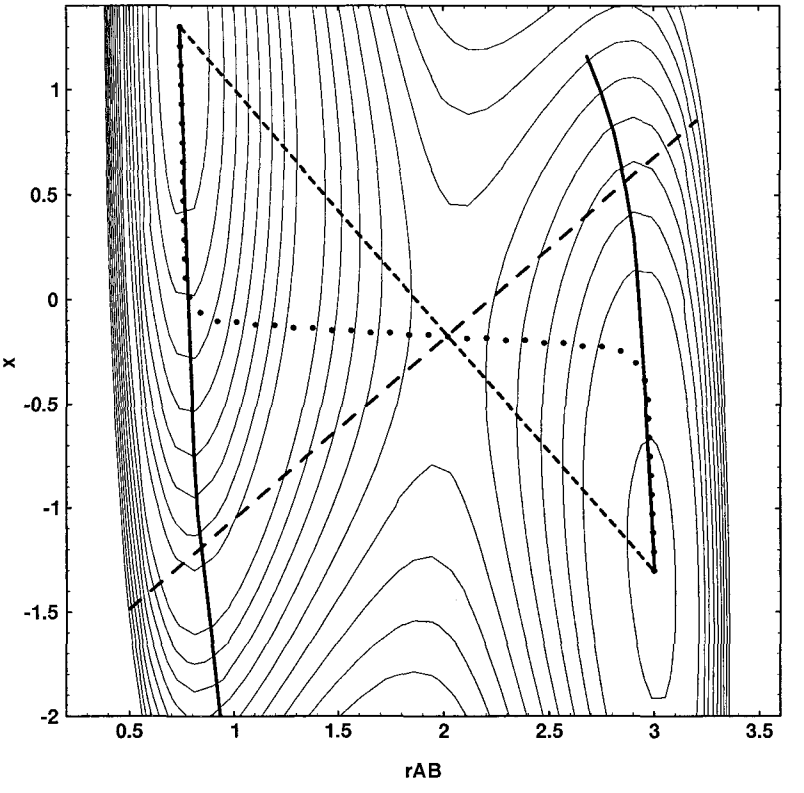


Fig. 1. – A contour plot of the potential energy surface for a simple test problem, model II (see Appendix), where an atom, B, can form a chemical bond with either one of two fixed atoms, A or C, (described by a LEPS potential) and is simultaneously coupled in a harmonic way to a fourth atom, D. The horizontal axis gives the A-B distance and the vertical axis the B-D distance. All atoms are confined to a line. The MEP is shown with a dotted line going through the saddle point (obtained with the NEB method using 50 images). A straight line interpolation between the initial and final state minima, defining a 'drag coordinate', is shown with small dashes. In the drag method, the system energy is minimized with respect to all remaining degrees of freedom, corresponding here to relaxation along a line perpendicular to the drag line. Such a line going through the saddle point is shown with large dashes. The drag method gives paths indicated by two thick, solid lines. The drag paths lie close to the slowest ascent paths, which do not lead to the saddle point. The result of the drag method is a discontinuous path which depends on the direction of the drag.

guess in the minimization procedure for the next value of the drag coordinate, then two different paths will be traced out depending on whether the calculation starts from the initial state or the final state. These paths 'overshoot' the saddle point until the potential energy contours become parallel to the line corresponding to the degree of freedom that is being optimized.

More recently another class of two point methods has emerged. There, a chain of images (or replicas, or 'states') of the system is generated between the end point configurations and all the intermediate images are optimized simultaneously in some

concerted way. An example calculation is shown in figure 2. Here, the initial images are located along a line joining the initial and final point, but after an optimization calculation acting on all the images simultaneously, the images lie near the MEP. The distribution of the images, which give a discrete representation of the path, can be controlled and has been set here to be twice as high in the barrier region as compared with the end regions. An important aspect of these types of methods is that parallel computers or simply a cluster of networked computers can be used very efficiently for finding the MEP, in contrast to methods that are based on a sequential 'walk' along the potential energy surface [4, 5, 6, 7, 8, 11, 12]. We will give an overview of these types of methods in section 2. We will then, in section 3, discuss a method we refer to as the 'Nudged Elastic Band' (NEB) method. This method has already been presented briefly elsewhere [13] and has been applied successfully to a wide range of problems, including studies of diffusion processes at metal surfaces[14], multiple atom exchange processes during sputter deposition[15], dissociative adsorption of a molecule on a surface[13], diffusion of rigid water molecules in ice[16], premelting of metal clusters[17], contact formation between metal tip and a surface[18], cross-slip of screw dislocations in a metal (a simulation requiring over 100,000 atoms in the system, and a total of over 2,000,000 atoms in the MEP calculation)[19], and atomic exchange processes at semiconductor surfaces (using a plane wave based DFT method to calculate the atomic forces) [20]. The NEB method is particularly simple and easy to implement and we give a detailed description for implementing the method in section 4. Section 5 describes a simple test calculation on an adatom hop on a surface, illustrating the importance of nudging the elastic band. Section 6 points out the importance of including spring interactions between the images. A discussion of the theoretical basis of the NEB method and possible extensions of it are discussed in section 7. A summary is give in section 8.

2. – Chain-of-states methods

In chain-of-states methods, several images (or 'states') of the system are connected together to trace out a path. If the images are connected with springs of zero natural length and the object function is defined as

$$S^{PEB}(\vec{R}_1,\ldots,\vec{R}_{P-1}) = \sum_{i=0}^{P} V(\vec{R}_i) + \sum_{i=1}^{P} \frac{Pk}{2}(\vec{R}_i - \vec{R}_{i-1})^2 \quad (1)$$

then the chain is mathematically analogous to a Feynman path integral [21] for an off-diagonal element of a density matrix describing a quantum particle. Kuki and Wolynes carried out thermal sampling of such paths to identify important tunneling paths of electrons in proteins [22]. In the context of finding MEPs for classical systems, one could envision minimizing the object function in eqn. (1) with respect to the intermediate images while keeping the end point images, R_0 and R_P, fixed. We will refer to this as the *plain* elastic band (PEB) method. Below, we will illustrate how it, in fact, fails to provide the MEP in most situations. We will also present modifications to eqn. (1) which solve those problems.

