

## A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives

Graeme Henkelman and Hannes Jónsson

*Department of Chemistry, Box 351700, University of Washington, Seattle, Washington 98195-1700*

(Received 3 June 1999; accepted 20 July 1999)

The problem of determining which activated (and slow) transitions can occur from a given initial state at a finite temperature is addressed. In the harmonic approximation to transition state theory this problem reduces to finding the set of low lying saddle points at the boundary of the potential energy basin associated with the initial state, as well as the relevant vibrational frequencies. Also, when full transition state theory calculations are carried out, it can be useful to know the location of the saddle points on the potential energy surface. A method for finding saddle points without knowledge of the final state of the transition is described. The method only makes use of first derivatives of the potential energy and is, therefore, applicable in situations where second derivatives are too costly or too tedious to evaluate, for example, in plane wave based density functional theory calculations. It is also designed to scale efficiently with the dimensionality of the system and can be applied to very large systems when empirical or semiempirical methods are used to obtain the atomic forces. The method can be started from the potential minimum representing the initial state, or from an initial guess closer to the saddle point. An application to Al adatom diffusion on an Al(100) surface described by an embedded atom method potential is presented. A large number of saddle points were found for adatom diffusion and dimer/vacancy formation. A surprisingly low energy four atom exchange process was found as well as processes indicative of local hex reconstruction of the surface layer. © 1999 American Institute of Physics. [S0021-9606(99)70638-0]

### I. INTRODUCTION

Most atomic scale transitions in the condensed phase, such as chemical reactions and diffusion, are activated processes, i.e., require surmounting a significant energy barrier. While under typical conditions the thermal energy is on the order of  $k_B T = 0.025$  eV, the barriers for such transitions are typically on the order of 0.5 eV or higher. A transition that occurs thousands of times per second is so slow on the scale of atomic vibrations that it would typically take several thousands of years of computational time on the fastest modern computer to simulate a classical trajectory with reasonable chance of observing a single transition. As a result, classical dynamics simulations of activated transitions are faced with an impossible time scale problem. It is not possible to observe such transitions by simply simulating the classical dynamics of the system. Arbitrarily raising the temperature of the system can lead to crossover to a different transition mechanisms. The problem is to find a simulation algorithm that can be used to find which transition would occur and at what rate, if the classical dynamics could be simulated for a long enough time.

Within transition state theory (TST) the problem becomes that of finding the free energy barrier for the transition.<sup>1</sup> This is a very challenging problem, especially when the mechanism of the transition is unknown. Within the harmonic approximation to transition state theory (hTST)<sup>2,3</sup> the problem becomes that of finding the saddle point on the potential energy surface corresponding to a maximum along a minimum energy path that takes the sys-

tem from one potential energy minimum to another. This is still a difficult problem when dealing with condensed matter systems because of the high dimensionality of the potential energy surface. When both the initial and final states of the transition are known, minimum energy path(s) for the transition can be found quite readily<sup>4</sup> (the problem of making sure the path with lowest activation energy has been found is still a difficult problem). When only the initial state of the transition is known, the problem of finding the relevant saddle point(s) becomes that of navigating in high dimensional space—a very challenging task. If the set of all relevant, low lying saddle points for transitions from a given initial state could be found reliably and the prefactor in the hTST rate constant expression evaluated, then long time activated dynamics of the system could, in principle, be simulated. This may be impossible for all but the simplest systems.

Recently, significant progress has been made towards this goal. In the activation-relaxation technique developed by Barkema and Mousseau, the system is driven from one potential energy basin to another by inverting the component of the force acting on the system along a line drawn from the instantaneous configuration to the initial configuration<sup>5,6</sup> (or to a trailing image of the system<sup>7</sup>). The new potential energy basin is then accepted or rejected based on Monte Carlo sampling. This has enabled equilibration of supercooled liquids down to much lower temperature than could be achieved with direct classical dynamics simulations. In principle the method could be used to estimate the rate of the transitions

observed using harmonic transition state theory, but the algorithm does not always take the system through the close vicinity of the saddle point, making the estimate of the activation energy uncertain. Furthermore, there is no guarantee the activation-relaxation technique will give the same transition as a long classical trajectory, i.e., the transition with the lowest saddle point.

A very different approach to long time simulations has been developed by Voter, the so-called hyperdynamics method.<sup>8,9</sup> There, a classical trajectory is generated for a modified potential with a reduced well depth. The activated transitions are thus made more probable and may be observed during the short time interval accessible by classical dynamics simulations. The relative rate of different transition mechanisms is preserved in the modified system (within the transition state theory approximation) so the hyperdynamics trajectory should reveal the most probable activated transitions. It can, furthermore, provide an estimate of the transition rate within the full, anharmonic transition state theory approximation. In general, the hyperdynamics trajectory simulation still requires a large number of force evaluations, more than can be handled at the present time with *ab initio* methods, and the acceleration of the transitions becomes smaller as the system gets larger.

Powerful methods have been developed for climbing up potential energy surfaces from minima to saddle points in *ab initio* calculations of molecules. These methods have become part of the standard tool kit for molecular calculations, but they require the evaluation and inversion of the Hessian matrix at each point along the search so as to find the local normal modes of the potential energy surface. The strategy of following local normal modes to find saddle points was apparently first described by Crippen and Scheraga,<sup>10</sup> and later by Hilderbrandt.<sup>11</sup> In these early algorithms, a small step is taken up the potential along a particular mode, followed by a step towards lower potential energy along all other modes. In the early 1980's, these methods were replaced by quasi-Newton methods, in which the eigenvalues of the Hessian matrix are shifted to ensure that the potential is maximized along one chosen mode and minimized along all others. The shift parameters, or Lagrange multipliers, were introduced by Cerjan and Miller<sup>12</sup> and later modified by Simons *et al.*,<sup>13,14</sup> and by Wales.<sup>15</sup> A summary of the early developments is given in Ref. 16. These methods have been used extensively in *ab initio* calculations of molecules and empirical potential calculations of atomic and molecular clusters.<sup>17,18</sup> We will refer to these methods collectively as mode following methods. They are derived by expanding the potential in a local quadratic form, and selecting one of the local harmonic modes as the direction for the climb. If the softest mode is chosen, this is analogous to walking up the slowest ascent of a valley. This does not necessarily lead to a saddle point (as will be illustrated in an example below), so an important property of these methods is the ability to search for a saddle point along different orthogonal modes leading away from a given initial configuration. But, since the mode following methods require the evaluation and inversion of the second derivative Hessian matrix, they scale poorly with the number of degrees of freedom in the system.

Furthermore, second derivatives are only available at rather low levels of *ab initio* calculations, and are also not available in plane wave based density functional theory (DFT) calculations. The dimer method presented here captures the most important qualities of the mode following algorithms, while using only first derivatives of the potential energy.

An illustration of the importance of having a method that can be used to systematically identify saddle points leading from a given initial state, is the discovery by Feibelman in 1990 that an Al adatom does not diffuse on the Al(100) surface by repeated hops from one site to another, as had previously been assumed, but rather by a concerted exchange mechanism involving concerted displacement of two atoms.<sup>19</sup> This illustrates well how the preconceived notion of a transition mechanism can be incorrect even for a simple system. With the rapid increase in computational power and increased sophistication of simulation software, the complexity of simulated systems has increased greatly. For many systems that are actively being studied, even with *ab initio* methods, it is difficult and, in any case risky to simply guess what the transition mechanism is.

## II. THE DIMER METHOD

The method presented here for finding saddle points involves working with two images (or two different replicas) of the system. We will refer to this pair of images as the "dimer." If the system has  $n$  atoms, each one of the images is specified by  $3n$  coordinates. The two replicas have almost the same set of  $3n$  coordinates, but are displaced slightly by a fixed distance. The saddle point search algorithm involves moving the dimer uphill on the potential energy surface, from the vicinity of the potential energy minimum of the initial state up towards a saddle point. Along the way, the dimer is rotated in order to find the lowest curvature mode of the potential energy at the point where the dimer is located. The strategy of estimating the lowest curvature mode at a point without having to evaluate the Hessian matrix was presented by Voter in his hyperdynamics method. There, it is used to construct a repulsive bias potential so as to accelerate classical dynamics of activated processes.<sup>9</sup> In the method presented here, we use his dimer strategy to make the search for saddle points more efficient.

