# Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points

Graeme Henkelman[a] and Hannes Jónsson[b]
*Department of Chemistry, Box 351700, University of Washington, Seattle, Washington 98195-1700*

(Received 21 August 2000; accepted 14 September 2000)

An improved way of estimating the local tangent in the nudged elastic band method for finding minimum energy paths is presented. In systems where the force along the minimum energy path is large compared to the restoring force perpendicular to the path and when many images of the system are included in the elastic band, kinks can develop and prevent the band from converging to the minimum energy path. We show how the kinks arise and present an improved way of estimating the local tangent which solves the problem. The task of finding an accurate energy and configuration for the saddle point is also discussed and examples given where a complementary method, the dimer method, is used to efficiently converge to the saddle point. Both methods only require the first derivative of the energy and can, therefore, easily be applied in plane wave based density-functional theory calculations. Examples are given from studies of the exchange diffusion mechanism in a Si crystal, Al addimer formation on the Al(100) surface, and dissociative adsorption of $CH_4$ on an Ir(111) surface. © *2000 American Institute of Physics.* [S0021-9606(00)70546-0]

## I. INTRODUCTION

The nudged elastic band (NEB) method is an efficient method for finding the minimum energy path (MEP) between a given initial and final state of a transition.[1–3] It has become widely used for estimating transition rates within the harmonic transition state theory (hTST) approximation. The method has been used both in conjunction with electronic structure calculations, in particular plane wave based density-functional theory (DFT) calculations (see, for example, Refs. 4–8), and in combination with empirical potentials.[9–12] Studies of very large systems, including over a million atoms in the calculation, have been conducted.[13] The MEP is found by constructing a set of images (replicas) of the system, typically on the order of 4–20, between the initial and final state. A spring interaction between adjacent images is added to ensure continuity of the path, thus mimicking an elastic band. An optimization of the band, involving the minimization of the force acting on the images, brings the band to the MEP. An essential feature of the method, which distinguishes it from other elastic band methods,[14–16] is a force projection which ensures that the spring forces do not interfere with the convergence of the elastic band to the MEP, as well as ensuring that the true force does not affect the distribution of images along the MEP. It is necessary to estimate the tangent to the path at each image and every iteration during the minimization, in order to decompose the true force and the spring force into components parallel and perpendicular to the path. Only the perpendicular component of the true force is included, and only the parallel component of the spring force. This force projection is referred to as ''nudging.'' The spring forces then only control the spacing of the images along the band. When this projection scheme is not used, the spring forces tend to prevent the band from following a curved MEP (because of ''corner-cutting''), and the true force along the path causes the images to slide away from the high energy regions towards the minima, thereby reducing the density of images where they are most needed (the ''sliding-down'' problem). In the NEB method, there is no such competition between the true forces and the spring forces; the strength of the spring forces can be varied by several orders of magnitude without effecting the equilibrium position of the band.

The MEP can be used to estimate the activation energy barrier for transitions between the initial and final states within the hTST approximation. Any maximum along the MEP is a saddle point on the potential surface, and the energy of the highest saddle point gives the activation energy needed for the hTST rate estimate. It is important to ensure that the highest saddle point is found, and therefore some information about the shape of the MEP is needed. It is quite common to have MEPs with one or more intermediate minima and the saddle point closest to the initial state may not be the highest saddle point for the transition.

While the NEB method is robust and has been successful, there are situations where the elastic band does not converge well to the MEP. When the force parallel to the MEP is large compared to the force perpendicular to the MEP, and when many images are used in the elastic band, kinks can form on the elastic band. As the minimization algorithm is applied, the kinks can continue to oscillate back and forth, preventing the band from converging to the MEP. Previously, this problem was reduced by including some fraction of the perpendicular spring force when the tangent was changing appreciably between adjacent images in the band.[3] A switching function was added which gradually introduced the parallel spring component as the change in the tangent increased. The problem with this approach is that the elastic band then tends to be pulled off the MEP when the path is

[a]Electronic mail: graeme@u.washington.edu
[b]Electronic mail: hannes@u.washington.edu

 

curved. If the path is curved at the saddle point, this can lead to an overestimate of the saddle point energy. This problem was, in particular, observed in calculations of the exchange diffusion process in a Si crystal (discussed below).

In this paper we present an analysis of the origin of the kinks and give a simple solution to the problem. With a different way of estimating the local tangent to the elastic band, the tendency to form kinks disappears. We also address the issue of finding a good estimate of the saddle point given the converged elastic band. An estimate of the saddle point configuration is obtained from the pair of images near the saddle point by interpolation. The dimer method[17] is then used to converge on the saddle point. In cases where the energy barrier is narrow, and few images are located near the saddle point, the estimate obtained from the elastic band alone may not be reliable. When the saddle point is found rigorously with the dimer method, it is not necessary to converge the elastic band all the way to the MEP. The optimal combination of partial minimization of the elastic band and subsequent dimer method calculations is discussed. The methodology is presented with reference to simple two-dimensional model potentials. We also present results of calculations on three realistic systems: the exchange of Si atoms in a Si crystal, the formation of an Al addimer on the Al(100) surface, and the dissociation of a $CH_4$ molecule on the Ir(111) surface. We have used both empirical potentials and DFT calculations in these studies.

## II. THE ORIGINAL IMPLEMENTATION OF NEB

An elastic band with $N+1$ images can be denoted by $[\mathbf{R}_0,\mathbf{R}_1,\mathbf{R}_2,\ldots,\mathbf{R}_N]$ where the endpoints, $\mathbf{R}_0$ and $\mathbf{R}_N$, are fixed and given by the energy minima corresponding to the initial and final states. The $N-1$ intermediate images are adjusted by the optimization algorithm.