An early attempt to use a chain-of-states method to analyse transitions in classical systems was made by Pratt [23]. He proposed a statistical procedure where a finite temperature Monte Carlo algorithm is used to sample a Markovian chain of images of the system in order to find transition state regions. This approach has recently been further developed by Chandler and coworkers [24] and opens the possibility of analysing

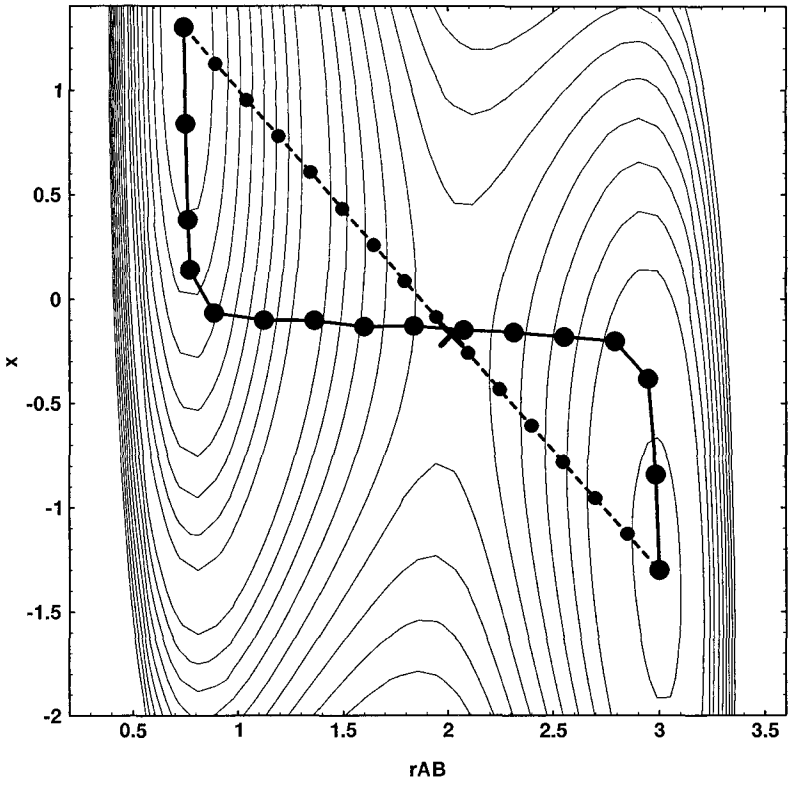


Fig. 2. – Same test problem as in Fig. 1. The initial and final configuration of a NEB with 16 images is shown. The straight line interpolation between the initial and final point is shown with dashes and the small, filled circles along the line indicate the initial configuration chosen for the images in the elastic band. The larger, filled circles lying close to the MEP show the images after convergen ce. The spring constant is $k = 0.5$ near the ends of the band and $k = 1.0$ in the middle so as to increase the density of images in the most relevant region.

transitions in rugged potential energy landscapes where a very large number of relevant MEPs need to be included in rate calculations. We will limit the discussion here to systems where one or at most a few MEPs need to be found, even though the systems are large and the transitions complex.

An algorithm for finding a reaction path based on optimization of a line integral over a discretized path was presented by Elber and Karplus [25]. They defined the object function as

$$S^{EK}(\vec{R}_1,\ldots,\vec{R}_{P-1}) = \frac{1}{\sqrt{\sum_j |\Delta\vec{l}_j|^2}} \sum_{j=1}^{P-1} V(\vec{R}_j)|\Delta\vec{l}_j| + \sum_{j=1}^{P} \lambda \left(|\Delta\vec{l}_j| - \sqrt{\sum_j^P |\Delta\vec{l}_j|^2/P} \right)^2 \quad (2)$$

where $|\Delta\vec{l}_j| = \sqrt{|\vec{R}_j - \vec{R}_{j-1}|^2}$. Here, the springs connecting adjacent images have a natural length, equal to the average separation between images along the current estimate of the optimal path. The object function is minimized with respect to the intermediate images, $\vec{R}_1, \vec{R}_2, \ldots, \vec{R}_{P-1}$ using a non-linear optimization algorithm. The algorithm has been applied to transitions in several biological systems. It can give a good indication of where the saddle point is, but, as noted by the authors, does not converge to the MEP and does not give a good estimate of the saddle point energy. Typically the results are refined using a Newton-Raphson method (which requires the second derivatives) or an iterative method based on the first derivatives only [11].

Czerminski and Elber presented an improved method for finding approximate MEPs, the Self-Penalty Walk algorithm (SPW) [26]. Noting that the minimization of S^{EK} can lead to aggregation of the images and crossings of the path with itself in the regions near minima, they added a repulsive term to the object function in eqn. (2) so as to keep the images apart.

Ulitsky and Elber proposed a different algorithm which can converge to the MEP but has a smaller radius of convergence than the SPW [27]. Choi and Elber later presented an improvement of this algorithm with faster convergence, the Locally Updated Planes (LUP) algorithm [28]. There, an initial guess for a MEP in terms of a sequence of states is improved on by first estimating a local tangent to the path as the line segment connecting the previous and later image in the chain

$$\hat{q}_i = \frac{\vec{R}_{i+1} - \vec{R}_{i-1}}{|\vec{R}_{i+1} - \vec{R}_{i-1}|} \tag{3}$$

and then minimizing the energy of each image, i, within the hyperplane with normal q_i, i.e. relaxing the system according to

$$\frac{\partial \vec{R}_i}{\partial t} = -\vec{\nabla} V(\vec{R}_i)[1 - \hat{q}_i \hat{q}_i]. \tag{4}$$

After every M steps (where M is on the order of 10) in the relaxation, the local tangents $\hat{q}_i$ are updated. This algorithm has been used in many different contexts, in particular in large DFT calculations of chemical processes [29]. Since the images are not connected, the LUP algorithm gives an uneven distribution of images along the path, and can even give a discontinuous path when two or more MEPs lie between the given initial and final states [28]. Choi and Elber point out that it is important to start with a good initial guess to avoid these problems, for example an approximate MEP obtained by the SPW method. The NEB method, discussed below, is closely related to both the LUP method and the elastic band methods where continuity of the path is guaranteed by including spring interactions. The NEB method incorporates the strong points of both of these approaches.