### A. Forces and energies

The dimer, depicted in Fig. 1, is a pair of images separated from their common midpoint  $\mathbf{R}$  by a distance  $\Delta R$ . The vector  $\hat{\mathbf{N}}$  which defines the dimer orientation is a unit vector pointing from one image at  $\mathbf{R}_2$  to the other image at  $\mathbf{R}_1$ . When a transition state search is launched from an initial configuration, with no prior knowledge of what  $\hat{\mathbf{N}}$  might be, a random unit vector is assigned to  $\hat{\mathbf{N}}$  and the corresponding dimer images are formed

$$\mathbf{R}_1 = \mathbf{R} + \Delta R \hat{\mathbf{N}}$$

and

$$\mathbf{R}_2 = \mathbf{R} - \Delta R \hat{\mathbf{N}}. \quad (1)$$

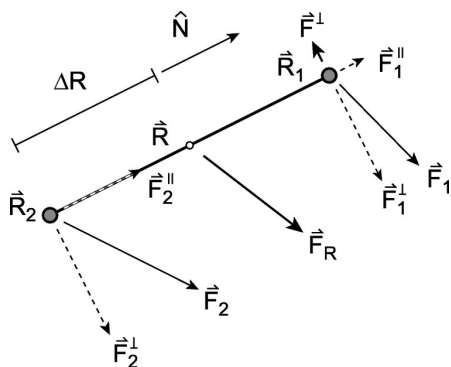


FIG. 1. Definition of the various position and force vectors of the dimer. The rotational force on the dimer,  $\mathbf{F}^\perp$ , is the net force acting on image 1 perpendicular to the direction of the dimer.

Initially, and whenever the dimer is moved to a new location, the forces acting on the dimer and the energy of the dimer are evaluated. These quantities are calculated from the energy and the force ( $E_1$ ,  $\mathbf{F}_1$ ,  $E_2$ , and  $\mathbf{F}_2$ ) acting on the two images. The energy of the dimer  $E = E_1 + E_2$  is the sum of the energy of the images. The energy and the force acting on the midpoint of the dimer are labeled as  $E_0$  and  $\mathbf{F}_R$  and are calculated by interpolating between the images. The force  $\mathbf{F}_R$  is simply the average force  $(\mathbf{F}_1 + \mathbf{F}_2)/2$ . The energy of the midpoint is estimated by using both the force and the energy of the two images. A relation for  $E_0$  can be derived from the finite difference formula for the curvature of the potential  $C$  along the dimer:

$$C = \frac{(\mathbf{F}_2 - \mathbf{F}_1) \cdot \hat{\mathbf{N}}}{2\Delta R} = \frac{E - 2E_0}{(\Delta R)^2}, \quad (2)$$

$E_0$  can be isolated from this expression in terms of the known forces on the images

$$E_0 = \frac{E}{2} + \frac{\Delta R}{4} (\mathbf{F}_1 - \mathbf{F}_2) \cdot \hat{\mathbf{N}}. \quad (3)$$

It should be emphasized that all the properties of the dimer are derived from the forces and energy of the two images. There is no need to evaluate energy and force at the midpoint between the two images. This is important for minimizing the total number of force evaluations required to find saddle points. An additional benefit of this strategy is that the method can be efficiently parallelized over two processors, the energy and force on each image being calculated on a separate processor. For *ab initio* calculations, in which force evaluations typically take a very long time compared with communication time, the execution time for each transition state search is effectively halved if two processors are used.

## B. Rotating the dimer

Each time the dimer is displaced, it is also rotated with a single iteration towards the minimum energy configuration. The practicality of the dimer method relies heavily on using an efficient algorithm for the rotation. Minimizing the dimer energy,  $E$ , is equivalent to finding the lowest curvature mode

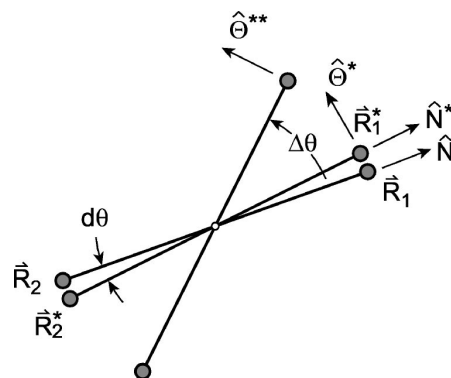


FIG. 2. Definition of the various quantities involved in rotating the dimer. All vectors are in the plane of rotation. The dimer is first rotated about a small angle  $d\theta$  to give a finite difference estimate of  $F'$  [given by Eq. (5)]. The dimer is then rotated by a calculated angle  $\Delta\theta$  [given by Eq. (13)] to zero the force within the plane of rotation.

at  $\mathbf{R}$ . The energy,  $E_0$ , at the fixed midpoint of the dimer is constant during the rotation. Since  $\Delta R$  is also constant, Eq. (2) shows that the dimer energy,  $E$ , is linearly related to the curvature,  $C$ , along the dimer. Therefore, the direction which minimizes  $E$  is along the minimum curvature mode at the midpoint  $\mathbf{R}$ .

## 1. Modified Newton method for rotation within a plane

The dimer rotation will first be discussed in the context of a modified Newton method. In the next section, the method will be extended to incorporate also a conjugate gradient approach. The dimer is rotated along the rotational force,  $\mathbf{F}^\perp = \mathbf{F}_1^\perp - \mathbf{F}_2^\perp$ , where  $\mathbf{F}_i^\perp \equiv \mathbf{F}_i - (\mathbf{F}_i \cdot \hat{\mathbf{N}}) \hat{\mathbf{N}}$  for  $i = 1, 2$ . The rotational force is taken to be the net force acting on image 1 (see Fig. 1). The rotation plane is spanned by  $\mathbf{F}^\perp$  and the dimer orientation  $\hat{\mathbf{N}}$ . It is useful to define a unit vector,  $\hat{\Theta}$ , within the plane of rotation, perpendicular to  $\hat{\mathbf{N}}$ . For the modified Newton method,  $\hat{\Theta}$  is just a unit vector parallel to  $\mathbf{F}^\perp$ . The vectors  $\hat{\Theta}$  and  $\hat{\mathbf{N}}$  form an orthonormal basis which spans the rotation plane. Given an angle of rotation,  $d\theta$ , image 1 moves from  $\mathbf{R}_1$  to  $\mathbf{R}_1^*$  (see Fig. 2)

$$\mathbf{R}_1^* = \mathbf{R} + (\hat{\mathbf{N}} \cos d\theta + \hat{\Theta} \sin d\theta) \Delta R. \quad (4)$$

After image 1 is moved to the new point  $\mathbf{R}_1^*$ , the new dimer orientation  $\hat{\mathbf{N}}^*$  is calculated, and image 2 is positioned at  $\mathbf{R}_2^*$  according to Eq. (1). The forces  $\mathbf{F}_1^*$ ,  $\mathbf{F}_2^*$ , and  $\mathbf{F}^* = \mathbf{F}_1^* - \mathbf{F}_2^*$  are then computed. A scalar rotational force  $F = \mathbf{F}^\perp \cdot \hat{\Theta} / \Delta R$  is used to describe the magnitude of the rotational force along the direction of rotation. Dividing by  $\Delta R$  scales the magnitude of  $F$  so that it is independent of the dimer separation. A finite difference approximation to the change in the rotational force,  $F$ , as the dimer rotates through the angle  $d\theta$  is given by

$$F' = \frac{dF}{d\theta} \approx \left| \frac{\mathbf{F}^* \cdot \hat{\Theta}^* - \mathbf{F} \cdot \hat{\Theta}}{d\theta} \right|_{\theta = d\theta/2}. \quad (5)$$

This approximation most accurately estimates the derivative for the midpoint of the finite rotation at  $\theta = d\theta/2$ .

A reasonable estimate of the rotation,  $\Delta\theta$ , required to bring  $F$  to zero can be obtained from Newton's method

$$\Delta\theta \approx \frac{(\mathbf{F} \cdot \hat{\Theta} + \mathbf{F}^* \cdot \hat{\Theta}^*)}{-2F'} \quad (6)$$

The rotation angles are illustrated in Fig. 2. Unfortunately, Newton's method systematically overestimates the angle  $\Delta\theta$  required to rotate the dimer to the minimum energy. An improvement on Eq. (6) can be made if the form of  $E(\theta)$ , the dimer energy as a function of rotation angle, is known. This can be accomplished in the following way. The first term in a Taylor expansion of the potential  $U$  in the neighborhood of  $\mathbf{R}$  is a hyperplane through the point  $U(\mathbf{R}) = E_0$ . This term alone produces no rotational force on the dimer because the dimer energy in this case is independent of orientation,  $E(\theta) = 2E_0$ . The quadratic term in the Taylor expansion introduces a rotational force. In order to write an analytic form of the quadratic approximation to the potential,  $\hat{x}$  and  $\hat{y}$  are defined to be the normal modes of the potential  $U$  within the plane of rotation. The plane is the two-dimensional subspace spanned by  $\hat{\Theta}$  and  $\hat{N}$ . Including terms up to second order in the Taylor expansion gives

$$U = E_0 - (F_x x + F_y y) + \frac{1}{2}(c_x x^2 + c_y y^2). \quad (7)$$

The forces  $F_x$  and  $F_y$  are  $-\partial U/\partial x$  and  $-\partial U/\partial y$  where  $x$  and  $y$  are distances along  $\hat{x}$  and  $\hat{y}$ . The curvature of the potential along  $\hat{x}$  and  $\hat{y}$  are  $c_x$  and  $c_y$ , respectively. The dimer energy  $E$  can be expressed within this quadratic approximation as a function of  $\theta$ :

$$E = 2E_0 + (\Delta R)^2 [c_x \cos^2(\theta - \theta_0) + c_y \sin^2(\theta - \theta_0)], \quad (8)$$

where  $\theta_0$  is some reference angle. As expected,  $F_x$  and  $F_y$  which define the linear change in  $U$  do not contribute to the energy of the dimer. Equation (8) can be rearranged using a trigonometric identity:

$$E = 2E_0 + \frac{1}{2}(\Delta R)^2 \{ (c_x - c_y) \cos[2(\theta - \theta_0)] + (c_x + c_y) \}. \quad (9)$$