In the original formulation of the NEB method, the tangent at an image $i$ was estimated from the two adjacent images along the path, $\mathbf{R}_{i+1}$ and $\mathbf{R}_{i-1}$. The simplest estimate is to use the normalized line segment between the two

$$\hat{\boldsymbol{\tau}}_i=\frac{\mathbf{R}_{i+1}-\mathbf{R}_{i-1}}{|\mathbf{R}_{i+1}-\mathbf{R}_{i-1}|}, \tag{1}$$

but a slightly better way is to bisect the two unit vectors

$$\boldsymbol{\tau}_i=\frac{\mathbf{R}_i-\mathbf{R}_{i-1}}{|\mathbf{R}_i-\mathbf{R}_{i-1}|}+\frac{\mathbf{R}_{i+1}-\mathbf{R}_i}{|\mathbf{R}_{i+1}-\mathbf{R}_i|}, \tag{2}$$

and then normalize $\hat{\boldsymbol{\tau}}=\boldsymbol{\tau}/|\boldsymbol{\tau}|$. This latter way of defining the tangent ensures the images are equispaced (when the spring constant is the same) even in regions of large curvature. The total force acting on an image is the sum of the spring force along the tangent and the true force perpendicular to the tangent

$$\mathbf{F}_i=\mathbf{F}_i^s|_{\parallel}-\nabla V(\mathbf{R}_i)|_{\perp}, \tag{3}$$

where the true force is given by

$$\nabla V(\mathbf{R}_i)|_{\perp}=\nabla V(\mathbf{R}_i)-\nabla V(\mathbf{R}_i)\cdot\hat{\boldsymbol{\tau}}_i, \tag{4}$$

and the spring force was, in the simplest version of the method, calculated as

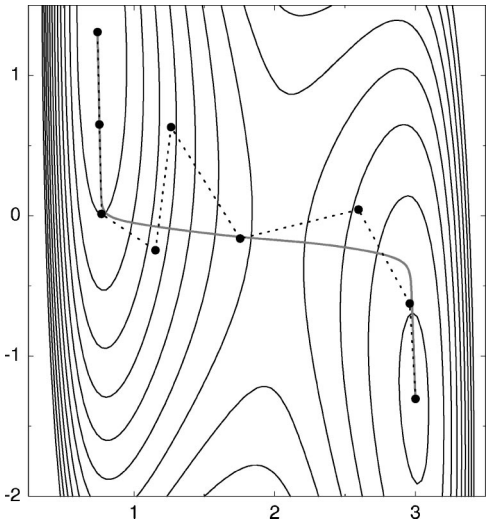


FIG. 1. The original nudged elastic band method as described by Eqs. (2) and (3) can develop kinks along the path as illustrated here for a two-dimensional LEPS model potential. The kinks do not get smaller in this case, but fluctuate as the minimization is continued. The band does not converge to the minimum energy path (solid line).

$$\mathbf{F}_i^s|_{\parallel}=k[(\mathbf{R}_{i+1}-\mathbf{R}_i)-(\mathbf{R}_i-\mathbf{R}_{i-1})]\cdot\hat{\boldsymbol{\tau}}_i\hat{\boldsymbol{\tau}}_i. \tag{5}$$

When the perpendicular component of the spring force is not included, as in Eq. (5), the path forms kinks in some cases. An illustration of this is given in Fig. 1. The potential surface is a two-dimensional combination of a LEPS potential[18] and a harmonic oscillator (the functional form and parameters are given in Ref. 3). Initially, the seven movable images were placed equispaced along the straight line between the minima. A minimization algorithm was then applied, moving the images in the direction of the force given by Eq. (3). We use a projected velocity Verlet algorithm for the minimization.[3] After a few iterations the images are close to the MEP, but after a certain level of convergence is reached, the magnitude of the force on the images does not drop any further. Kinks form in regions where the parallel component of the force is large compared with the perpendicular component, often near the inflection points of the energy curve for the MEP. As the minimization algorithm is continued, the kinks simply fluctuate back and forth. The saddle point region, and thereby the estimate of the saddle point energy, were typically not affected significantly by the kinks. In some systems, however, this turned out to be a serious problem.

In order to reduce the kinks and enable the minimization to reduce the magnitude of the forces to a prescribed tolerance, some fraction of the perpendicular component of the spring force was included when the angle between the vectors $\mathbf{R}_i-\mathbf{R}_{i-1}$ and $\mathbf{R}_{i+1}-\mathbf{R}_i$ deviated appreciably from zero. A switching function of the angle was used to multiply the perpendicular spring force component.[3] At kinks the angle is large, and addition of some of the perpendicular spring force tends to straighten the elastic band out. The problem is that this leads to corner-cutting in regions where the MEP is curved. If the saddle point is in such a curved region, then the addition of a fraction of the perpendicular spring force

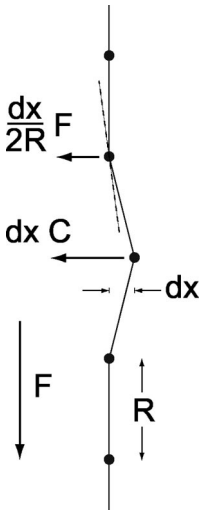


FIG. 2. An illustration of the cause of the kinks that can develop on the elastic band in the original formulation when the force component parallel to the path is large while the perpendicular component is small. The sketch indicates the critical condition for stability as given by Eq. (6).