Sevick, Bell and Theodorou [30] proposed a chain of states method for finding the MEP, but their optimization method, which includes explicit constraints for rigidly fixing the distance between images, requires evaluation of the matrix of second derivatives of the potential and is, therefore, not as applicable to large systems and complex interactions.

Chain-of-states methods have also been used for finding classical dynamical paths [31, 32]. Gillilan and Wilson [32] suggested using an object function similar to S^{PEB} eqn. (1) for finding saddle points, but the method suffers from problems which will be illustrated in the next section.

3. – The NEB method

To motivate the NEB method, we first discuss the shortcomings of the plain elastic band method where the object function is defined as in eqn. (1). Figure 3 shows the result for a two-dimensional LEPS potential (see appendix for details on the model). In this method, the force acting on image i is

$$\vec{F}_i = -\vec{\nabla}V(\vec{R}_i) + \vec{F}_i^s \tag{5}$$

where

$$\vec{F}_i^s \equiv k_{i+1}\left(\vec{R}_{i+1} - \vec{R}_i\right) - k_i\left(\vec{R}_i - \vec{R}_{i-1}\right). \tag{6}$$

Figure 3a shows the results of a calculation where all spring constants are $k = 1.0$. Here,

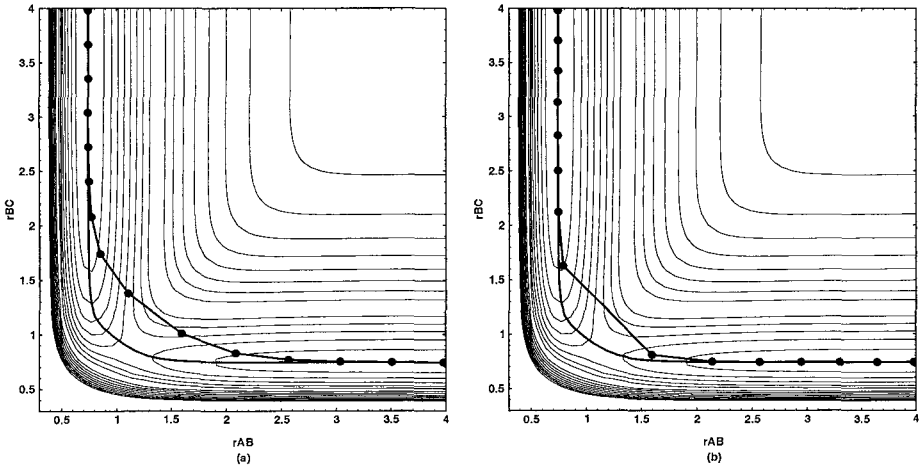


Fig. 3. – A contour plot of the potential energy surface for a simple test problem, model I (see Appendix), where an atom, B, can form a chemical bond with either one of two movable atoms, A or C, as described by a LEPS potential. The horizontal axis gives the A-B distance and the vertical axis the B-C distance. (a) A *plain* elastic band with spring constant $k = 1.0$ is shown with filled circles connected by a solid line. It cuts the corner and leads to an over estimate of the saddle point energy. The MEP as obtained by the NEB method is shown with a solid line going through the saddle point. The inset shows the variation of the potential energy along the plain elastic band (dashed line) and the MEP (solid line). (b) Same as (a) but with a spring constant of $k = 0.1$. The corner cutting is diminished, but now the images slide down from the barrier region towards the minima at the endpoints thus reducing the resolution of the path in the region of greatest impo rtance.

the elastic band is too stiff and the path cuts the corner and therefore misses the saddle point region. Figure 3b shows the result when a smaller spring constant is used $k = 0.1$. Now the elastic band comes closer to the saddle point, but the images manage to slide down and avoid the barrier region, thus reducing the resolution of the path in the most critical region. These problems have been noted previously [32]. In the continuum limit,

the object function becomes

$$S^{PEB} = P\int d\lambda \left(V(\vec{R}(\lambda)) + \frac{k(P)}{2P} \left| \frac{d\vec{R}}{d\lambda} \right|^2 \right) \tag{7}$$

where the number of images, P, is going to infinity, and where we can specify some dependence $k(P)$ of the spring constant on the number of images. If $\frac{k(P)}{2P}$ goes to a finite (but non-zero) value as the number of images increases, the stiffness of the band leads to 'corner-cutting' as described above. However, even if the coefficient $\frac{k(P)}{2P}$ vanishes as P gets larger the resulting path may not go through the saddle point. This can be seen by noting, that the object function is analogous to the action of a classical particle of unit mass moving on the inverted potential energy surface with 'time' defined as $\tau = \sqrt{P/k(P)}\lambda$. Then, minimizing S^{PEB} corresponds to the dynamics $d^2\vec{R}/d\tau^2 = \vec{\nabla}V$. As the classical particle moves through the saddle point region it will have a finite velocity. If the path is curved there, the particle will be affected by a force with components perpendicular to the path, *i.e.* the particle cannot follow the MEP.

We now present a simple way to solve the problems with the plain elastic band method, enabling convergence to the MEP when sufficiently many images are used in the elastic band. The problem with corner cutting results from the component of the spring force which is perpendicular to the path and tends to pull images off the MEP. The problem with the sliding down results from the component of the true force $\vec{\nabla}V(R_i)$ in the direction of the path. The distance between images becomes uneven so the net spring force can balance out the parallel component of the true potential force.

The cure for this is very simple. In the NEB method, a minimization of an elastic band is carried out where the perpendicular component of the spring force and the parallel component of the true force are projected out. The force on image i then becomes

$$\vec{F}_i^0 = -\vec{\nabla}V(\vec{R}_i)|_\perp + \vec{F}_i^s \cdot \hat{\tau}_\parallel \, \hat{\tau}_\parallel \tag{8}$$

where $\hat{\tau}_\parallel$ is the unit tangent to the path and $\vec{\nabla}V(\vec{R}_i)|_\perp = \vec{\nabla}V(\vec{R}_i) - \vec{\nabla}V(\vec{R}_i) \cdot \hat{\tau}_\parallel \, \hat{\tau}_\parallel$. We refer to this projection of the perpendicular component of $\vec{\nabla}V$ and the parallel component of the spring force as 'nudging'. These force projections decouple the dynamics of the path itself from the particular distribution of images chosen in the discrete representation of the path. The spring force then does not interfere with the relaxation of the images perpendicular to the path and the relaxed configuration of the images satisfies $\vec{\nabla}V(\vec{R}_i)|_\perp = 0$, i.e. they lie on the MEP. Furthermore, since the spring force only affects the distribution of the images within the path, the choice of the spring constant is quite arbitrary. This decoupling of the relaxation of the path and the discrete representation of the path is essential to ensure convergence to the MEP.