The derivative of this potential yields an analytical expression for the scalar rotational force on the dimer

$$F = A \sin[2(\theta - \theta_0)]. \quad (10)$$

The constant  $A = (c_x - c_y)$  does not, in practice, have to be evaluated. The energy and the rotational force on the dimer are invariant to rotations of  $\pi$ . Equation (10) shows that  $\theta_0$  can be interpreted as the angle at which the force on the dimer is zero within the rotation plane. The difference  $\theta - \theta_0$  is the necessary angle of rotation required to reach a zero force. It is now possible to obtain an analytic form of the derivative  $F'$  defined in Eq. (5) within this harmonic approximation to the energy

$$F' = \frac{dF}{d\theta} = 2A \cos[2(\theta - \theta_0)]. \quad (11)$$

In a simulation,  $F$  and  $F'$  are evaluated at some orientation of the dimer. If this point is labeled as  $\theta=0$ , then the angle through which the dimer should be rotated to reach a

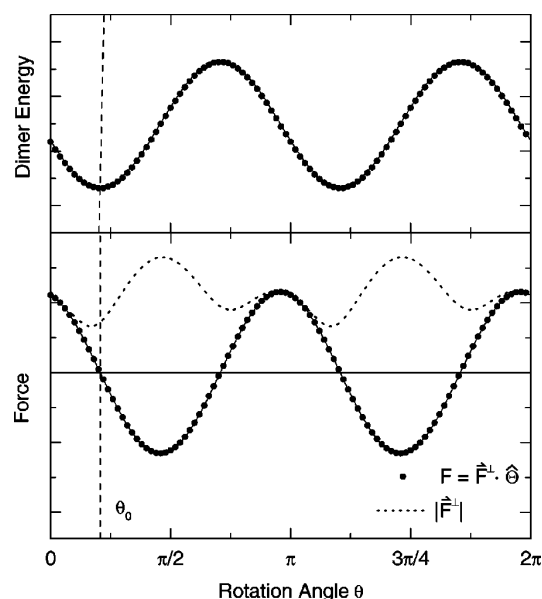


FIG. 3. Illustration of the modified Newton's method for orienting the dimer. The force and the energy of the dimer for an Al adatom on Al(100) are shown for a full rotation. The success of a sinusoidal fit to the dimer energy indicates that a quadratic approximation [Eq. (9)] is a good approximation. A fit [Eq. (10)] to the force acting on the dimer yields a minimum dimer energy within the plane at  $\theta_0 = 0.64937$ , in good agreement with that obtained from Eq. (13) using only  $F$  and  $F'$  calculated at  $\theta=0$ . The dashed line shows the magnitude of the total rotational force. At  $\theta=\theta_0$  this force has no component in the plane of rotation.

zero in the force is  $\theta_0$ . Let  $F_0$  and  $F'_0$  be the values of  $F$  and  $F'$  evaluated at  $\theta=0$ . Then the ratio of Eqs. (10) and (11):

$$\frac{F_0}{F'_0} = \frac{1}{2} \tan(-2\theta_0), \quad (12)$$

yields a simple expression in which the desired rotation angle can be isolated in terms of known quantities

$$\Delta\theta = \theta_0 = -\frac{1}{2} \arctan\left(\frac{2F_0}{F'_0}\right). \quad (13)$$

Equation (13) has better behavior than Eq. (6) for rotation in the limits of  $F \rightarrow 0$  and  $F' \rightarrow 0$ . This approach to the minimization is similar to a method used to minimize the electronic degrees of freedom in plane wave based density functional theory (DFT) calculations.<sup>20</sup>

As an example, we will discuss an application of the modified Newton method to a system representing an aluminum adatom on an Al(100) surface. Here we focus on the properties of the rotation only. The results of the saddle point searches for this system are presented in Sec. III B. A dimer is placed on the potential surface and incrementally rotated through an angle of  $2\pi$ . Figure 3 shows the force and energy of the dimer. The energy shows the expected sinusoidal behavior in the local quadratic approximation to the potential. The period of  $\pi$  is due to the symmetry of the dimer. The sinusoidal curve evaluated using Eq. (9) agrees well with the energy data. The force is also well represented by the sinusoidal Eq. (10). This is to be expected because  $F$  is simply



proportional to the derivative of the dimer energy with respect to  $\theta$ . A fit to these data (shown in Fig. 3) gives a value of  $\theta_0 \approx 0.64937$ . This value, indicated by the vertical dashed line, corresponds well with the minimum in the dimer energy within the rotation plane. In a simulation,  $\Delta\theta$  is determined from Eq. (13). For this example,  $F_0$  was found to be 4.453 12, and  $F'_0$  was  $-2.4847$ , from which  $\Delta\theta$  was calculated as 0.649 36, in good agreement with the observed value.

In summary, the quadratic approximation to the potential provides a formula for rotating the dimer and zeroing the rotational force within the plane. This is done by evaluating the magnitude of the rotational force  $F$ , the curvature of the dimer energy  $F'$ , and evaluating  $\Delta\theta$  by substituting into Eq. (13). Figure 3 shows the magnitude of the total rotational force. Once the dimer is rotated by  $\Delta\theta$ , the rotational force has essentially no component within the plane of rotation. It is a bit disconcerting to see that the magnitude of the total force drops by only 35% in the first iteration. This ratio is typical of the modified Newton method. This can be improved upon by considering conjugate gradients.

## 2. Conjugate gradient choice of rotation plane

Conjugate gradient methods tend to be more efficient than steepest descent methods because the force at both the current iteration and the previous iteration are used to determine an optimal direction of minimization.<sup>21</sup> We use a conjugate gradient algorithm for choosing the plane of rotation, while the minimization of the force on the dimer within a plane is carried out using the modified Newton method described in the previous section. The conjugate gradient method as described in Ref. 21 cannot be applied directly to the dimer energy minimization. For rotation, the dimer's midpoint and separation must be held fixed, which adds an additional constraint on the system. In this section, the traditional conjugate gradient method is first reviewed, and then a modification for the constrained dimer rotation is described.

The first step in the conjugate gradient minimization is a steepest descent step in which the direction of displacement is given by the gradient, (or force),  $\mathbf{F}$ . The energy along  $\mathbf{F}$  is then minimized. For subsequent iterations, the direction of displacements,  $\mathbf{G}$  is taken to be a linear combination of the current force,  $\mathbf{F}_i$ , and the force at the previous iteration  $\mathbf{F}_{i-1}$ . The vector  $\mathbf{G}$  at iteration  $i$  is defined recursively:

$$\mathbf{G}_i = \mathbf{F}_i + \gamma_i \mathbf{G}_{i-1}, \quad (14)$$

where  $\gamma_i$  is the weighting factor

$$\gamma_i = \frac{(\mathbf{F}_i - \mathbf{F}_{i-1}) \cdot \mathbf{F}_i}{\mathbf{F}_i \cdot \mathbf{F}_i}. \quad (15)$$

In the dimer method, the traditional conjugate gradient method is modified in several ways to accommodate the constraints implicit in the dimer rotation. Each minimization direction,  $\mathbf{G}$ , becomes a plane of rotation spanned by the unit vector  $\hat{\Theta}$  which is parallel to  $\mathbf{G}^\perp$ , and the dimer orientation  $\hat{\mathbf{N}}$ . The line minimization step is implemented with the modified Newton's method of the previous section. The difference is that, for every step other than the first,  $\hat{\Theta}$  is not along the force  $\mathbf{F}^\perp$  as it was, but rather along the conjugate vector  $\mathbf{G}^\perp$ .