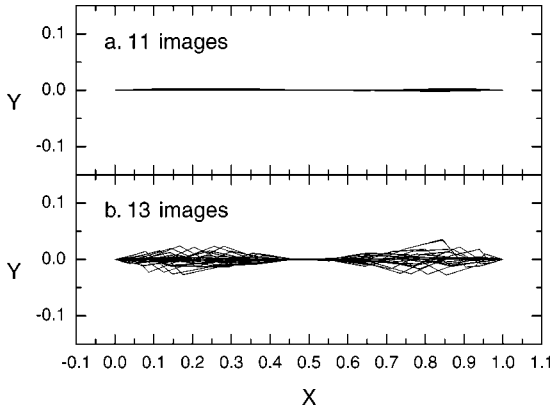


FIG. 3. The effect of the number of images on the stability of the nudged elastic band in the original formulation. The calculations were performed on the cosine potential given in Eq. (7). (a) When 11 images are used, the stability condition given by Eq. (6) is satisfied and the elastic band converges to the minimum energy path. The initial configuration, which deviates slightly from the minimum energy path, is shown as well as a few configurations taken from the minimization. (b) With 13 images, the stability is borderline and the condition given by Eq. (7) violated. The small initial deviations of the band from the minimum energy path are magnified as large kinks form in the neighborhood of the inflection points. The kinks oscillate during the minimization until the band eventually settles to the minimum energy path.

will lead to an overestimate of the saddle point energy. In many cases this overestimate is not a serious problem. Nevertheless, it is still an inconvenience; the switching function means that there is an extra parameter in the implementation of the NEB and there is some uncontrollable error in the estimate of the activation energy. The exchange diffusion process in Si is one example where the MEP is so curved at the saddle point that the switching function leads to a significant overestimate of the activation energy, as discussed below. Before discussing a way to solve this problem, we analyze the origin of the kinks in the following section.

## III. ANALYSIS OF KINKS ON THE ELASTIC BAND

We now examine more closely the origin of the kinks on the elastic band, shown for example in Fig. 1. In many calculations the original NEB method converges very well, even without a switching function, but in some situations it is unstable. Figure 2 shows five images, separated by a distance $R$, in an idealized NEB calculation. It is assumed that additional images are located above and below the segment shown. The MEP is assumed to be straight in this region, and the top two and bottom two images shown are lying on the MEP. The potential energy surface is assumed to have a constant slope in the direction of the MEP and the force in that direction is given by the constant $F$. The potential energy perpendicular to the MEP is assumed to be quadratic with a restoring force of $-xC$ where $x$ is the perpendicular distance from the MEP, and $C$ is the curvature. The middle image in Fig. 2 is displaced a small distance $dx$ from the MEP. If the original way of estimating the tangent [Eq. (2)] is used, there are two competing effects acting on the middle image. The first is the tendency for the displaced image to move back under the restoring force $-dxC$. The second is for the higher energy neighboring image to move away from the MEP because the estimated tangent at the higher image is no longer along the MEP and the force $F$ now has a small perpendicular component, $dx/2R$. The band becomes unstable when the restoring force $-dxC$ is less than the destabilizing force $dx/2RF$. This leads to the stability condition

$$F<2CR. \tag{6}$$

Since the spacing between the images, $R$, depends on the number of images used in the band, this stability condition predicts that the band always becomes unstable if the number of images is made large enough (in the absence of the switching function).

The stability condition was tested on a simple two-dimensional system with a straight MEP. The potential energy function is

$$V(x,y)=A_x\cos(2\pi x)+A_y\cos(2\pi y). \tag{7}$$

An MEP lies along the straight line between (0,0) and (1,0). The maximum force along the path is $F=2\pi A_x$ and the curvature perpendicular to the path is $C=4\pi^2A_y$. Choosing $A_x=A_y=1$ yields the stability condition $R>\frac{1}{4}\pi\approx\frac{1}{13}$. This corresponds to a maximum of 12 images in the elastic band before it becomes unstable.

This prediction is born out by the calculations. A chain with 11 images with small initial perturbations from the MEP, proved to be stable and converged monotonically to the straight path as shown in Fig. 3(a). A chain with 13 images, on the other hand, took much longer to converge and developed large fluctuations, bigger than the initial displacement, before eventually converging to the MEP. Several snapshots of the calculation are shown in Fig. 3(b), and illustrate the typical magnitude of oscillations during the simulation.

When 25 images were used in the band, very large kinks developed so that images were able to slide down into the minima at (0,0) and (1,0) during the minimization. This continued until the number of images located between the

minima had dropped and the spacing between them increased enough to satisfy the stability condition, Eq. (6). Then the images at the minima were slowly pulled into place by the spring force. In this way, chains with up to $\sim 80$ images were able to (slowly) converge. For more images, the minimization would not converge and some images were always left in a jumble at the minima.

In order to further test the stability condition, the restoring force perpendicular to the path was doubled by setting $A_y = 2$. The stability condition predicts that a band with twice the number of images, 24, will remain stable. Calculations showed that the band becomes unstable at 21 images. The good agreement between the simple prediction of Eq. (6) and these simulations suggest that this is the correct origin of the kinks. The modified NEB method with the new tangent, presented in the next section, converges well for both small and large numbers of images.