When the energy of the system changes rapidly along the path, but the restoring force on the images perpendicular to the path is weak, as can be the case, for example, when covalent bonds are broken and formed (as in Si adatom diffusion on a Si surface), the paths can get 'kinky' and the convergence of the minimization slowed down. Images caught in the region of large parallel force try to slide down but since the nudging ensures equal spacing of the images, this can only occur by lengthening and, thereby, buckling of the chain in another, nearly force free region. The estimate of the local tangent can then become problematic. An effective remedy for this has been the introduction of a smooth switching function that gradually turns on the perpendicular component of the

spring force where the path becomes kinky. The force on image i then becomes

$$\vec{F}_i^{NEB} = \vec{F}_i^0 + f(\phi_i)\left(\vec{F}_i^s - \vec{F}_i^s \cdot \hat{\tau}_{\|}\, \hat{\tau}_{\|}\right) \tag{9}$$

where $f(\phi)$ is a switching function which goes from 0 for a straight path to 1 if adjacent segments of the path form a right angle, for example

$$f(\phi) = \frac{1}{2}\left(1 + \cos\left(\pi(\cos\phi)\right)\right) \tag{10}$$

when $\pi/2 < \phi < \pi/2$ and 1 else. Here $\cos\phi_i = (\vec{R}_{i+1} - \vec{R}_i)\cdot(\vec{R}_i - \vec{R}_{i-1})/(|\vec{R}_{i+1} - \vec{R}_i||\vec{R}_i - \vec{R}_{i-1}|)$. The full perpendicular spring force is kept if the angle between adjacent path segments becomes 90° but none of the perpendicular spring force is kept if the angle is zero (i.e. the three images are in a straight line). A small amount of the perpendicular spring force is often enough to straighten out the path and significantly improve convergence. One should be careful not to include too much of the perpendicular spring force, since it can lead to corner cutting and an overestimate of the saddle point energy when few images are used to represent the path.

Figure 3 shows the result of NEB calculations for the LEPS problem, model I. Clearly, both the corner-cutting and the sliding-down problems exhibited by the plain elastic band are solved. In these calculations, the spring constant was chosen to be the same in all cases, so the images become equally spaced along the path. Fig. 2 shows the results for model II. Here the spring constant was chosen to be twice as strong in the region of the saddle point, to illustrate the flexibility the method offers in distributing images along the path. As a result, the resolution of the path is twice as high in the most critical region.

4. – Implementation of the NEB method

The implementation of the NEB method in a molecular dynamics program is quite simple. We have, in particular, implemented NEB with many-body potentials for silicon [20], point-charge models for rigid water molecules[16], Effective Medium[18] and Embedded Atom Method potentials for metals [13, 14] and plane wave based density functional theory calculations[20]. First, the energy and gradient need to be evaluated for each image in the elastic band using some description of the energetics of the system; a first principles calculation or an empirical force field. Then, for each image, the coordinates of the two adjacent images are required in order to estimate the local tangent to the path, project out the perpendicular component of the gradient and add the parallel component of the spring force (according to eqn. (8)) as well as some of the perpendicular component of the spring force if the path is kinky (eqns. (8) and (9)). We have approximated the tangent as the vector that bisects the angle formed between the vectors $\vec{R}_{i+1} - \vec{R}_i$ and $\vec{R}_i - \vec{R}_{i-1}$. The computation of $\vec{\nabla}V$ for the various images of the system can be done in parallel, for example with a separate node handling each one of the images. Each node then needs to receive coordinates of adjacent images only to evaluate the spring force and to carry out the force projections.

Using $P+1$ images of the system to represent the path in between the end points in a system described by N coordinates, the magnitude of the forces needs to be minimized with respect to $N(P-1)$ degrees of freedom. Various techniques can be used for the minimization. We have used a simple but quite efficient method based on the velocity

Verlet algorithm for simulating classical dynamics [33]. At each timestep the coordinates and velocities are updated from the coupled first order equations of motion based on the force evaluated at the current coordinates. If the velocity is zeroed at each step, the algorithm gives a steepest descent minimization. A more efficient procedure is to keep the component of the velocity parallel to the force at the current step, unless it is pointing in a direction opposite to the force. More specifically, if $\hat{F}$ is the $N(P-1)$ dimensional unit vector in the direction of the force and $\vec{u}$ is the $N(P-1)$ dimensional velocity vector, then

$$\vec{u}^{new}(t) = \begin{cases} (\vec{u}(t) \cdot \hat{F})\ \hat{F} & if \quad \vec{u}(t) \cdot \hat{F} > 0 \\ 0 & else. \end{cases} \tag{11}$$

If the force is consistently pointing in a similar direction, the system picks up velocity. This is equivalent to increasing the timestep. When the system overshoots the minimum, the velocity is zeroed.

To start the path finding algorithm, an initial guess is required. We have found a simple linear interpolation between the initial and final point adequate in many cases. When multiple MEPs are present, the optimization leads to convergence to the MEP closest to the initial guess. In order to find the optimal MEP in such a situation, it may be necessary to use a simulated annealing procedure, as discussed in section 7. A typical path finding calculation involves between 10 and 30 images and requires a few hundred iterations to converge.

Finally, it is important to eliminate overall translation and rotation of the system during the optimization of the path. This can be accomplished by fixing six degrees of freedom in each image of the system. We have implemented this in a simple way by fixing one atom (i.e. zeroing all forces acting on one of the atoms in the system), constraining another atom to only move along a line (zeroing, for example, the x and y components of the force), and constraining a third atom to move only in a plane (zeroing, for example, the x component of the force).