Equations analogous to Eqs. (14) and (15) are used in the conjugate gradient rotation scheme. For the first iteration  $\mathbf{G}^\perp = \mathbf{F}^\perp$ . For the  $i$ th iteration

$$\mathbf{G}_i^\perp = \mathbf{F}_i^\perp + \gamma_i |\mathbf{G}_{i-1}^\perp| \hat{\Theta}_{i-1}^{**} \quad (16)$$

where

$$\gamma_i = \frac{(\mathbf{F}_i^\perp - \mathbf{F}_{i-1}^\perp) \cdot \mathbf{F}_i^\perp}{\mathbf{F}_i^\perp \cdot \mathbf{F}_i^\perp}, \quad (17)$$

is the weighting factor between the rotational force  $\mathbf{F}_i^\perp$  and the old modified force vector  $\mathbf{G}_{i-1}^\perp$ . The vector  $\mathbf{G}_{i-1}$  in Eq. (14) has been replaced by the vector  $|\mathbf{G}_{i-1}^\perp| \hat{\Theta}_{i-1}^{**}$  (Fig. 2 shows how  $\hat{\Theta}^{**}$  is found). This difference is due to the fact that  $\mathbf{G}_{i-1}^\perp$  was aligned along  $\hat{\Theta}_{i-1}$ . As described in the previous paragraph, a vector which is perpendicular to  $\hat{\mathbf{N}}_i$  within the old rotation plane is needed. This is simply a vector along  $\hat{\Theta}_{i-1}^{**}$ , with a magnitude equal to  $\mathbf{G}_{i-1}^\perp$ . The conjugate gradient approach [Eqs. (16) and (17)], including the modified Newton algorithm for line minimization [Eq. (13)], represents a significant improvement over a straightforward application of the standard conjugate gradient constraint for the dimer. The average number of force calls required to find a saddle point in the Al/Al(100) study was reduced by a factor of about 2.5.

## C. Translating the dimer

Compared to rotation, the translation of the dimer is relatively straightforward. The saddle point is a maximum along the lowest curvature mode, the reaction coordinate, and a minimum along all other modes. The dimer will orient itself along the lowest curvature mode when the energy of the dimer is minimized by rotation. The net translational force acting on the two images in the dimer,  $\mathbf{F}_R$ , tends to pull the dimer towards a minimum, however. Therefore, a modified force,  $\mathbf{F}^\dagger$ , is defined where the force component along the dimer is inverted:

$$\mathbf{F}^\dagger = \mathbf{F}_R - 2\mathbf{F}^\parallel. \quad (18)$$

Movement of the dimer along this modified force will bring it to a saddle point. This is illustrated in Fig. 4. In principle, any optimization algorithm depending only on first derivatives can be used to move the dimer along the effective force to the saddle point.

We have used two different algorithms to translate the dimer. The first is similar to an algorithm that has been used by others to find potential energy minima.<sup>22</sup> We will refer to this algorithm as the "quick-min" algorithm. A time step size,  $\Delta t$ , is selected. This should be as large as possible, while still allowing the system to reach the convergence criteria for the saddle point. The system is propagated from its initial position using a classical dynamics algorithm, with the modification that only the projection of the velocity at the previous step along the current force is kept. Additionally, if the dot product of the force and the velocity becomes negative, the cumulative velocity is set to zero

$$\Delta \mathbf{V}_i = \mathbf{F}_i^\dagger \Delta t / m \quad (19)$$

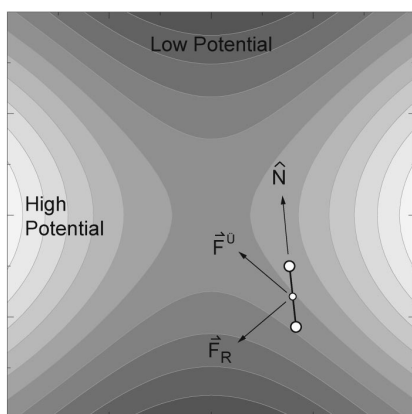


FIG. 4. The effective force  $\mathbf{F}^\dagger$  acting on the center of the dimer is the true force  $\mathbf{F}_R$  with the component along the lowest curvature mode  $\hat{\mathbf{N}}$  inverted. In the neighborhood of a saddle point, the effective force points towards the saddle point.

$$\mathbf{V}_i = \begin{cases} \Delta \mathbf{V}_i (1 + \Delta \mathbf{V}_i \cdot \mathbf{V}_{i-1} / \Delta \mathbf{V}_i^2) & \text{if } \mathbf{V}_i \cdot \mathbf{F}_i^\dagger > 0 \\ \Delta \mathbf{V}_i & \text{if } \mathbf{V}_i \cdot \mathbf{F}_i^\dagger < 0 \end{cases} \quad (20)$$

There is a problem with the algorithm described thus far when the dimer is started from a shallow minimum. If the lowest curvature mode is along a contour of the potential energy basin, the dimer can take a very long time to leave the basin, or even possibly become trapped there forever. A solution to this problem is to treat regions where all modes have positive curvature, the *convex regions*, differently from the regions where at least one mode has a negative curvature, the *nonconvex regions*. The neighborhood of potential minima falls into the first category while the saddle point region falls into the second category. Equation (18) is modified in the following way to ensure that the dimer quickly leaves convex regions:

$$\mathbf{F}^\dagger = \begin{cases} -\mathbf{F}^\parallel & \text{if } C > 0 \\ \mathbf{F}_R - 2\mathbf{F}^\parallel & \text{if } C < 0 \end{cases} \quad (21)$$

where  $C$  is the minimum curvature. In the convex regions  $C > 0$ , and the dimer follows this mode up the potential surface until the lowest curvature becomes negative. It is possible that  $C$  never becomes negative (an example of that in a two-dimensional case will be given below), in which case the dimer continues to climb up the potential forever. This problem is unlikely to occur in large atomic systems. We never encountered it in the Al/Al(100) calculations described below.

The second method we tried for translating the dimer was the conjugate gradient method. This was found to perform better than quick-min in the Al/Al(100) calculations. In the initial step, the system is minimized along a line defined by the initial force. Analogous to the rotation algorithm, the system is moved a small distance along the line (keeping the dimer orientation fixed), and the derivative in the magnitude of the effective forces was calculated. Newton's method is used to estimate the zero in the effective force along the line and the dimer is moved to that point. If the effective force in the line increases in the small step, the dimer is still in the

minimum region, and Newton's method calculates a step backwards against the effective force, pulling the dimer back into the minimum. In this case, the dimer is simply moved with the effective force a predefined step size. This algorithm tends to move the dimer out of the convex region quickly and in practice speeds up convergence to a saddle point. After each translation, the dimer is reoriented and then moved along a direction conjugate to the previous line minimization.<sup>21</sup>

#### D. Selecting initial configurations

In most systems, there can be a large number of saddle points leading out of the potential energy basin of interest. A single saddle point search will generally not be enough to address the question of how the system tends to leave the basin. In general, it is necessary to know *all* low lying saddle points (to within a few  $k_B T$  from the lowest energy saddle point) leading from a potential energy basin. While no existing method can guarantee that all relevant saddle points will be found, reasonable progress may be made if there is a way to search for new saddle points in a manner that minimizes the number of duplications.

One simple approach is to start with a collection of initial configurations, scattered about the potential energy minimum in the initial basin. In order to avoid high energy configurations, which might be spatially near the potential minimum, a system can be evolved by classical dynamics at some finite temperature and configurations of the atoms corresponding to maximal displacements from the potential energy minimum can be saved as initial configurations for saddle point searches. In other words, different saddle points can be found if the initial configurations are drawn from the high potential energy images within a thermal ensemble in the potential energy basin. This approach turned out to be quite successful. But, it is important to realize that some saddle points can be systematically excluded when only this method is used. The configurations generated tend to be along low energy modes around the minimum and the dimer searches from these configurations tend to converge to the same saddle points, the saddle point lying at the end of a low curvature mode. These are, however, often the lowest energy saddle points. Starting with a random set of images displaced from the minimum, for example, a Gaussian distribution of displacements amounting to  $\approx 0.1$  Å in each coordinate, gave a greater variety of saddle points and therefore better sampling. The following section describes how different saddle points can be found when starting from the same initial configuration.

#### E. Orthogonalization

The various mode following algorithms can converge to a variety of saddle points starting from the same initial configuration by following different normal modes.<sup>12-14,16</sup> There is, however, no inherent relationship between the number of normal modes and the number of saddle points. The important aspect of the mode following algorithms is the orthogonality of the modes, which tends to lead the system in different directions towards different saddle points. This kind of

orthogonality constraint can be built into the dimer method quite easily and with little increase in computational cost.