## IV. THE NEW IMPLEMENTATION OF NEB

By using a different definition of the local tangent at image *i*, the kinks can be eliminated. Instead of using both the adjacent images, $i+1$ and $i-1$, only the image with higher energy and the image *i* are used in the estimate. The new tangent, which replaces Eq. (2), is

$$\boldsymbol{\tau}_i = \begin{cases} \boldsymbol{\tau}_i^+ & \text{if} \quad V_{i+1} > V_i > V_{i-1} \\ \boldsymbol{\tau}_i^- & \text{if} \quad V_{i+1} < V_i < V_{i-1} \end{cases}, \tag{8}$$

where

$$\boldsymbol{\tau}_i^+ = \mathbf{R}_{i+1} - \mathbf{R}_i, \quad \text{and} \quad \boldsymbol{\tau}_i^- = \mathbf{R}_i - \mathbf{R}_{i-1}, \tag{9}$$

and $V_i$ is the energy of image *i*, $V(\mathbf{R}_i)$. If both of the adjacent images are either lower in energy, or both are higher in energy than image *i*, the tangent is taken to be a weighted average of the vectors to the two neighboring images. The weight is determined from the energy. The weighted average only plays a role at extrema along the MEP and it serves to smoothly switch between the two possible tangents $\boldsymbol{\tau}_i^+$ and $\boldsymbol{\tau}_i^-$. Otherwise, there is an abrupt change in the tangent as one image becomes higher in energy than another and this can lead to convergence problems. If image *i* is at a minimum $V_{i+1} > V_i < V_{i-1}$ or at a maximum $V_{i+1} < V_i > V_{i-1}$, the tangent estimate becomes

$$\boldsymbol{\tau}_i = \begin{cases} \boldsymbol{\tau}_i^+ \Delta V_i^{\max} + \boldsymbol{\tau}_i^- \Delta V_i^{\min} & \text{if} \quad V_{i+1} > V_{i-1} \\ \boldsymbol{\tau}_i^+ \Delta V_i^{\min} + \boldsymbol{\tau}_i^- \Delta V_i^{\max} & \text{if} \quad V_{i+1} < V_{i-1} \end{cases}, \tag{10}$$

where

$$\Delta V_i^{\max} = \max(|V_{i+1} - V_i|, |V_{i-1} - V_i|)$$

and (11)

$$\Delta V_i^{\min} = \min(|V_{i+1} - V_i|, |V_{i-1} - V_i|).$$

Finally, the tangent vector needs to be normalized. With this modified tangent, the elastic band is well behaved and converges rigorously to the MEP if sufficient number of images are included in the band.

Another small change from the original implementation of the NEB is to evaluate the spring force as

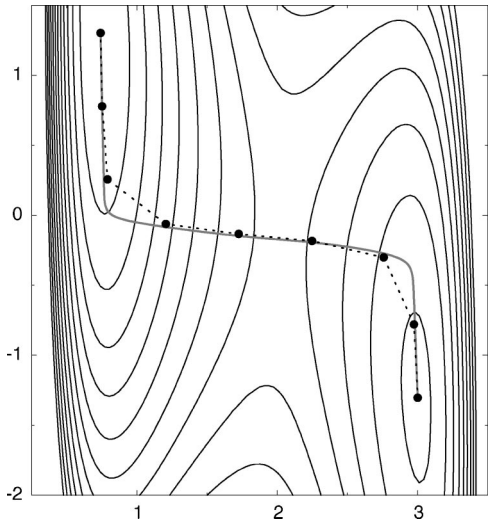


FIG. 4. With the modified tangent, Eqs. (8)–(12), the nudged elastic band method does not develop kinks and converges smoothly to the minimum energy path (solid line).

$$\mathbf{F}_i^s|_{\parallel} = k(|\mathbf{R}_{i+1} - \mathbf{R}_i| - |\mathbf{R}_i - \mathbf{R}_{i-1}|)\hat{\boldsymbol{\tau}}_i \tag{12}$$

instead of Eq. (5). This ensures equal spacing of the images (when the same spring constant, *k*, is used for the springs) even in regions of high curvature where the angle between $\mathbf{R}_i - \mathbf{R}_{i-1}$ and $\mathbf{R}_{i+1} - \mathbf{R}_i$ is large.

When this modified NEB method is applied to the system of Fig. 1 the band is well behaved as shown in Fig. 4. The energy and force of the NEB images is shown in Fig. 5 along with the exact MEP obtained by moving along the gradient from the saddle point. The thin line though the points is an interpolation which involves both the energy and the force along the elastic band. The details of the interpolation procedure is discussed in the Appendix.

The motivation for this modified tangent came from a stable method for finding the MEP from a given saddle point. A good way to do this is to displace the system by some increment from the saddle point and then minimize the energy while keeping the distance between the system and the saddle point configuration fixed. This gives one more point along the MEP, say $M_1$. Then, the system is displaced again by some increment and minimized keeping the distance to the point $M_1$ fixed, etc. The important fact is that the MEP can be found by following force lines down the potential from the saddle point, but never up from a minimum. If a force line is followed up from a minimum, it will most likely not go close to the saddle point. This motivated the choice for the local tangent to be determined by the higher energy neighboring image in the NEB method.

### A. Exchange diffusion process in Si crystal

A particularly severe problem with kinks had been noticed in calculations of self diffusion in a Si crystal. For example, one possible mechanism is the exchange of two atoms on adjacent lattice sites.[20] Both calculations using DFT and empirical potentials could not converge the force because of kinks. In calculations using 16 images and the Tersoff potential,[19] the force fluctuations remained at a level

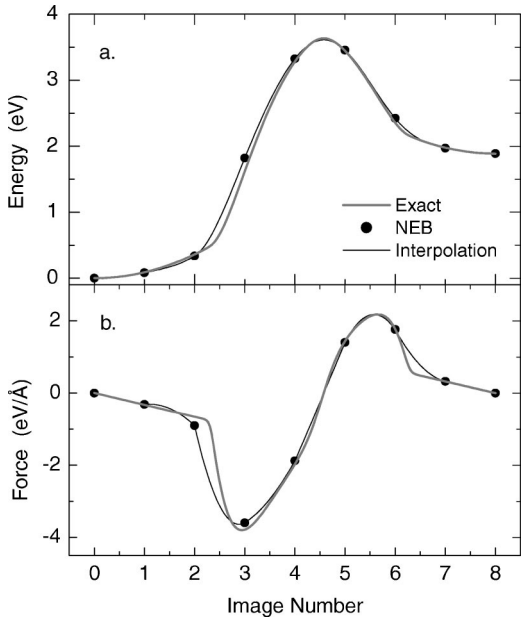


FIG. 5. The energy and force of the converged elastic band for the two-dimensional LEPS potential shown in Fig. 4. An interpolation between the images (thin line), see the Appendix, agrees well with the exact minimum energy path (thick, gray line).