5. – Application to an adatom hop on a surface

We now discuss another application of the method to a simple system to illustrate the convergence of the NEB method and to contrast it with the plain elastic band method. We calculated the activation barrier for a diffusion hop of a Cu atom over the bridge site on the $Cu(100)$ surface using a potential of the EAM type [34] fitted to various Cu crystal properties and the Cu dimer [35]. This is a simple test case where the energy of the saddle point can be evaluated simply because of the symmetry of the saddle point configuration.

When only a small subset of the atoms in the system is displaced substantially during the transition (as is the case here), and when the interaction can be broken down into contributions arising from pairwise, three-body, etc. contributions, then the calculation of the energy and force can be implemented very efficiently. Instead of representing every atom in the system with P+1 images in the elastic band, the atoms outside a certain radius away from the 'active' region (roughly equal to the range of the pairwise or three-body interaction potential) can effectively be taken to be in the same position for the whole transition (i.e. same coordinates in all images in the elastic band). We have implemented EAM [34] and Tersoff [36] type interactions potentials in this way. The computations required to determine the full transition path can be of the same order as a relaxation of a single configuration along the path. For example, in a calculation

of the diffusion barrier for a metal atom on a metal surface (such as the one presented below) a total of 500 to 1000 atoms are required for each configuration of the system. Typically only 30 to 50 atoms need to be represented with multiple images in the elastic band, even when complicated exchange processes are simulated. With 20 images in the elastic band (which is typically sufficient) the whole transition path can be treated as a single configuration with about twice the number of atoms. The determination of the MEP then only takes a little over twice the time required to relax a single point along the path! This type of implementation, however, requires additional bookkeeping in the force routine which can be difficult if the interactions are complicated.

Figure 4 shows the results of several calculations using the *plain* elastic band method. When the chain consists of 20 images the estimated barrier clearly increases as the springs are made stiffer, giving a substantial overestimate when $k > 1\ eV/\text{Å}^2$. The atoms do not have enough freedom to relax their positions, because the springs hold the atoms too tightly in place. On the other hand, when $k < 0.1\ eV/\text{Å}^2$ the images tend to slide down the potential barrier, leading to an underestimate of the barrier height and poor resolution of the MEP near the saddle point. This sliding down tendency is evident in comparing the highest energy image to the next-highest energy image. The optimal spring constant is one which gives a balance between the restoring force in the springs and the downward pull from the curvature of the potential barrier. Representing the potential barrier in the vicinity of the saddle point as $V = V_0 - \frac{1}{2}Kx^2$, the value of K is close to being the optimal choice for the constant k. In this particular case, $K = 0.48\ eV/\text{Å}^2$. When the number of images in the chain is increased to 50 a better estimate of the barrier is obtained. Note that the optimal value for k is still the same. The longer chain gives an estimate of the barrier height which varies by about $0.02\ eV$ as k is varied in the range $0.2 - 2.0\ eV/\text{Å}^2$.

The results for this test problem show that the plain elastic band method can, in simple cases, give a reasonable approximation to the saddle point energy if the constant k is chosen to have a value in the right range, despite the fact that there is no guarantee of convergence to the MEP. However, for some systems the plain elastic band method has been found to give very poor results. One example is illustrated in figure 3. This also turned out to be the case in calculations of the reaction path for more complex diffusion processes on metal surfaces [14] and dissociative sticking of H_2 on $Cu(110)$ [13].

Figure 5 shows the calculated results using the NEB method. The images are now equally spaced in the N dimensional configuration space. There is no sliding down the barrier even at low k values. Furthermore, since the perpendicular component of the spring force is zeroed, the springs do not prevent the atoms from relaxing to the MEP. The estimated barrier height, therefore, does not increase with k. In fact, over the whole range of k used in the calculations, from $k = 0.01$ to $k = 20$, the converged energy of the highest energy image was the same to five significant figures. Eventually, a problem does evolve as the springs are made extremely stiff. Very large spring forces make the minimization scheme inefficient since a very small time step is required. Also, an extremely small spring constant will make the convergence to the MEP slow. But, over three orders of magnitude for the value of the spring constant, the calculation works well and gives a good estimate of the barrier. Since we on purpose chose an even number of images in the chain and the barrier is symmetric, the highest energy image will never be sitting at the saddle point position. The energy of the highest image is a lower bound on the barrier height. An approximate upper bound can be obtained by calculating the difference between the highest and next-highest image and extrapolating. For higher

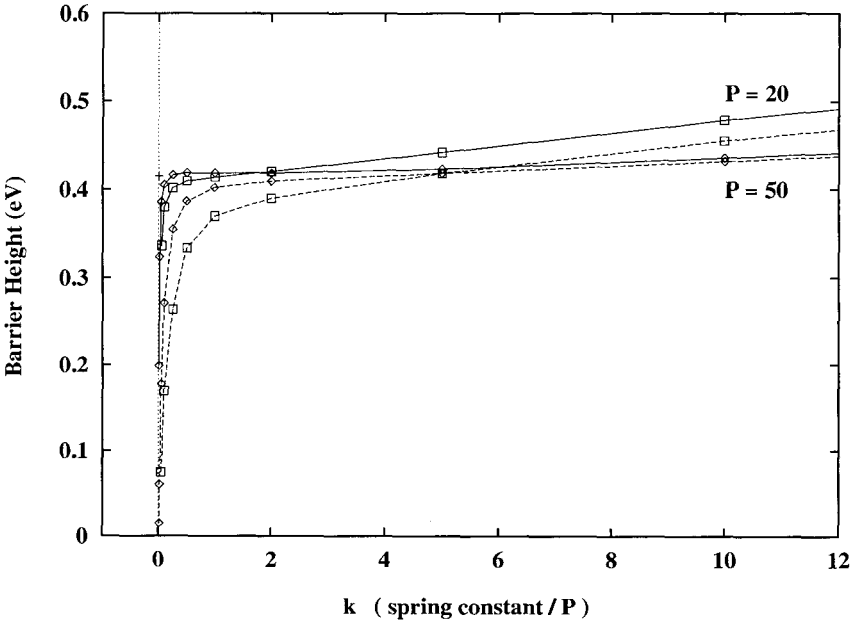


Fig. 4. – Calculation of the activation energy barrier for an adatom diffusion hop over a bridge site on the $Cu(100)$ surface using minimization of the *plain* elastic band. Results for various values of the spring constant, k, and number of images in t he chain, $P+1$, are shown. Squares are for $P = 19$ and diamonds for $P = 49$. The highest energy image (solid lines) and the next-highest one (dashed lines) are shown. The optimal spring constant, k, is close to the curvature of the potential ener gy barrier along the transition path. A mark on the $x = 0$ line shows the value obtained by relaxing the adatom at the bridge site, 4.15 eV. For large values of k, corner cutting makes the estimate of the activation energy too high. For small values of k, the images slide down from the saddle point region.

resolution, more images need to be included in the chain or a stiffer spring used near the barrier than near the ends of the path.