From any initial configuration, the lowest curvature mode can be found with the dimer rotation method described in Sec. II B. When the dimer is rotated to its minimum energy configuration, the dimer orientation is along the lowest curvature mode  $\hat{\mathbf{N}}_1$ . Let the curvature along this mode be denoted by  $C_1$ . The first saddle point search will typically be launched with the dimer initially oriented along this mode. It is also possible to find the next lowest mode by again rotating a dimer to its minimum energy configuration but now maintaining orthogonality to the vector  $\hat{\mathbf{N}}_1$ . Orthogonalization is carried out by simply subtracting any component along  $\hat{\mathbf{N}}_1$  from the vectors  $\hat{\mathbf{N}}$ ,  $\mathbf{F}_1$ , and  $\mathbf{F}_2$ , while rotating the dimer to a minimum energy. A new saddle point search can then be launched from the same initial configuration with the dimer initially oriented along the second lowest mode  $\hat{\mathbf{N}}_2$ . In this second search, the orthogonality condition between the current orientation of the dimer  $\hat{\mathbf{N}}$  and the initial orientation in the first search  $\hat{\mathbf{N}}_1$  is maintained until the curvature  $C$  along the dimer becomes lower than the curvature measured along the direction  $\hat{\mathbf{N}}_1$ . This is not the same as requiring that the curvature along  $\hat{\mathbf{N}}$  be lower than  $C_1$ , the initial curvature along  $\hat{\mathbf{N}}_1$ . The requirement is that  $\hat{\mathbf{N}}$  gives the lowest curvature mode at the point on the potential energy surface when the orthogonality constraint is dropped. When  $\hat{\mathbf{N}}$  gives the lowest curvature mode, there is no need to maintain the orthogonality condition to the vector  $\hat{\mathbf{N}}_1$  because the dimer has no tendency to rotate into this now higher curvature direction.

It is straightforward to continue this procedure to follow systematically different directions. As long as the curvature along the current lowest mode  $\hat{\mathbf{N}}$  is greater than the curvature in one of the initial directions of earlier searches, then the orthogonality condition is maintained. These different saddle point searches can be carried out in parallel after the initial set of low lying modes ( $\hat{\mathbf{N}}_1, \hat{\mathbf{N}}_2, \dots$ ) have been found. The cost of these subsequent searches increases somewhat, because as long as the dimer does not lie along the lowest curvature mode, the force and dimer energy must be computed for every initially lower mode. Two things save this potentially poor scaling. First, there is no need to compare the curvatures very often, and second, it has turned out that the dimer tends to escape the region where previous modes have lower curvature quite quickly. The combination of a distribution of initial configurations and orthogonalization provides an efficient, highly parallelizable, method for searching for saddle points leading out of a given potential energy basin.

### III. RESULTS

The characteristics of the dimer method have been studied using two model potentials. The first is a two-dimensional model potential. The second is a system representing an Al adatom on an Al(100) surface—a system containing 301 atoms.

TABLE I. Parameters for the Gaussian functions added to the two-dimensional test potential.

$i$	1	2
$A_i$	1.5	6.0
$x_{0_i}$	2.020 83	0.8
$y_{0_i}$	−0.172 881	2.0
$\sigma_{x_i}$	0.1	5.0
$\sigma_{y_i}$	0.35	0.7

#### A. Two-dimensional test system

Simple potential surfaces which can be visualized easily are important test cases for any saddle point search method. We have chosen a LEPS potential coupled to a harmonic oscillator because it illuminates some of the problems that can be encountered in saddle point searches. The analytic form and justification for the potential are described elsewhere.<sup>4</sup> The slowest ascent direction along  $-\hat{\mathbf{y}}$  from the initial basin centered at (0.7655, 0.2490) does not lead to a saddle point if a steepest ascent search is carried out along this mode. Without modification, the potential has a single saddle point between the two large basins defined approximately by  $x < 2$  and  $x > 2$ . We have added two Gaussian functions to this potential to increase the number of saddle points leading out of the initial basin from one to four

$$G_i(x, y) = A_i e^{-(x-x_{0_i})^2/2\sigma_{x_i}} e^{-(y-y_{0_i})^2/2\sigma_{y_i}}. \quad (22)$$

The parameters of the Gaussian functions are given in Table I. Figure 5 illustrates how the dimer method works on this potential surface.

A classical trajectory calculation was used to generate three starting configurations for the saddle point searches in Fig. 5(a). Dimers are placed at these initial points (the dimer separation is much too small to be resolved), and the subsequent saddle point searches following the lowest mode and using the quick-min algorithm are shown. There is a fairly sharp change in the two paths initially following the nearly vertical directions. These are the points at which the curvature along the dimer has switched from positive to negative (the boundary of convex and nonconvex regions), as dictated by Eq. (21). To begin with, in the convex region, the dimer is only moved up the potential surface along the lowest curvature orientation. After entering the nonconvex region, the dimer is also being moved down along the force perpendicular to the dimer orientation.

Figure 5(b) shows the results of saddle point searches using the orthogonalization algorithm. Two initial configurations were used and for each one the two lowest normal modes were used in saddle point searches. The two searches along  $+\hat{\mathbf{x}}$  and  $+\hat{\mathbf{y}}$  quickly converge to saddle points. The search following the softest mode along  $-\hat{\mathbf{y}}$  goes down to about  $y = -10$  before locking onto the negative curvature mode along  $\hat{\mathbf{x}}$  and rotating by  $90^\circ$ . The minimization perpendicular to the dimer brings it back towards the saddle point. The second search from this same initial point brings the dimer in the  $-\hat{\mathbf{x}}$  direction, where no saddle point exists. It is convenient to specify both a maximum potential energy, and

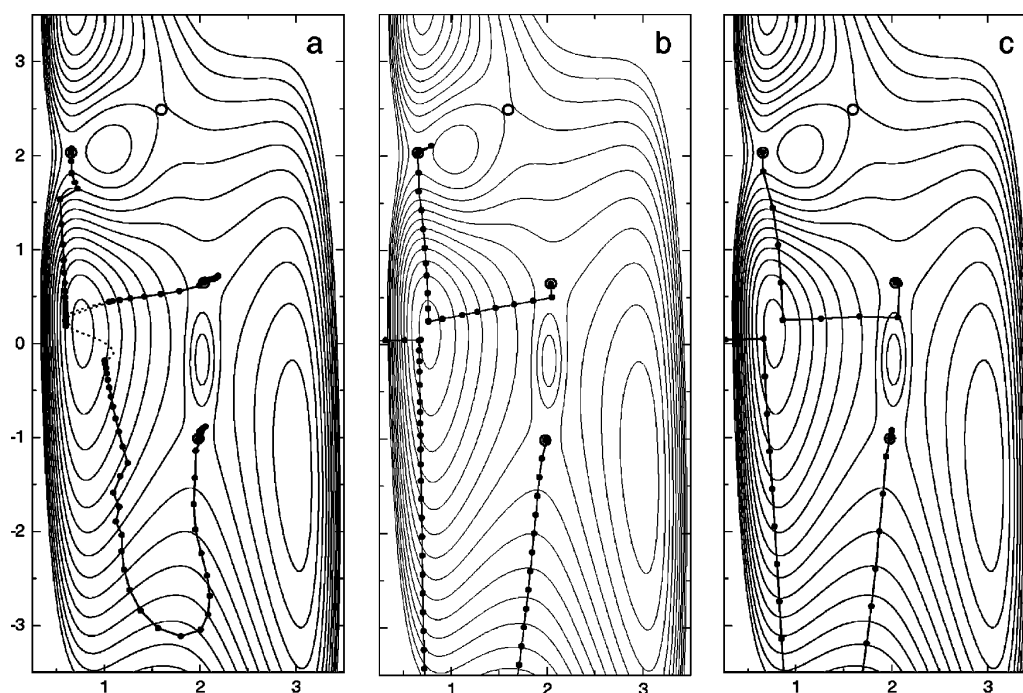


FIG. 5. Two algorithms for moving the dimer are compared, along with a mode following algorithm (Ref. 16). The left-hand side figure shows a short molecular dynamics trajectory (dashed line), and three initial configurations generated from the trajectory. The quick-min algorithm was used to move the dimer along the lowest curvature direction. In the middle figure, two initial configurations were chosen on opposite sides of the minimum. For each configuration, two dimer search calculations were carried out, in one case following the lowest curvature direction, in the other following the next lowest curvature. The dimer moving off the plot to the right-hand side did not converge to a saddle point. The dimer which moved off the bottom of the plot returned after moving to  $y = -10$  and finally did converge to a saddle point. In this calculation, the conjugate gradient method was used to move the dimer. The right-hand side figure shows results of a mode following method requiring the diagonalization of the Hessian matrix. Qualitatively, the results of the dimer method with orthogonalization are very similar to the results of the mode following method. This comparison is only feasible in a few dimensions because the mode following algorithm requires inversion of the Hessian matrix. None of the methods located one of the saddle points (in the upper right-hand side corner of the basin).

a maximum number of allowed iterations so that a failed search, such as this one, can be aborted without wasting an unreasonable amount of effort.

Figure 5(c) shows the results of a mode following method<sup>16</sup> which relies on computing the Hessian matrix. It performs very efficiently on this test potential, finding three modes in an average of 16 moves each. In two dimensions, the cost of each move is small, but as the number of dimensions increase the cost of evaluating and inverting the Hessian matrix increases rapidly (as  $n^3$ ). Even if the number of steps remains small, the total cost becomes prohibitive in large systems. It is reassuring to see that the orthogonalized dimer searches mimic the attractive features of the mode following algorithm without having to evaluate the Hessian matrix.

