An enhanced splined saddle method

S. Alireza Ghasemi[a)] and Stefan Goedecker
Department of Physics, Universität Basel, Klingelbergstr. 82, 4056 Basel, Switzerland

(Received 3 February 2011; accepted 9 June 2011; published online 6 July 2011)

We present modifications for the method recently developed by Granot and Baer [J. Chem. Phys. **128**, 184111 (2008)]. These modifications significantly enhance the efficiency and reliability of the method. In addition, we discuss some specific features of this method. These features provide important flexibilities which are crucial for a double-ended saddle point search method in order to be applicable to complex reaction mechanisms. Furthermore, it is discussed under what circumstances this methods might fail to find the transition state and remedies to avoid such situations are provided. We demonstrate the performance of the enhanced splined saddle method on several examples with increasing complexity, isomerization of ammonia, ethane and cyclopropane molecules, tautomerization of cytosine, the ring opening of cyclobutene, the Stone-Wales transformation of the C_{60} fullerene, and finally rolling a small NaCl cube on NaCl(001) surface. All of these calculations are based on density functional theory. The efficiency of the method is remarkable in regard to the reduction of the total computational time. © *2011 American Institute of Physics*. [doi:10.1063/1.3605539]

I. INTRODUCTION

Understanding the dynamics of chemical reactions is of great importance in chemistry, biology, physics, etc.[1,2] Transition state properties of an energy landscape play a crucial role in the dynamics of the system. Transition states are saddle points of the potential energy surface (PES). Although in a real reaction the system may not cross the saddle point exactly, it generally passes nearby this point. Transition state theory provides us information of the reaction dynamics by relating the barrier height to the rate of the chemical reaction. A barrier height is the energy difference of a saddle point and one of its neighboring minima. Therefore, in order to understand the dynamics of a system, it is important to determine the location of the saddle points of the corresponding PES. There are two classes of methods for the determination of transition states configurations referred to as single-ended (also known as surface walking) and double-ended (also known as interpolation) saddle point search methods.[3,4] Interpolation methods require knowledge of both the reactant and product configurations in order to generate a sequence of structures between them. The nudged elastic band[5,6] (NEB) and the string method[3,7,8] are the two most well-known and successful schemes among double-ended methods. Usually these methods are used to find an estimate for the minimum energy pathway (MEP). The highest point of the resulting rough approximate MEP is indeed a good starting point for a single-ended saddle point search method. This approach is commonly used and is shown to be satisfactory. But the very large numerical cost of this approach makes the saddle point search sometimes prohibitive.[3] Typically the initial pathway in the double-ended methods is the linear synchronous transit (LST) joining the two configurations. The difficulty with this choice is that they usually fail to converge if the MEP lies far from the LST. In addition, electronic structure calculations may fail to converge for some interior configurations of the LST if they lie on high energy portions of the potential energy surface. While the interpolation methods serve as a precursor calculation, the surface-walking methods complement the task to find the saddle point energy and coordinates with prescribed accuracy. The well-known dimer method[9] and its improved successors[4,10–12] and Lanczos method[13–15] replace the gradient by a modified gradient which its projection along the negative curvature is inverted. The accuracy of the modified gradient depends on the accuracy of the negative curvature direction defined in such methods, in case of the dimer method approximated by the dimer direction. These methods currently belong to preferred choices among scientists. However, in recent years the climbing image NEB (Ref. 16) (CI-NEB) has proved to be an efficient method to determine saddle points.[17] In this approach one avoids to run two separate optimization, i.e., one NEB run followed by a surface walking calculation from the image with maximum point to find the saddle point. Instead, the highest energy image, during a CI-NEB run, feels no spring forces and climbs to the saddle point via a reflection in the force along the tangent of the pathway. In fact this scheme is designed for understanding the mechanism of reactions qualitatively and the saddle point configuration and its energy are given with a higher accuracy compared to a raw NEB calculation. If highly accurate MEP is required, it is indeed more efficient to find the saddle point then the accurate description of the reaction pathway can be followed by descending from saddle point to minima using high order integration schemes such as fourth order Runga-Kutta integrator.[17] One possible problem can occur in the following situation: if the image with the highest energy switches to other ones, then an image which was not feeling spring forces, abruptly feels spring forces which might be large. Since an overview of available saddle point search methods goes beyond the scope of this paper, we thus refer to Ref. 18. In addition, several comparative

[a)]Author to whom correspondence should be addressed. Electronic mail: alireza.ghasemi@unibas.ch.

 © 2011 American Institute of Physics

Downloaded 16 Feb 2012 to 131.152.33.215. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

studies of saddle point search methods have been published by different groups, e.g., Refs. 10, 17, 19, and 20.