6. – What happens if the springs are skipped?

Since the action of the springs is only felt along the path, the value of the spring constant is not critical. For example, an equal distribution of the images will be obtained for any non-zero, constant value of the spring constant, as illustrated in the previous section. But, if the spring constant is set to zero in an NEB calculation, the control of the distribution of the images is lost. Figure 6 shows the result for model system II (see appendix). After 100 iterations with a timestep of 0.1 (fig 6a), the images have relaxed to the vicinity of the MEP, but the path is kinky and the spacing between images is uneven. The kinks fluctuate with time and a clear convergence in the forces is not obtained. This perpetual motion of the images leads to gradual sliding down from the high barrier region. After 200 steps there is a sizable gap in the barrier region (figure 6b) and after 1000 steps most of the images have slid down into the potential wells. The inclusion of the springs, which increases the computational effort insignificantly, is therefore essential for reaching

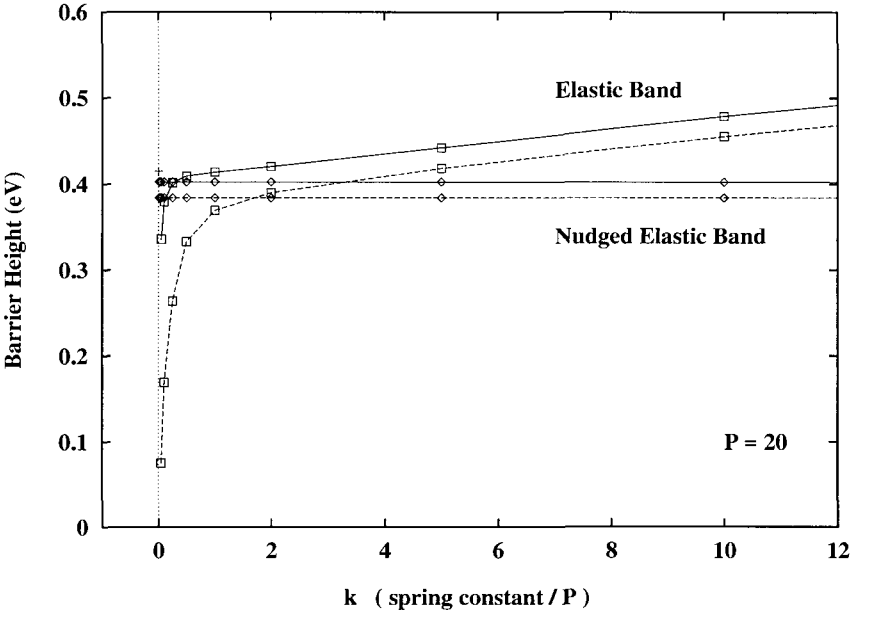


Fig. 5. – Same diffusion process as in fig. 4, but only the $P = 19$ case is shown. Results of both the *plain* elastic band method and results of the NEB method are shown. The nudging algorithm in NEB ensures that the images are evenly spaced along the path in the 3N dimensional space if the stiffness of all the springs is chosen to be the same. Also, the nudging algorithm eliminates corner cutting and leads to convergence to practically the same barrier height for values of k ranging over three orders of magnitude.

a clean convergence to an MEP and for ensuring continuity of the path especially when two or more MEPs are present.

7. – An object function for NEB

The question arises whether the NEB optimization can be written as a minimization of some object function. This is, in particular, relevant for extending the method to finite temperature simulated annealing procedures allowing searches among different MEPs to find the optimal one, i.e. the one with the lowest saddle point energy, rather than just convergence to the MEP closest to the initial guess.

We now consider a continuum representation of the path, and consider the process of relaxing the path to the MEP. The path is described by $\vec{R}(\lambda)$, a N dimensional position vector parametrized by a scalar $\lambda = [0, 1]$. N is the number of degrees of freedom participating in the transition (possibly coordinates of all the atoms in the system). Any property of the path should be 'gauge invariant', i.e. independent of the particular parametrization used in its description, such as the rate at which the path is traced out by $\vec{R}(\lambda)$ as λ is varied. In searching for an appropriate 'action', i.e. integral over the path, which can serve as an object function for the optimization, we can choose any function of the potential energy, $V(\vec{R})$, and the length, l_0, of the path. In particular, we

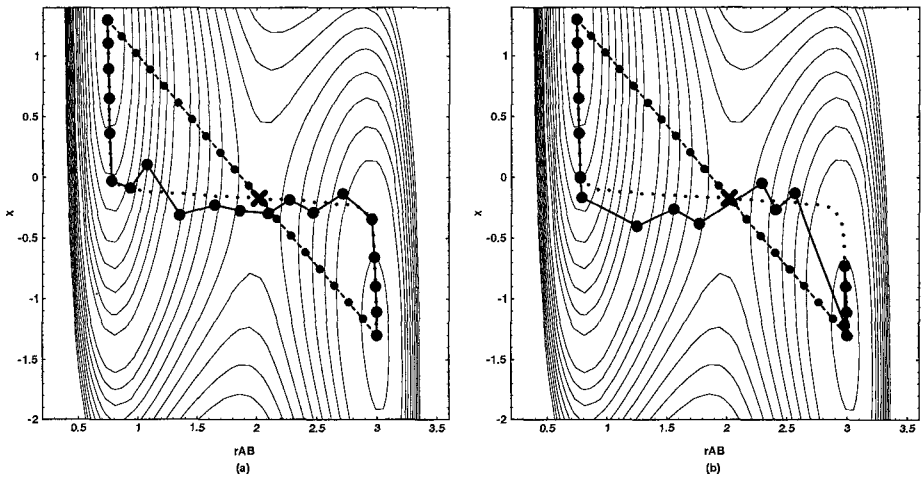


Fig. 6. – Same test problem as in figures 1 and 2. Results of optimization with no springs (k set to zero) but including the force projections of NEB is shown. The initial configuration of the images was chosen to be along the straight line connecting initial and final points. (a) After 100 time steps (using a stepsize of 1 fs). (b) After 200 time steps. Because of the kinks on the path and by a wiggling motion of the path, the images manage to slowly slide down from the barrier region.