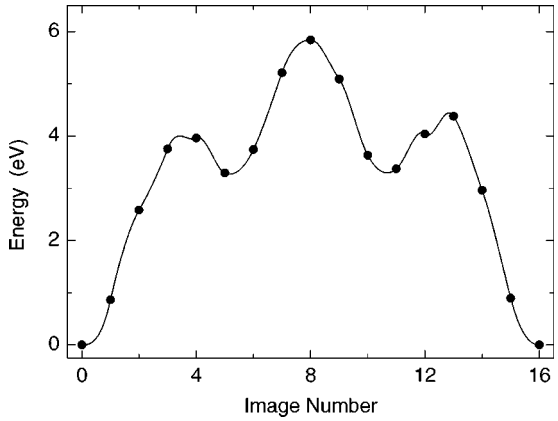


FIG. 6. The minimum energy path for the Pandey exchange process in silicon crystal modeled with the Tersoff potential. While the original formulation of the nudged elastic band did not converge to the minimum energy path in this case unless a large fraction of the perpendicular spring force was included (leading to an overestimate of the activation energy), the modified formulation with the new tangent converges well. The potential energy surface has several small ripples and local minima. The local minima near images 4 and 12, suggested by the interpolated energy curve, are indeed present on the Tersoff potential energy surface.

of 0.5 eV/Å (the average total force on all the images). When a switching function was used to add a fraction of the perpendicular component of the spring force, the saddle point estimate was clearly strongly dependent on how much spring force was added, indicating a problem with corner-cutting. The reason why diffusion processes in Si are particularly difficult is that the energy changes rapidly as the covalent bonds get broken, but the lattice is quite open and there is a small restoring force for bending bonds. The problem becomes worse when the Tersoff potential function is used because the potential-energy surface has several bumps and minima, presumably because of sharply varying cut-off functions. The MEP takes sharp detours into these local minima and this can lead to large angles between neighboring images.

Figure 6 shows a calculation carried out with the new formulation of the NEB. The new tangent was found to be resilient to the large angles on the MEP, some larger than 90° between adjacent images. The calculation converged to $<0.01$ eV/Å in 500 iterations. The minima near images 5 and 11 in Fig. 6 are similar to results obtained by DFT, and probably correspond to true metastable configurations of the rotating pair of Si atoms. The small dips predicted by the interpolation near images 4 and 12 are indeed minima on the Tersoff potential surface, but do not show up in DFT calculations and are likely artifacts in the empirical potential. The presence of the intermediate minima illustrates the importance of getting information about the full MEP, rather than just converging on the saddle point nearest to the initial state of the transition.

## V. COMBINING THE DIMER AND NEB METHODS

When estimating rates in hTST, the activation energy is determined only by the highest energy along the MEP. The full MEP is only needed to locate the highest maximum, when two or more maxima are present. Finding the precise value of the energy at the maximum saddle point can be tedious with the NEB method. Enough images need to be included to get high enough resolution near the maximum for the interpolation to give an accurate value. Also, many force evaluations can be wasted converging images far from the saddle point, which in the end are not relevant. It can, therefore, be advantageous to first use only a few iterations of the NEB method, enough to get a rough estimate of the shape of the MEP, and then turn to another method which can efficiently converge on the highest saddle point.

The dimer method[17] is a good candidate for this cooperative strategy. Similar to the NEB method, it requires only forces to find a saddle point from an initial configuration. We have used the interpolation scheme discussed in the Appendix to estimate the coordinates of the saddle point configuration from the two images closest to the maximum. This configuration was then used as a starting point for the dimer method. The tangent to the path at the estimated saddle point (obtained from the interpolation) was also used as the initial orientation of the dimer, i.e., as a guess for the direction of the lowest curvature normal mode. The dimer method was implemented as described in Ref. 17.

### A. Formation of Al addimer on Al(100) surface

The strategy of combining the NEB and dimer methods was tested on a process where an addimer forms on the Al(100) surface. A surface atom gets pulled out of the surface layer and ends up forming an addimer with an adatom that is present in the initial state. The atomic interactions were described by an empirical embedded atom method (EAM)[21] potential. An extensive study of the various transitions involving an adatom on this surface have been calculated with the dimer method and are described elsewhere.[17] We chose this process because it is complex and the saddle

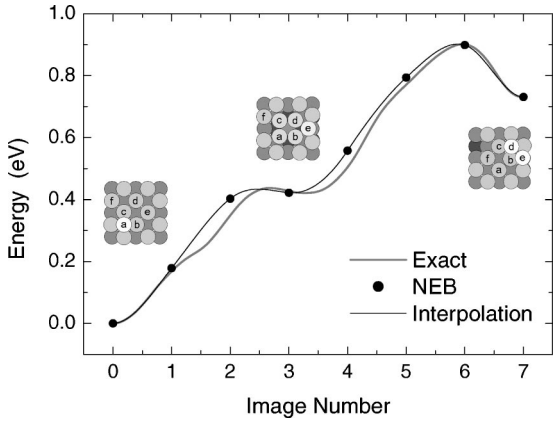


FIG. 7. The minimum energy path for a process where an Al adatom on an Al(100) surface bonds with one of the surface atoms to form an addimer. An EAM potential is used to describe the interatomic interactions. The path is highly curved. The insets show the atomic configuration in initial, final and intermediate stable state. The best estimate of the minimum energy path, obtained by sliding down the steepest descent path from the saddle point is shown (thick, gray line) as well as the energy of an elastic band with six movable images (filled circles) and interpolation between the images (solid line).