Saddle points are like minima stationary points of PES. Nevertheless finding a saddle point is much more difficult than finding a minimum. So why is finding saddle points a challenging task compared to finding minima? The fundamental reason is due to the fact that to approach a minimum it is known that one must always go down in energy landscape and there exists an indispensable tool to accomplish it, namely negative of the potential gradient. This aspect makes performing and controlling a minimization procedure to be much easier than that of a saddle point optimization, in which depending on the current location on the PES one should move upward and sometimes downward through the energy landscape. In 2008 Granot and Baer[21] introduced a novel approach that reformulates the saddle point search to a minimization by introducing an energy functional in which a minimum point is associated to a saddle point of the original PES. Even though the idea of splitting a saddle point optimization into two orthogonal directions was first introduced by Halgren and Lipscomb,[22] the approach used by Granot and Baer is different and it benefits from the advantage of having a target function with a well-defined gradient for minimization. This method looks promising since in its first version it was already competitive with methods which have been developed and improved for decades.

We present an enhanced version of their method and we will refer to their method in this paper as the splined saddle method. Our modifications significantly enhance its efficiency and reliability. Furthermore, we discuss about specific features of the splined saddle method that make this method complicated and a potential for future improvements. Three separate improvements are introduced: (1) replacement of finding the root of the first derivative by maximization of the energy along the pathway using Newton method utilized by splines, (2) the use of a preliminary NEB run to reduce the possibility of partial overlap of the pathway with an equipotential curve, (3) the use of a hybrid approach for the evaluation of the quantum mechanical potential energy/forces. These improvements concern the efficiency of the method in order to minimize calls to the quantum mechanical code. In addition, we discuss about the choice of an appropriate optimizer which increases the reliability of the splined saddle method. The performance of the enhanced splined saddle method is then assessed for several molecular reactions with a variety of complexity. These examples demonstrate that a significant reduction in computational effort can be achieved without a loss in accuracy using the enhanced splined saddle method.

II. THE METHOD

Here we briefly reiterate the derivation of the splined saddle method. Having the energy functional v_{max} defined as the maximal energy along the path $\mathbf{q}_p$,

$$v_{max}[\mathbf{q}_p] = \max_{t\in[0,1]} v(\mathbf{q}(t)) = v(\mathbf{q}(t_{max}[\mathbf{q}_p])),$$

where $v(\mathbf{q})$ is the potential energy function and $t_{max}[\mathbf{q}_p]$ is the pathway parameter at the point with the maximum potential along the path. Notice that $\mathbf{q}$ stands for the coordinates of points along the pathway, while $\mathbf{q}_p$ stands for the pathway and its parametrization.

$$\frac{\delta v_{max}[\mathbf{q}_p]}{\delta \mathbf{q}_p(t)} = \nabla v(\mathbf{q}(t_{max}[\mathbf{q}_p])).\left[\frac{\delta \mathbf{q}(t_{max}[\mathbf{q}_p])}{\delta \mathbf{q}_p}\right],$$

where the following identity is incorporated:

$$\nabla v(\mathbf{q}(t_{max})).\dot{\mathbf{q}}(t_{max}) = \left.\frac{d}{dt} v(\mathbf{q}(t_{max}))\right|_{t=t_{max}} = 0.$$

Using spline to represent the path, $\mathbf{q}_p(t)$ is replaced by the spline path $\tilde{\mathbf{q}}(t)$ which passes through the anchor points Q_α where $\alpha = 0, \ldots, N_a$. Q_0 and Q_{N_a} represent one end (say reactant) and the other end (say product), respectively, and both are fixed. The approximate maximal energy functional is thus

$$\begin{aligned}\tilde{v}_{max}(Q_1, \ldots, Q_{N_a-1}) &= v_{max}[\tilde{q}] \\ &= v(q(t_{max}(Q_1, \ldots, Q_{N_a-1}))).\end{aligned} \tag{1}$$

The gradient of $\tilde{v}_{max}(Q_1, \ldots, Q_{N_a-1})$ is given by

$$\frac{\partial \tilde{v}_{max}}{\partial Q^i_\alpha} = \sum_{j=1}^{N} \frac{\partial v(\tilde{\mathbf{q}}(t_{max}))}{\partial q^j} J_i^{j,\alpha}, \qquad \begin{cases}\alpha = 1, \ldots, N_a - 1 \\ i = 1, \ldots, N\end{cases}, \tag{2}$$

where the Jacobian is defined as

$$J_i^{j,\alpha} \equiv \frac{\partial \tilde{q}^j(t_{max})}{\partial Q^i_\alpha}, \qquad \begin{cases}\alpha = 1, \ldots, N_a - 1 \\ i, j = 1, \ldots, N\end{cases}. \tag{3}$$