will choose

$$S = \int_0^{l_0} dl\, G(V(\vec{R})) = \int_0^1 d\lambda\, v(\lambda)\, G(V(\vec{R}_{(\lambda)})) = \int_0^1 d\lambda\, L(\lambda) \tag{11}$$

where G is some as yet unspecified function and v is the rate at which the path is traced out, $v = \left|d\vec{R}/d\lambda\right|$. The Lagrangian $L = vG$ only includes a potential energy term at this point and gives a generalized force acting on the path

$$F_i = -\frac{\partial L}{\partial R_i} + \frac{d}{d\lambda}\left(\frac{\partial L}{\partial(dR_i/d\lambda)}\right). \tag{12}$$

Here i denotes any one of the N components describing the configuration of the atoms involved in the transition. Carrying out the differentiation gives

$$F_i = -vG'\left(\sum_{j=1}^{N}(\delta_{ij} - \tau_{\|i}\,\tau_{\|j})\frac{\partial V}{\partial R_j} - \frac{G}{G'}\,\omega_i\right) \tag{13}$$

where $\hat{\tau}_{\|}$ is the unit tangent vector to the path, $\tau_{\|} = d\vec{R}/dl = (1/v)d\vec{R}/d\lambda$, and $\hat{\omega}$ is the curvature vector, $\hat{\omega} = d\hat{\tau}_{\|}/dl$. The invariance of the action to the parametrization of the path ensures that only the perpendicular component of the potential energy gradient enters the force

$$\vec{F} = -vG'\left(\vec{\nabla}V(\vec{R})|_{\perp} - \frac{G}{G'}\,\hat{\omega}\right). \tag{14}$$

The function G should be chosen in such a way that the action is minimal for the MEP. One possible choice is

$$G(V) = e^{V/V_0}$$

where V_0 is a scaling parameter which can be used to tune the potential energy landscape. Then

$$\vec{F} = -\frac{v}{V_0} e^{V/V_0} \left(\vec{\nabla} V|_\perp - V_0 \hat{\omega}\right). \tag{15}$$

For any finite V_0, the force only involves the perpendicular component of the potential energy gradient and by minimizing S one obtains a path for which $\vec{\nabla} V|_\perp = V_0 \vec{\omega}$. By making V_0 small enough, the path can be brought arbitrarily close to the MEP, i.e. $\vec{\nabla} V|_\perp = 0$. As V_0 is lowered the forces are scaled by larger and larger prefactor $(v/V_0)\, e^{V/V_0}$ which can be thought of as a rescaling of the effective mass of the path but does not have any practical consequence (it simply determines the appropriate timestep size in the minimization).

The continuum formulation can, therefore, lead directly to the 'nudging' of the true system forces, i.e. the removal of the parallel component of the force. For practical numerical calculations, discretization of the path is needed. This means choosing a particular representation of the path. Each point in the discretization is a full image or replica of the systems N coordinates. The distribution of these images along the path is a problem which should be separate and completely decoupled from the relaxation or dynamics of the path itself. A given distribution can be obtained by introducing harmonic spring interactions between the images and choosing the spring constant between adjacent images accordingly. The spring force should act only parallel to the path, since its purpose is only to distribute the images along the path. Therefore, the force on each image, j, becomes

$$\vec{F}_j = -\frac{v}{V_0} e^{V(\vec{R}_j)/V_0} \left(\vec{\nabla} V(\vec{R}_j)|_\perp - V_0 \hat{\omega} + \vec{F}_i^s \cdot \hat{\tau}_\parallel \, \hat{\tau}_\parallel\right) \tag{16}$$

where F_i^s is given by eqn. (6). At low enough V_0 where $V_0 \hat{\omega}$ can be neglected and after absorbing the factor $(v/V_0)\, e^{V/V_0}$ into an effective mass, eqn (8) is recovered.

In large and complex systems, two or more MEPs may exist between the given initial and final points. The minimization technique described above will likely converge to the MEP closest to the initial guess. The various paths can in such cases be sampled by running simulated annealing or finite temperature molecular dynamics.

8. – Summary

To summarize, The NEB method has several desirable qualities, including (1) it converges to a MEP, given sufficient resolution in the discrete representation of the path, i.e. when enough images are included in the chain. (2) It only requires evaluation of the interaction energy and the first derivative of the energy with respect to coordinates. (3) The convergence to the MEP is decoupled from the discrete representation of the path, making the former robust and the latter flexible. (4) The method is guaranteed to give a continuous path even when multiple MEPs exist. (5) The algorithm inherently involves parallel calculations and can easily make use of parallel computers or, simply a linked cluster of workstations since very little communication is required between the computing nodes.

* * *

We gratefully acknowledge helpful discussions with Giovanni Ciccotti, Horia Metiu, Jens Nørskov and Art Voter. HJ would like to thank CAMP (Center for Atomic-scale Materials Physics) and the Physics Department of the Technical University of Denmark for hospitality and support during a sabbatical visit. This work was supported by the Division of Chemical Sciences, Office of Basic Energy Research, U. S. Department of Energy, under grant No. DE-FG06-91ER14224.

APPENDIX A.