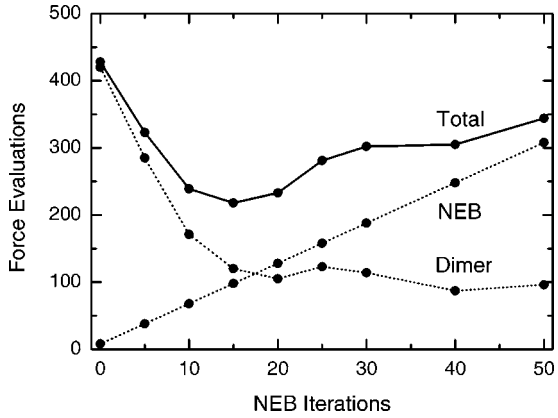


FIG. 8. The computational effort (given as number of force evaluations) involved in finding the saddle point for the Al addimer formation process shown in Fig. 7. The $x$ axis shows the number of iterations in a nudged elastic band calculation with six movable images, starting from a straight line interpolation between the initial and final states. The slope of the straight dashed line labeled ''NEB'' is the six force evaluations needed per iteration in the minimization of the elastic band. After every five iterations the interpolation procedure described in the Appendix was used to obtain an estimate of the saddle point energy and configuration, and this was used as a starting point for a search for the saddle point using the dimer method. The dashed curve labeled ''dimer'' gives the number of force evaluations needed to converge on the saddle point in each case. The total number of force evaluations, from the initial $5n$ steps of the elastic band minimization plus the dimer calculation, is given by the solid line labeled ''Total.'' The results show that an optimal strategy for finding the saddle point in this case is to carry out 15 iterations of the elastic band minimization, and then a dimer search.

point is quite far removed from a straight line interpolation between the initial and final states. The process involves an intermediate state that is barely stable (see Fig. 7). First, the adatom gets incorporated into the surface to form a shallow but stable structure where a group of four surface atoms forming a square has rotated by 45°. Then, an addimer is formed leaving a vacancy in the surface layer. Figure 7 shows a fully converged NEB with eight images, the interpolated energy and the energy obtained by descending from the saddle point along the MEP. The eight image NEB clearly gives a good representation of the MEP.

We carried out a systematic study of the efficiency of combining the NEB and dimer methods. Figure 8 shows the results. First, the NEB method was run to convergence, starting from a straight line interpolation between the initial and final states. This took 50 iterations. The computational effort is measured by the number of force evaluations, since over 90% of the computer time is used to evaluate the force on the various configurations. The straight line in Fig. 8 simply shows how the total number of force evaluations increases as more iterations are performed in the NEB calculation (since there are six movable images in the band, it takes six force evaluations to carry out one iteration of the NEB method). Configurations were saved every five iterations of the NEB calculation, and for each the interpolation was used to obtain an estimate of the saddle point configuration. From this configuration a dimer calculation was started and run until the saddle point was found to within the given tolerance of 1 meV/Å. The number of force evaluations needed in the dimer calculation is given in Fig. 8, each point located along the $x$ axis at the number of NEB iterations that had been carried out before the dimer calculation was started. The first point, corresponding to zero NEB iterations, corresponds to a calculation where the energy and force were calculated for the eight images along the straight line interpolation from the initial to the final states (including end points). The saddle point configuration was estimated from this data and it took 420 force evaluations for the dimer method to converge to the saddle point. The total number of force evaluations for this calculation was, therefore, 428. This entire procedure was repeated starting from an elastic band configuration obtained after five NEB iterations, etc. The most efficient procedure for this particular case, corresponding to a little over 200 force evaluations, involved 15 NEB iterations and then convergence to the saddle point with the dimer method.

The two methods complement each other well. When the final state is known, this provides important information for the saddle point search. The NEB method incorporates this information. The dimer method can be used to search for a saddle point starting in the vicinity of the initial state minimum, but in a high-dimensional system with multiple intermediate minima, the calculation of a complete MEP with the dimer method is tedious. The dimer method would converge on the first saddle point near the initial state, then a search for the intermediate minimum has to be carried out, followed by another dimer search for the next saddle point, etc. When the final state is known, using the NEB method for at least a few iterations is better than solely using the dimer method from the initial or final state. The NEB gives a nonlocal picture of the potential energy landscape, and can reveal intermediate stable states which can be important for understanding the basic mechanism of the transition and possibly suggest different mechanisms which lead to alternate final states. The interpolation using the force along the elastic band makes the NEB method able to reveal such intermediate minima even when very few images are included in the

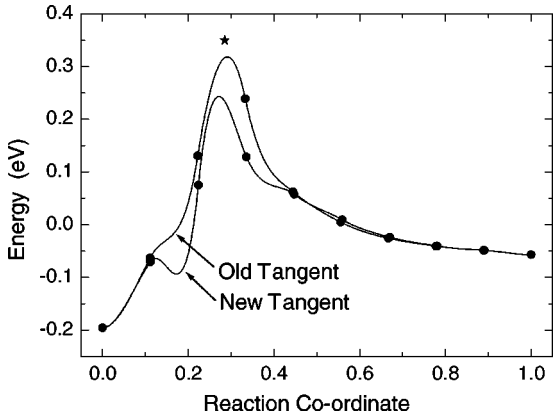


FIG. 9. A density-functional theory calculation of the dissociative adsorption of a $CH_4$ molecule on an Ir(111) surface. Converged elastic bands with eight movable images are shown using both the original way of estimating the tangent, and using the new way of estimating the tangent. An interpolation using the method described in the Appendix is shown with a solid line. With the new tangent, the interpolation correctly predicts an intermediate minimum (near reaction coordinate of 0.2). The resolution of the barrier is, however, too poor and the estimate of the saddle point obtained from the interpolation is much too low. The saddle point energy ($\star$) was found with a dimer calculation started from the estimated saddle point configuration given by the interpolation.