Having the function value $\tilde{v}_{max}$ and its gradient, in principle any conventional minimization method such as steepest descent (SD), conjugated gradient, Broyden-Fletcher-Goldfarb-Shanno (BFGS), fast inertial relaxation engine[23] (FIRE), which need only the function value and its derivative, can be used. In Sec. IV we argue that the use of conjugated gradient, BFGS and the limited memory BFGS (Ref. 24) (LBFGS) methods, if used in connection with line searches, for the peculiar function $\tilde{v}_{max}$ can be problematic.

In fact, the splined saddle method splits the optimization problem of finding a saddle point into two optimizations consisting of a global maximization in the inner loop and a minimization in the outer loop. This reformulation of the saddle point optimization may seem to be far from being efficient since two interdependent optimizations are involved. This is indeed not true because the inner loop is a global maximization of a univariate function which can be done efficiently and presumably safely.

It must be emphasized that the calculation of the Jacobian in Eq. (3) scales quadratically with respect to the number of particles in the system and the prefactor is proportional to the number anchor points. As a consequence, it is not recommended to use the splined saddle method for calculations where the energy/force evaluations are performed with force fields in which energy/forces are typically computed linearly with respect to the number of particles. In contrast, the *ab initio* methods have larger scaling than $\mathcal{O}(N^2)$, typically $\mathcal{O}(N^3)$, or the calculation of the Jacobian in the splined saddle method is computationally negligible compared to the computational cost of the

Downloaded 16 Feb 2012 to 131.152.33.215. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

quantum mechanical energy/forces evaluations. Hence this problem does not concern them. All calculation details of the Jacobian, when the pathway is represented by natural cubic spline and the parametrization is based on the cumulative inter-anchor point distance, are given in Appendix A. The calculation details of the Jacobian is provided because it is not given in the original paper and it can differ depending on the choice of the spline and the parametrization.

III. PRONOUNCED FEATURES OF THE SPLINED SADDLE METHOD

An interesting feature of this method is that the pathway solution is not unique. In fact, any pathway which starts at the reactant and while passing through the saddle point ends at the product is a solution, provided there is no point along the pathway which has potential value higher than that of the saddle point. Theoretically, there can be infinite number of such pathways, i.e., minima of the energy functional $v_{max}[\mathbf{q}_p]$, all of which have the same energy value, namely the potential value at the saddle point. On the other hand pathways in this method are represented by splines. Thus, those pathways of the functional $v_{max}[\mathbf{q}_p]$, that can be represented by piecewise polynomials, are obtained in our calculations. In other words we can find the minima of the function $\tilde{v}_{max}(Q_1, \ldots, Q_{N_a-1})$ and such minima are a subset of solutions for the functional $v_{max}[\mathbf{q}_p]$.

To examine this feature numerically, we test it for a simple model potential, which has been also used in Ref. 8 as an example, given by the following equation:

$$v(x, y) = (1 - x^2 - y^2)^2 + y^2/(x^2 + y^2). \tag{4}$$

This model potential has two minima at $(-1, 0)$ and $(1, 0)$ and one saddle point in upper semispace $y > 0$ at $(0, 1)$. The MEP passing through this saddle point is the semi-circle with radius one centered at $(0, 0)$. Since the non-uniqueness of $\tilde{v}_{max}$ minima should in principle be valid for different number of anchor points, we initiated two sets of runs with $N_a - 1 = 1$ and 2. Each set consists of 100 runs with random initial positions of anchor points, restricted to some regions discussed in details in Appendix B. In all 200 runs the splined saddle method succeeded to find the saddle point and the final anchor points are all different, each one corresponds to a different pathway, five of those with $N_a - 1 = 2$ are shown in Fig. 1 and five of those with $N_a - 1 = 1$ are shown in Fig. 2. In one run, the solid (green) curve shown in Fig. 1, it took many force evaluations (311 which is an order of magnitude larger than the average value given in Table I) to find the saddle point. The reason is that the high energy portions of the mentioned pathway is passing by an equipotential curve so that finding the maximum point along the pathway is severely badly conditioned.