It is instructive to identify which initial points lead the dimer to a given saddle point. Figure 6 shows the basins of attraction for the various saddle points of the two-dimensional test problem when only the lowest mode is followed. A test dimer was placed at an array of initial points. The shaded areas are the regions of the potential in which the dimer rotated into a negative curvature mode, indicating that there is at least one negative curvature mode at that point. Each dimer was then moved in the direction of the effective force to a saddle point. The regions are shaded according to

which saddle point the dimer moved to. In this way, the negative curvature region is further divided up into four basins of attraction, one for each saddle point. If the dimer is started outside of the shaded region, it can still converge to a saddle point as shown in Fig. 5, by moving up the potential along the lowest (positive) curvature mode until it reaches a basin of attraction. Once the dimer reaches a basin of attraction of a saddle point, it is most likely to converge to that saddle point. Therefore, images which start around the left minimum will not reach saddle point 3.

It is fairly clear then why none of the images started around the left minimum reached saddle point 3. They would have to pass through basins 2 or 4 before getting to the basin around saddle point 3. This illustrates a limitation of the dimer method. If a saddle point has a basin which is surrounded or separated from the initial configuration by other saddle point basins, it is very unlikely for the dimer method to find it. Another limitation of the method can be seen by considering saddle point 3. This saddle point connects the two minima on the left by a rather curved minimum energy path. Dimers that are started from the minimum on the right following the  $+\hat{y}$  mode converge to saddle point 3. But this is not a saddle point leading out of that minimum. This is a problem when the goal is to find all saddle points leading out of a given minimum. It is easy enough to check if a given



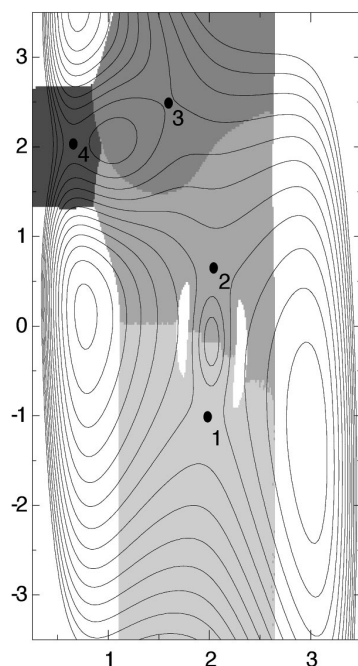


FIG. 6. Regions of attraction around each saddle point. The shaded regions correspond to points with at least one negative curvature mode. The different shades of gray indicate which saddle point the dimer will converge to. A limitation of the method is apparent here. A dimer starting from the initial basin on the left-hand side will not be able to find saddle point 3, which represents one of the escape routes from the basin, without first visiting the basin around saddle point 2 or 4.

saddle point is relevant for a given initial state, but the cost of finding it can be a wasted effort. This effect is observed in the aluminum system as well as this two-dimensional potential.

## B. Al adatom on an Al(100) surface

We now turn to a study of transition mechanisms in a system where an Al adatom is initially sitting in the fourfold hollow site on Al(100) surface. The goal is to find all mechanisms by which the system can escape from this initial state with an activation energy less than about 1 eV. An embedded atom potential, similar to that of Voter and Chen,<sup>23</sup> was used to model the atomic interactions. The energy of the two lowest saddle points predicted by this potential turn out to be close to the results of density functional theory calculations,<sup>19</sup> indicating that the potential function is reasonably accurate. The substrate consists of 300 atoms, 50 per layer in six layers. The bottom two layers are held frozen and the top surface is left open to vacuum. A single Al adatom is placed on the surface bringing the total number of degrees of freedom to 603.

The dimer method was run starting from 1000 randomly chosen configurations around the minimum. A cluster including the adatom and 25 nearby substrate atoms was displaced according to a Gaussian probability distribution with a width of 0.1 Å along each coordinate. The conjugate gradient method was used both to rotate and translate the dimer. Values of the finite difference steps for rotation and transla-

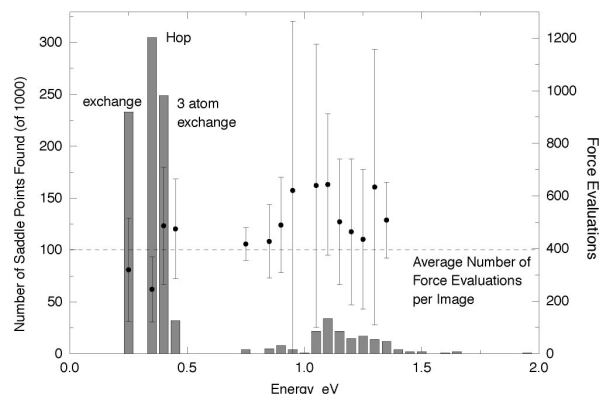


FIG. 7. The result of 1000 saddle point searches using the dimer method for an Al adatom on an Al(100) surface. Each search starts from a point randomly displaced from the potential energy minimum, using a Gaussian distribution with a width of 0.1 Å in each coordinate of 25 atoms including the adatom and its neighbors. A large number of saddle points was found. The histogram shows how many of the dimers converged on saddle points in the given energy range. The processes corresponding to the ten lowest energy saddle points are shown in Fig. 8. A total of 60 processes below 2.0 eV were found, most of them in the range of 0.8–1.5 eV. The filled circle shows the average number of force evaluations required to converge to the saddle points within each bin. The bars show the range between the shortest and longest search. On average 400 force evaluations (200 per image in the dimer) were needed to converge to a saddle point (dashed line).

tion were  $d\theta = 10^{-4}$  rad and  $dR = 10^{-3}$  Å, respectively. A maximum translation distance was set at 0.1 Å, and the dimer separation was set at  $\Delta R = 10^{-2}$  Å. The tolerance for convergence to a saddle point was set at  $|F| < 10^{-4}$  where  $\mathbf{F}$  is the  $3n$  dimensional force vector.

A summary of the results is shown in Fig. 7. Of the 1000 dimer searches, 990 converged to saddle points below 2 eV. Three searches failed to converge within the imposed limit of 2000 iterations. Seven of the searches converged to the long wavelength, high energy mode (with activation energy of  $\approx 5$  eV) corresponding to a concerted shift of the 51 surface atoms by one lattice constant. The energy of the saddle points were binned and the number of searches within each bin is shown in the histogram in Fig. 7. The average computational cost to find the saddle points within each bin is given in terms of the number of force evaluations. The average number of force evaluations needed to converge on a saddle point was 400, that is 200 per image in the dimer. The three lowest energy saddle points attracted 78% of the 1000 dimers (saddle points that are equivalent by symmetry are grouped together). These are the exchange process involving two atoms found by Feibelman,<sup>19</sup> the hop, and a remarkably low *four* atom exchange process (see Fig. 8). A three atom exchange process has the fourth lowest saddle point. No transition mechanisms were found with saddle point energy in the range 0.44–0.76 eV, but above this energy gap there is a large number of different processes with saddle point energy up to about 1.5 eV. Figure 8 shows the ten lowest energy transitions. The initial state, saddle point, and final state are shown. The energies listed below each transition number are the energy of the saddle point configuration with respect to a configuration of an adatom on a flat surface.

Once the dimer has converged to a saddle point, it is

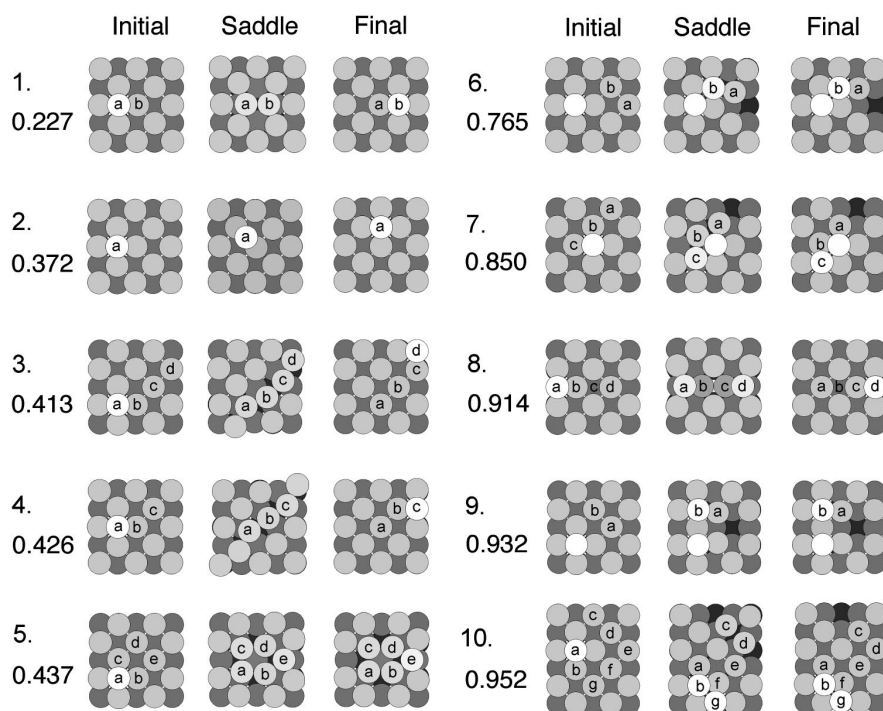


FIG. 8. The ten lowest energy transition mechanisms found in 1000 searches using the dimer method. An on-top view of a small section of the surface is shown in the initial state (left-hand side), saddle point (middle), and final state (right-hand side). The atoms are shaded by depth, and the atoms that move the most are labeled. The energy of the saddle point configurations is given in eV with respect to an Al adatom in the fourfold hollow site on the flat Al(100) surface. In addition to the two atom exchange process (process 1, discovered by Feibelman) and the hop (process 2), a four atom exchange mechanism (process 3) and a three atom exchange mechanism (process 4) are found to be low energy diffusion mechanisms. Other processes correspond to an adatom getting buried in the surface (process 5), vacancy formation (processes 6, 7, 9, and 10), and a four atom exchange diffusion mechanism involving a second layer atom (process 8).