band. Ultimately, it is the energy of the highest saddle point that is needed, and unless one of the NEB images happens to land very close to the saddle point, the estimate of the saddle point energy from the NEB calculation relies on interpolation. Then, a more rigorous convergence to the saddle point can be obtained with the dimer method.

### B. Dissociation of $CH_4$ on Ir(111)

We have used the combination of NEB and dimer methods in plane wave based DFT[22,23] calculations of the dissociative adsorption of a $CH_4$ molecule on an Ir(111) surface. The VASP code was used for the calculations.[24] We used the PW91 functional[25] and ultrasoft pseudopotentials.[26] A parallel implementation of the original formulation of the NEB method is included in the current VASP code. We modified the code to incorporate the new formulation [Eqs. (8), (10), and (12)]. We have also implemented the dimer method in the VASP code, making use of the natural parallelization over two processors (one for each image in the dimer). An Ir slab with five close packed layers was simulated with the top three layers of the slab allowed to relax while the bottom two were kept fixed in bulk crystal configuration. Each layer contained 12 atoms.

The results of calculations using the original formulation of the NEB and the new formulation are shown in Fig. 9. The $H^-$ and $CH_3^-$ adsorbed at adjacent on-top sites correspond to reaction coordinate of 0.0 in the figure. The $CH_4$ molecule is 4.0 Å above the Ir surface when the reaction coordinate is 1.0. Moving the molecule further away from the surface did not change the energy. The DFT/PW91 calculations include correlation of the electrons only semilocally. The long range dispersion interaction at distances where the overlap of the electron density of the Ir surface and the $CH_4$ molecule is small and is not included in the DFT/PW91 calculation. There is, therefore, no physisorption well evident in the results shown in Fig. 9. The activation energy barrier, however, occurs when the electron clouds overlap strongly, and the semilocal description of correlation should be adequate there. The height of the activation energy barrier, 0.4 eV, is in reasonable agreement with the activation energy deduced from experimental measurements.[27] A more detailed analysis of the calculation and comparison with experiment will be given elsewhere.[28]

A comparison of the results from the new and old formulation of the NEB is shown in Fig. 9. Since the barrier is quite narrow compared with the length of the path from the free molecule to the dissociated fragments, the number of images in the barrier region is low. The elastic band obtained from the new formulation predicts a stable local minimum in the region of reaction coordinate 0.2. There is a minimum in the interpolated MEP due to the value of the parallel force at the images. This minimum is indeed there, only deeper, and was verified by a separate calculation where the system coordinates where simply minimized starting from the prediction obtained from the interpolation between the elastic band images. We have carried out several calculations of the dissociation with varying number of layers in the slab and various numbers of atoms in each slab, both of which have significant effect on the barrier height, and in all those cases the minimum is predicted by the new formulation but not by the old one. This illustrates how much closer the elastic band with the new tangent follows the MEP.

The number of images in the barrier region is, however, too small to predict accurately the barrier height from the interpolation. This is a good example of where the combined NEB and dimer approach is advantageous. The dimer method gave a saddle point energy that is higher than either interpolation predicts, as shown in Fig. 9.

## VI. SUMMARY

We have presented a new formulation of the NEB method which is more stable than the previous formulation. With a different way of estimating the local tangent to the elastic band, the problems with kinks on the elastic band can be eliminated. While in many systems this is was not a serious problem, there were examples where the previous procedure of including part of the perpendicular spring force led to serious overestimate of the saddle point energy. The exchange of two Si atoms in the Si crystal is one example. With the new formulation, the calculation of the MEP for the exchange process is straight forward.

We have also described a strategy where the NEB is combined with the dimer method. A few iterations using the NEB, enough to get a rough estimate of the shape of the MEP, followed by interpolation between the images and then a search for the saddle point using the dimer method was found to cut the overall computational effort in half in a test problem involving dimer formation on an Al(100) surface. This strategy was also applied to a calculation of methane dissociation on an Ir(111) surface using plane wave based DFT calculations.

## ACKNOWLEDGMENTS

We would like to thank Blas Uberuaga for helpful discussions, in particular regarding the implementation of the dimer method in VASP. We thank Roland Stumpf for making his interpolation scheme (which is slightly different from what we describe here) and graphics routines for analysis of NEB calculations in VASP available to us, as well as many useful suggestions on NEB calculations with DFT. This work was funded by the National Science Foundation Grant No. CHE-9710995 and by the Petroleum Research Fund Grant No. PRF#32626-AC5/REF#104788.

## APPENDIX

The analysis of the results of an NEB calculation requires an interpolation between the images in order to get estimates of the coordinates of atoms and the energy at maxima and minima along the MEP. In addition to the energy at each image, the force parallel to the band is also known and provides useful information for the interpolation. From a good interpolation scheme, the presence of maxima and minima on the MEP can be deduced from remarkably few images. In some cases, the use of the force can lead to the identification of extrema in the MEP which could not be seen when only the energy data were used in the interpolation.