Although these solutions of $\tilde{v}_{max}$ are probably continuous, it cannot be confirmed exactly by numerical calculations. Nevertheless, we performed a large number of runs with $N_a - 1 = 1$. All the points on the solid (black) curve shown in Fig. 2, which is in fact made by dense points, are minima of the function $\tilde{v}_{max}$. Even though the function $\tilde{v}_{max}$ with $N_a - 1 = 1$ has two degrees of freedom, the corresponding

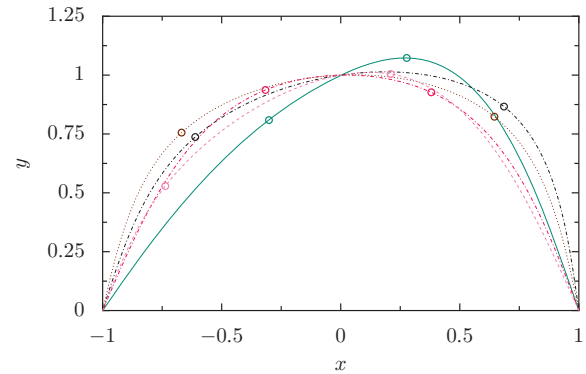

FIG. 1. Illustration of five pathways (the minima of $v(\tilde{\mathbf{q}}(t_{max}(Q_1, \ldots, Q_{N_a-1})))$ with $N_a - 1 = 2$) which all pass through the same saddle point of the two-dimensional model potential, i.e., the point (0, 1). The solid (green) curve nearly overlaps with an equipotential curve so that maximization becomes formidable.

Hessian of the $\tilde{v}_{max}$ for $N_a - 1 = 1$ lead to one zero eigenvalue. The other eigenvalue varies drastically through the solid curve (black). Notice that the original model potential in Eq. (4) has two degrees of freedom and its Hessian at the saddle point has two non-zero eigenvalues. The eigenvector associated with zero eigenvalue is parallel to the tangent of the solid (black) curve in Fig. 2. Those runs which end at an anchor point associated with a large condition number, took many iterations to converge. All splined saddle runs in this section were perform using the SD method.

The two problems of the splined saddle method, the badly conditioned maximization of the inner loop which corresponds to overlapping of the pathway with equipotential curves and the occasional large condition number of $\tilde{v}_{max}$, can affect the efficiency of the method drastically. In Sec. IV B, we present a remedy to avoid the former problem. In a completely different point of view, the latter can be considered as an advantageous feature which recommends to locate at least one anchor point near the saddle point. Since the saddle point is not known in advance, locating one anchor point at the pathway maximum before starting minimization can significantly increase the efficiency of the method. This is the third step of the algorithm as shown in the flowchart of Fig. 3.

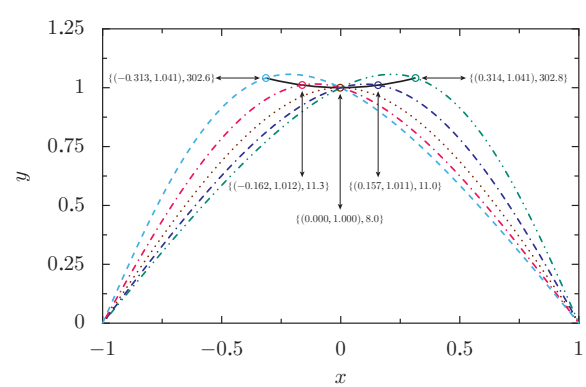

FIG. 2. Illustration of five pathways with $N_a - 1 = 1$ in which the saddle point (0, 1) has the highest energy along these pathways. The anchor point of each pathway is shown by circle and its coordinate (the first two numbers) and the corresponding nonzero Hessian eigenvalue of the function $\tilde{v}_{max}$ (the third number) are given in brackets.

Downloaded 16 Feb 2012 to 131.152.33.215. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

TABLE I. The average, minimum, and maximum number of energy/force calls used to converge to the saddle point of the model potential (presented in Sec. III) for two tolerance values $tol = 10^{-5}, 10^{-7}$, where tol is defined in Sec. V A. 100 runs are performed with $N_a - 1 = 1$ while the initial anchor points are located randomly with restrictions described in Appendix B.

Tol	Minimum	Maximum	Average
10^{-5}	25	59	35
10^{-7}	26	71	42

IV. IMPROVEMENTS

The energy landscape of the function $\tilde{v}_{max}$ is peculiar and for this reason particular attention must be paid for the choice of minimization method to be used for the splined saddle method. To some extent, the authors agree with Ref. 21 that the limited memory BFGS (LBFGS) can perform efficiently. But if it is used with line searches, it becomes fragile for this peculiar energy landscape. In regard to reliability, BFGS or any method based on line search not only might fail but also can deteriorate the efficiency in the inner loop, i.e., the maximization part. Methods such as SD and FIRE are in essence appropriate for the splined saddle method but not necessarily efficient. An alternative choice can be the use of an efficient method such as BFGS and replacing the underlying line search by a simple feedback mechanism based on the target function, i.e., $\tilde{v}_{max}$. In our implementation of BFGS, the quasi-Hessian of the BFGS method is reset to unit matrix times some constant, whenever the energy rises more than some prescribed amount.