easy to trace out the minimum energy path. The dimer is first allowed to rotate into the lowest curvature mode, the unstable mode, so that it is aligned along the reaction coordinate. An image is placed on one side of the dimer along the direction  $\hat{N}_1$ . The distance of the image from the midpoint of the dimer (the saddle point) is chosen according to the desired resolution of the path. In a manner very similar to the algorithm used to rotate the dimer, the energy of this image is minimized while keeping its distance from the previous image (in this case the saddle point) fixed. This procedure is repeated, each time placing a new image initially along the local path (the line between the two previous images) to minimize the number of function calls required to zero the tangential force on the new image. The process is stopped when the minimum energy of an image is greater than that of the previous image. After the path is traced out in one direction from the saddle point to a minimum, the opposite direction must be followed to complete the minimum energy path. This method was used for the ten saddle points of Fig. 8 and the energy paths are shown in Fig. 9. The reaction coordinate is scaled so that the distance between the minimum and the saddle point is 1 unit. The paths which do not terminate back at an energy of zero, the energy of a single Al adatom on the Al(100) surface, lead to other local minima on the potential energy surface. The final state of path 5 corresponds to a stable arrangement with the adatom buried in the surface. A group of four atoms in the surface layer has rotated by  $45^\circ$ . The final states of paths 6, 7, and 10 at approximately 0.7 eV are arrangements in which an addimer/vacancy pair has been created. Path 8 corresponds to a four atom exchange diffusion mechanism involving a second layer atom.

Several orthogonal dimer searches were then carried out from the 1000 initial configurations described above. For

each initial configuration, a total of eight orthogonal dimer searches were carried out. In each successive run, the dimer orientation is orthogonalized to the initial orientation of the dimer in the previous runs. After each search, the saddle point obtained was compared to those found in previous searches started from the same initial configuration. In this way, an average chance of finding a new saddle point in a subsequent, orthogonal search was estimated. The results are shown in Fig. 10. On average, the second search from a given initial configuration led to a new saddle point 60% of the time. A new saddle point was found in the third search

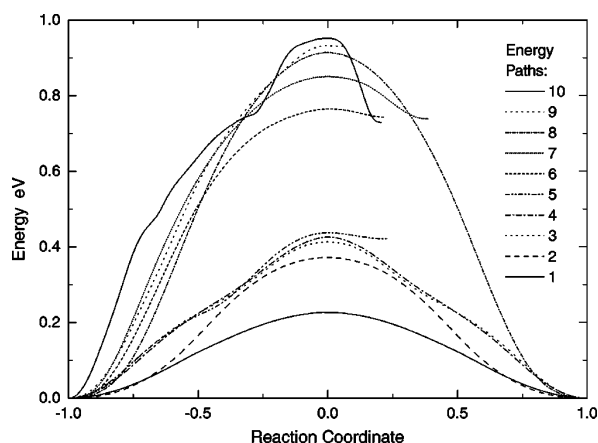


FIG. 9. The minimum energy paths corresponding to the ten saddle points identified and shown in Fig. 8. The reaction coordinate has been scaled so that  $-1$  represents the initial minimum and  $0$  the saddle point. Transitions 5, 6, 7, 9, and 10 lead to final state local minima which do not correspond to a single adatom on the Al(100) surface and are therefore asymmetric.

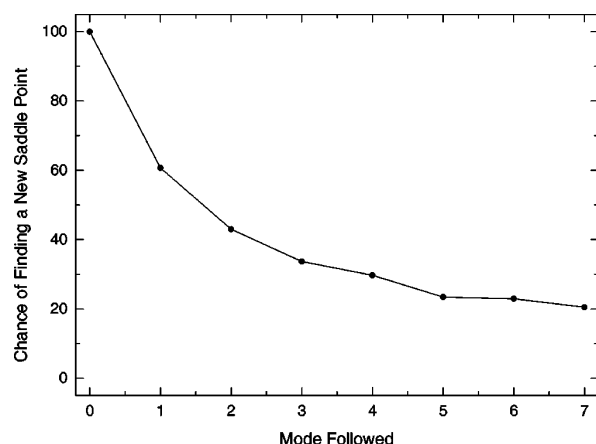


FIG. 10. For each one of the 1000 random initial configurations, a total of eight dimer searches were carried out with each subsequent search orthogonalized to the initial orientation of the dimer in previous searches. The saddle point obtained in each run was compared to the saddle points found in previous searches started from the same initial configuration. The figure shows the average fraction of searches which lead to a new saddle point. For example, after a saddle point was found by following the lowest curvature direction, the chance of finding a new saddle point when a second lowest, orthogonal direction was followed is approximately 60%. The chance of finding a new saddle point when the third lowest mode is subsequently followed is approximately 40%.

about 40% of the time. During the course of these simulations, many new transition mechanisms for the Al/Al(100) system were found. Some of the more interesting low energy saddle points are shown in Fig. 11. A transition involving the formation of a local hex reconstruction in the surface layer at the saddle point was found. In the final state of the transition, a group of four surface atoms has rotated so as to exchange places (process 1). A similar rotation of four surface atoms also occurs in the second transition shown in Fig. 11. An addimer/vacancy pair forms in several of the transitions (pro-

cesses 3, 4, and 7). For some of the transitions, neither the initial nor the final state corresponds to an adatom on a flat surface (processes 5, 6, and 8), illustrating that the dimer does not always converge on saddle points corresponding to escape routes for the potential energy basin where the initial point of the search is located.

### C. Scaling with system size

The motivation behind the dimer method is to develop an algorithm which scales well with system size. Most of the savings over the mode following methods, such as the Cerjan-Miller method, is a result of not having to evaluate and invert the Hessian matrix. The dominant computational effort in the dimer method involves computation of the force on the atoms, so the effort can effectively be measured by the number of force evaluations required to converge on a saddle point. The more degrees of freedom there are in the system, the more iterations are needed to orient and translate the dimer. An important question is how the computational effort scales with the number of degrees of freedom. The maximum number of degrees of freedom was obtained in the Al/Al(100) system by freezing only 55 atoms at the bottom of the substrate leaving the method to explore the remaining 738 degrees of freedom. In the other extreme, only the adatom was allowed to move, making 3 the minimum number of degrees of freedom. The average cost of finding an ensemble of saddle points was computed for eight configurations with more and more of the substrate atoms frozen. For the most restricted systems with fewer than 70 degrees of freedom, the hop was the only process found. The dimer method converged with about 70 force evaluations in these simulations. For simulations with fewer frozen atoms the full range of saddle points was found. This distribution was fairly insensitive to the number of degrees of freedom beyond 70. The average number of force evaluations for these runs is approximately 400, the same as what was shown in Fig. 7. The

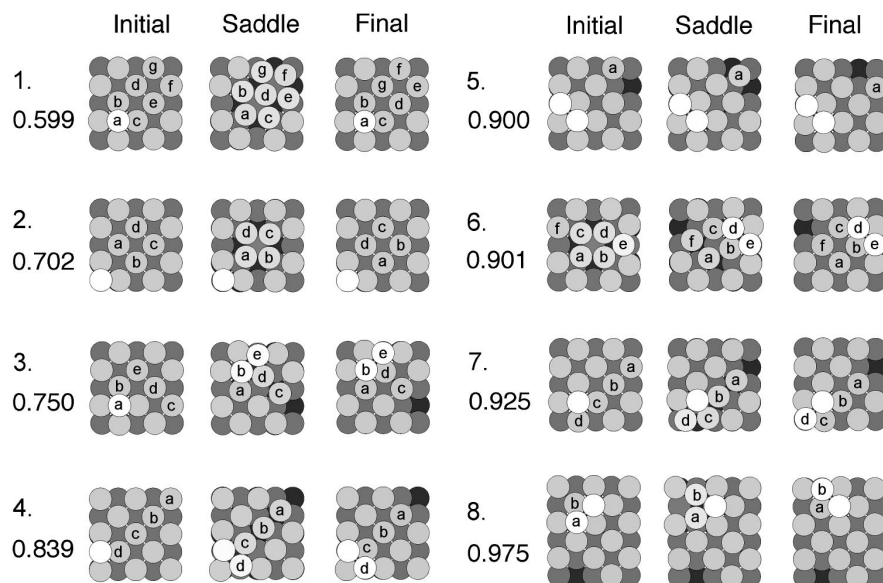


FIG. 11. Searches using orthogonal directions led to many new transitions for the Al/Al(100) system. Some of the more interesting low energy transitions are shown in the figure. The first transition shows the formation of a local hex reconstruction in the surface plane at the saddle point. In the end, a rotation of a group of four surface atoms has occurred. A similar rotation also occurs in the second transition. An addimer/vacancy pair forms in transitions 3, 4, and 7. Neither the initial nor the final state in transitions 5, 6, and 8 corresponds to an adatom on a flat surface, illustrating that the dimer does not always converge on saddle points corresponding to escape routes for the potential energy basin where the initial point of the search is located.