We have used a cubic polynomial to represent the MEP between each pair of adjacent images. There are four parameters in the polynomial which are determined by matching the energy and force at ends of the interval. Letting the length of the $i$th segment be $R_i$ and writing the polynomial as $a_i x^3 + b_i x^2 + c_i x + d_i$ the parameters are

$$\begin{aligned} a_i &= \frac{2V_{i+1} - V_i}{R^3} - \frac{F_i + F_{i+1}}{R^2}, \\ b_i &= \frac{3V_{i+1} - V_i}{R^2} + \frac{2F_i + F_{i+1}}{R}, \\ c_i &= -F_i, \\ d_i &= V_i, \end{aligned} \tag{A1}$$

where the energy at the images is $V_i$ and $V_{i+1}$ and parallel force is $F_i$ and $F_{i+1}$.

This interpolation is typically quite smooth. However, unlike traditional cubic spline interpolations,[29] the second derivative is not continuous at the images. A close look at the interpolation in Fig. 5 shows the discontinuity in the slope of the force. A possible improvement in the scheme described above is to generate a quintic polynomial interpolation so that the second derivatives can also be made continuous (and set to zero at the end points for a natural spline). It was found that although this did generate a smoother interpolation, it sometimes added spurious oscillations in between the images. If a continuous second derivative is not essential, the simple cubic interpolation described above is probably preferred.

The interpolation scheme described above can also be used to estimate the configuration of the system at various points along the MEP. Each atomic coordinate, $\mathbf{R}_i^j$, is a function of the distance along the path, $s$. The slope $d\mathbf{R}_i/ds$ is taken to be along the force at the NEB images, $d\mathbf{R}_i^j/ds = \mathbf{F}_i^j/|\mathbf{F}_i|$, except for the endpoints. At the endpoints, the force is zero and the direction of the path is, therefore, not well defined. There, we assumed the path was pointing towards the neighboring images. The equations for the coordinate interpolation are analogous to those given above for the energy interpolation, with the modifications $V_i \rightarrow \mathbf{R}_i$, and $F_i \rightarrow -d\mathbf{R}_i/ds = -\mathbf{F}_i/|\mathbf{F}_i|$. Each line in Eq. (A1) becomes a vector equation.

[1] G. Mills and H. Jónsson, Phys. Rev. Lett. **72**, 1124 (1994).
[2] G. Mills, H. Jónsson, and G. K. Schenter, Surf. Sci. **324**, 305 (1995).
[3] H. Jónsson, G. Mills, and K. W. Jacobsen, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne, G. Ciccotti, and D. F. Coker (World Scientific, Singapore, 1998), p. 385.
[4] B. Uberuaga, M. Levskovar, A. P. Smith, H. Jónsson, and M. Olmstead, Phys. Rev. Lett. **84**, 2441 (2000).
[5] J. Song, L. R. Corrales, G. Kresse, and H. Jónsson (to be published).
[6] W. Windl, M. M. Bunea, R. Stumpf, S. T. Dunham, and M. P. Masquelier, Phys. Rev. Lett. **83**, 4345 (1999).
[7] R. Stumpf, C. L. Liu, and C. Tracy, Phys. Rev. B **59**, 16047 (1999).
[8] T. C. Shen, J. A. Steckel, and K. D. Jordan, Surf. Sci. **446**, 211 (2000).
[9] M. Villarba and H. Jónsson, Surf. Sci. **317**, 15 (1994).
[10] M. Villarba and H. Jónsson, Surf. Sci. **324**, 35 (1995).
[11] E. Batista and H. Jónsson, Computational Materials Science (to be published).
[12] M. R. Sørensen, K. W. Jacobsen, and H. Jónsson, Phys. Rev. Lett. **77**, 5067 (1996).
[13] T. Rasmussen, K. W. Jacobsen, T. Leffers, O. B. Pedersen, S. G. Srinivasan, and H. Jónsson, Phys. Rev. Lett. **79**, 3676 (1997).
[14] R. Elber and M. Karplus, Chem. Phys. Lett. **139**, 375 (1987).
[15] R. Czerminski and R. Elber, Int. J. Quantum Chem. **24**, 167 (1990); R. Czerminski and R. Elber, J. Chem. Phys. **92**, 5580 (1990).
[16] R. E. Gillilan and K. R. Wilson, J. Chem. Phys. **97**, 1757 (1992).
[17] G. Henkelman and H. Jónsson, J. Chem. Phys. **111**, 7010 (1999).
[18] J. C. Polanyi and W. H. Wong, J. Chem. Phys. **51**, 1439 (1969).
[19] K. C. Pandey, Phys. Rev. Lett. **57**, 2287 (1986).
[20] J. Tersoff, Phys. Rev. B **39**, 5566 (1989).
[21] A. F. Voter and S. P. Chen, Mater. Res. Soc. Symp. Proc. **82**, 2384 (1987).
[22] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964); W. Kohn and L. J. Sham, *ibid*. **140**, A1133 (1965).
[23] W. Kohn, A. D. Becke, and R. G. Parr, J. Phys. Chem. **100**, 12974 (1996).
[24] G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993); G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 16 (1996); Phys. Rev. B **54**, 11169 (1996).
[25] J. P. Perdew, in *Electronic Structure of Solids*, edited by P. Ziesche and H. Eschrig (Akademie Verlag, Berlin, 1991).
[26] D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990).
[27] D. C. Seets, C. T. Reeves, B. A. Ferguson, M. C. Wheeler, and C. B. Mullins, J. Chem. Phys. **107**, 10229 (1997).
[28] G. Henkelman and H. Jónsson (in preparation).
[29] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computation*, 2nd ed. (Cambridge University Press, Cambridge, 1992), p. 420.