The two-dimensional test problems

A.1. *Model I: LEPS potential.* – This model mimics a reaction involving three atoms confined to motion along a line. Only one bond can be formed, either between atoms A and B or between atoms B and C. The potential function is of the LEPS form [37]

$$V^{LEPS}(r_{AB}, r_{BC}) = \frac{Q_{AB}}{1+a} + \frac{Q_{BC}}{1+b} + \frac{Q_{AC}}{1+c} - \left[\frac{J_{AB}^2}{(1+a)^2} + \frac{J_{BC}^2}{(1+b)^2} + \frac{J_{AC}^2}{(1+c)^2} - \frac{J_{AB}J_{BC}}{(1+a)(1+b)} - \frac{J_{BC}J_{AC}}{(1+b)(1+c)} - \frac{J_{AB}J_{AC}}{(1+a)(1+c)}\right]^{\frac{1}{2}} \tag{17}$$

where the Q functions represent Coulomb interactions between the electron clouds and the nuclei and the J functions represent the quantum mechanical exchange interactions. The form of these functions is

$$Q(r) = \frac{d}{2}\left(\frac{3}{2}e^{-2\alpha(r-r_0)} - e^{-\alpha(r-r_0)}\right)$$

and

$$J(r) = \frac{d}{4}\left(e^{-2\alpha(r-r_0)} - 6e^{-\alpha(r-r_0)}\right).$$

The parameters were chosen to be $a = 0.05$, $b = 0.30$, $c = 0.05$, $d_{AB} = 4.746$, $d_{BC} = 4.746$, $d_{AC} = 3.445$, and for all three pairs we use $r0 = 0.742$ and $\alpha = 1.942$.

A contour plot of the potential surface is given in figure 3.

A.2. *Model II: LEPS + Harmonic oscillator potential.* – In model II, we fix the location of the end point atoms A and C, and only allow atom B to move. An additional degree of freedom is introduced which can be interpreted as a fourth atom which is coupled in a harmonic way to atom B.

$$V(r_{AB}, x) = V^{LEPS}(r_{AB}, r_{AC} - r_{AB}) + 2k_c(r_{AB} - (r_{AC}/2 - x/c))^2 \tag{18}$$

where $r_{AC} = 3.742$, $k_c = 0.2025$, and $c = 1.154$. Other parameters are the same as for model I, except $b = 0.80$ in model II. This type of model has frequently been used as a simple representation of an activated process coupled to a medium, such as a chemical reaction in a liquid solution or in a solid matrix.

A contour plot of the potential surface is given in figures 1,2 and 6.

REFERENCES

[1] R. Marcus, *J. Chem. Phys.* **45**, 4493 (1966).
[2] G. H. Vineyard, *J. Phys. Chem. Solids* **3** 121 (1957).
[3] M.L. McKee and M. Page, *Reviews in Computational Chemistry* Vol. IV, K.B. Lipkowitz and D.B. Boyd, Eds., (VCH Publishers Inc., New York, 1993).
[4] C. J. Cerjan and W. H. Miller, *J. Chem. Phys.* **75**, 2800 (1981).
[5] D. T. Nguyen and D. A. Case, *J. Phys. Chem.* **89**, 4020 (1985).
[6] W. Quapp, *Chem. Phys. Lett.*, **253**, 286 (1996).
[7] H. Taylor and J. Simons *J. Phys. Chem.* **89**, 684 (1985).
[8] J. Baker, *J. Comput. Chem.* **7**, 385 (1986).
[9] T. A. Halgren and W. N. Lipscomb, *Chem. Phys. Lett.*, **49**, 225 (1977).
[10] M. J. Rothman and L. L. Lohr, *Chem. Phys. Lett.* **70**, 405 (1980).
[11] S. Fischer and M. Karplus, *Chem. Phys. Lett.*, **194**, 252 (1992).
[12] I. V. Ionova and E. A. Carter, *J. Chem. Phys.*, **98**, 6377 (1993).
[13] G. Mills and H. Jónsson, *Phys. Rev. Letters* **72**, 1124 (1994); G. Mills, H. Jónsson and G. Schenter, *Surf. Sci.* **324**, 305 (1995).
[14] M. Villarba and H. Jónsson, *Surf. Sci.* **317**, 15 (1994).
[15] M. Villarba and H. Jónsson, *Surf. Sci.* **324**, 35 (1995).
[16] E. Batista and H. Jónsson, (in preparation).
[17] S. Cozzini, H. Daigler and H. Jónsson, (in preparation).
[18] M. R. Sørensen, K. W. Jacobsen and H. Jónsson, *Phys. Rev. Letters* **77** 5067 (1996).
[19] T. Rasmussen, K. W. Jacobsen, T. Leffers, O. B. Pedersen, S. G. Srinivasan, and H. Jónsson, *Phys. Rev. Letters* (in press).
[20] B. Uberuaga and H. Jónsson, (in preparation).
[21] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals*, (McGraw Hill, New York, 1965).
[22] A. Kuki and P. G. Wolynes, *Science* **236**, 1647 (1986).
[23] L. R. Pratt, *J. Chem. Phys.* **85**, 5045 (1986).
[24] C. Dellago, P. G. Bolhuis, F. S. Csajka and D. Chandler, (preprint).
[25] R. Elber and M. Karplus, *Chem. Phys. Lett.* **139**, 375 (1987).
[26] R. Czerminski and R. Elber, *Int. J. Quantum Chem.*, **24**, 167 (1990).
[27] A. Ulitsky and R. Elber, *J. Chem. Phys.*, **92**, 1510 (1990).
[28] C. Choi and R. Elber, *J. Chem. Phys.*, **94**, 751 (1991).
[29] J. J. Mortensen, Y. Morikawa, B. Hammer, and J. K. Nørskov, *Journal of Catalysis* **169**, 85 (1997).
[30] E. M. Sevick, A. T. Bell and D. N. Theodorou, *J. Chem. Phys.*, **98**, 3196 (1993).
[31] T. L. Beck, J. D. Doll and D. L. Freeman, *J. Chem. Phys.* **90**, 3183 (1989).
[32] R. E. Gillilan and K. R. Wilson, *J. Chem. Phys.*, **97**, 1757 (1992).
[33] H. C. Andersen, *J. Chem. Phys.*, **72**, 2384 (1980).
[34] A. F. Voter and S. P. Chen, *Mat. Res. Soc. Symp. Proc.* **82** 175 (1987).
[35] A. Goldstein (Ph.D. thesis, University of Washington, 1995).
[36] J. Tersoff, *Phys. Rev. B* **39**, 5566 (1989).
[37] Polanyi and Wong, *J. Chem. Phys*, **51**, 1439 (1969).