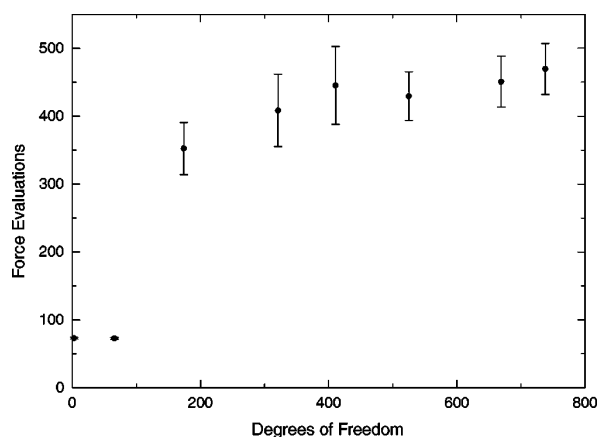


FIG. 12. The scaling of the computational effort for the dimer method, is measured by number of force evaluations, as a function of the system size. These data are taken from the Al/Al(100) system. The substrate consisted of 300 atoms, 50 atoms per layer. Starting with all movable atoms, the number of degrees of freedom was gradually reduced by freezing more and more of the substrate atoms so they became effectively removed from the calculation. If an atom is frozen the dimer cannot be oriented in these degrees of freedom, and the forces on frozen atoms do not affect the dimer. For the smallest number of degrees of freedom, the only mechanism found was the hop. When more than 20 atoms were unfrozen, the full range of saddle points was found. In this region, the plot shows a remarkably slow increase in cost with system size, especially if compared to the  $n^3$  scaling of mode following algorithms which involve the inversion of the Hessian matrix.

data in Fig. 12 show that the method is relatively insensitive to increasing the number of dimensions in the problem, as long as all processes are available to the system. This is very encouraging since it indicates that the dimer method should be useful for finding saddle points in large and complex systems.

#### IV. DISCUSSION

Finding the mechanism and estimating the rate of activated transitions from a given initial state boils down to locating the low energy saddle points at the rim of the potential energy basin corresponding to the initial state (if the harmonic approximation to transition state theory is a good approximation). For problems involving diffusion of atoms in and on crystals this problem is tractable, but far from trivial. Preconceived notions of the transition mechanism can be incorrect, as exemplified by the diffusion of an adatom on an Al(100) surface, which was thought to occur by simple hops of the adatom from one surface site to another until Feibelman discovered that an exchange mechanism is significantly lower in energy.<sup>19</sup> A systematic procedure for finding saddle points is needed for such systems. *Ab initio* calculations of small molecules have over the last decade made use of an efficient mode-following algorithm.<sup>12–14,16</sup> This method has also been applied to small clusters described by empirical potentials, but the mode following algorithm scales poorly with size, making it inefficient for solid state applications. It also requires knowledge of second derivatives of the potential energy with respect to the coordinates of the atoms, preventing its use in the highly successful plane wave based

DFT calculations of solids and surfaces of solids. (Some examples of DFT studies of surface diffusion are Refs. 19, 24–27.) The dimer method presented here can be applied to large systems and since it only relies on first derivatives of the energy, it can be used in conjunction with plane wave based DFT calculations of the atomic forces. We are currently implementing the dimer method in a plane wave DFT code.

An essential aspect of the dimer method is a highly optimized algorithm for rotating the dimer into the lowest energy orientation. This makes it feasible to use the method in conjunction with *ab initio* atomic forces. The calculations for the Al/Al(100) system took on average 400 force evaluations to converge on a saddle point. Since the force calculations on the two images in the dimer are independent, the dimer method parallelizes almost perfectly over two processors if the force evaluation is computationally intensive as in *ab initio* calculations. This means the computational effort is 200 force evaluations per processor. In order to find the set of low lying saddle points, a few saddle point searches have to be carried out. In the Al/Al(100) system, approximately ten searches would suffice to find the three to four lowest saddle points. Using either a collection of randomly chosen initial points or orthogonal searches from a given initial point (or a combination of both), the different saddle point searches can be carried out in parallel. The algorithm will, therefore, be particularly useful when a cluster of processors is available for the computations.

*Note added in proof.* After submitting our manuscript we have learned of a new method by Munro and Wales [Phys. Rev. B **59**, 3969 (1999)] which also enables saddle point searches with only first derivatives.

#### ACKNOWLEDGMENTS

This work was funded by the National Science Foundation, Grant No. CHE-9710995. We would like to thank Peter Feibelman for helpful comments on the manuscript. We would also like to acknowledge Marcus Hudritsch for the excellent graphics library SceneLib which made the output of the simulations so much easier to visualize.

- <sup>1</sup> P. Pechukas, in *Dynamics of Molecular Collisions*, Pt. B, edited by W. H. Miller (Plenum, New York, 1976).
- <sup>2</sup> C. Wert and C. Zener, Phys. Rev. **76**, 1169 (1949).
- <sup>3</sup> G. H. Vineyard, J. Phys. Chem. Solids **3**, 121 (1957).
- <sup>4</sup> H. Jónsson, G. Mills, and K. W. Jacobsen, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne, G. Cicciotti, and D. F. Coker (World Scientific, Singapore, 1998).
- <sup>5</sup> G. T. Barkema and N. Mousseau, Phys. Rev. Lett. **77**, 4358 (1996).
- <sup>6</sup> N. Mousseau and G. T. Barkema, Phys. Rev. E **57**, 2419 (1998).
- <sup>7</sup> G. T. Barkema and N. Mousseau (private communication).
- <sup>8</sup> A. F. Voter, J. Chem. Phys. **106**, 4665 (1997).
- <sup>9</sup> A. F. Voter, Phys. Rev. Lett. **78**, 3908 (1997).
- <sup>10</sup> G. M. Crippen and H. A. Scheraga, Arch. Biochem. Biophys. **144**, 462 (1971).
- <sup>11</sup> R. L. Hilderbrandt, Comput. Chem. **1**, 179 (1977).
- <sup>12</sup> C. J. Cerjan and W. H. Miller, J. Chem. Phys. **75**, 2800 (1981).
- <sup>13</sup> J. Simons, P. Jørgensen, H. Taylor, and J. Ozment, J. Phys. Chem. **87**, 2745 (1983).
- <sup>14</sup> J. Nichols, H. Taylor, P. Schmidt, and J. Simons, J. Chem. Phys. **92**, 340 (1990).
- <sup>15</sup> D. J. Wales, J. Chem. Phys. **91**, 7002 (1989).



- <sup>16</sup>J. Baker, J. Comput. Chem. **9**, 465 (1988).
- <sup>17</sup>D. J. Wales, J. Chem. Phys. **91**, 7002 (1989).
- <sup>18</sup>C. J. Tsai and K. D. Jordan, J. Phys. Chem. **97**, 11227 (1993).
- <sup>19</sup>P. J. Feibelman, Phys. Rev. Lett. **65**, 729 (1990).
- <sup>20</sup>M. C. Payne, M. P. Teter, D. C. Allen, T. A. Arias, and J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).
- <sup>21</sup>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.
- <sup>22</sup>R. G. Della Valle and H. C. Andersen, J. Chem. Phys. **97**, 2682 (1992).
- <sup>23</sup>A. F. Voter and S. P. Chen, Mater. Res. Soc. Symp. Proc. **82**, 2384 (1987).
- <sup>24</sup>G. Brocks, P. J. Kelly, and R. Car, Phys. Rev. Lett. **66**, 1729 (1991).
- <sup>25</sup>A. P. Smith and H. Jónsson, Phys. Rev. Lett. **77**, 1326 (1996).
- <sup>26</sup>R. Stumpf and M. Scheffler, Phys. Rev. B **53**, 4958 (1996).
- <sup>27</sup>P. J. Feibelman, Phys. Rev. Lett. **81**, 168 (1998).