A. Hybrid approach

Unlike empirical potentials in which the energy/forces are usually obtained with a fixed computational cost and accuracy, the cost can strongly vary in *ab initio* calculations depending on the level of the theory and the settings of the various parameters which influence the accuracy. For example, energy/forces in *ab initio* methods are typically obtained in a self-consistent calculation in which the corresponding convergence tolerance can affect the required number of iterations thereby computational cost. Another decisive parameter both with respect to computational cost and accuracy is the size of the basis set. Above all, the level of theory influences considerably the accuracy of the energy/forces and the computational cost. Therefore, in the context of electronic structure calculation, developing and improving methods in order to reduce the total number of energy/force evaluations is not the only strategy to minimize the computational burden of determining transition state configurations. Instead, an adequate strategy for determining transition state configurations is to perform calculations at a low level of theory with minimal basis sets and loose convergence tolerance whenever a low accuracy is sufficient and perform calculations at a higher level of theory with larger basis sets and tight convergence tolerance whenever a high accuracy is required. This hybrid strategy has already been used by Goodrow and co-workers[25,26] and it is shown to be efficient.

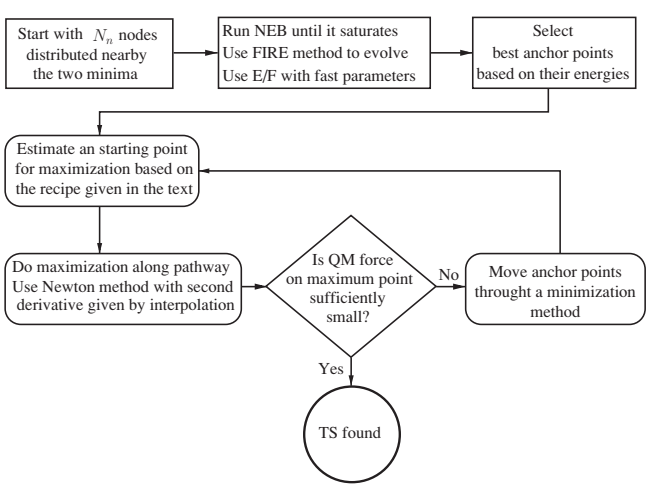

FIG. 3. A flow sheet of our strategy for finding the saddle point. It consists of two major parts that is a preliminary NEB run followed by the splined saddle method. In the hybrid strategy all these calculations are performed with low-accuracy calculations except the energy/forces at the maximum point. E/F stands for energy/force.

In other methods such as NEB and string methods, the potential force at the movable images are used to iterate to the next step. In the spline saddle method, the potential/forces values at anchor points play no role in moving them, and only the force at the maximum point along the pathway is involved. This is a profound difference which distinguishes the splined saddle method to other methods. However, to find the maximum, we need to calculate the potential (and forces if it is planned to be used) values at some points along the pathway. Assuming that the location of the maximum point along the pathway displaces only slightly when using lower accuracy calculations such that it does not affect the outer loop, one can perform all the calculations in the process of finding the maximum point using lower accuracy calculations and only at the maximum point recalculate the energy/forces using high accuracy calculations. This assumption may be severely violated if one performs calculations with different levels of theory. As an extreme example, it is known that at some circumstances the saddle points can even disappear in density functional theory (DFT) with local density approximation compared to higher quantum mechanical level of theory.[27] Therefore, one should use the hybrid strategy with special care whenever the inner and outer loops are performed with different levels of the theory. To a reasonable confidence, the aforementioned assumption can be valid with a high probability if only one level of theory is used but different method parameters such as the size of the basis set. For systematically extendable basis sets such as wavelets and plane waves, this procedure can be readily applied. In conclusion, we propose that in the inner loop *ab initio* calculations to be performed with a looser convergence criteria and a less fine basis set than the target accuracy requires. Subsequently the maximum point must be recalculated with the tight convergence criteria and fine basis set appropriate for the target accuracy. This is what we have successfully accomplished for our test systems, as it is demonstrated in Sec. V.

Downloaded 16 Feb 2012 to 131.152.33.215. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

B. Adding NEB

The 2D model potential has been helpful to reveal possible challenges which one might encounter in the splined saddle method. For example, if the high-energy portions of the pathway coincides with an equipotential curve, then the problem is ill-conditioned in two respects: one is the maximization will be harder; second the functional will have a large condition number as it is illustrated for the 2D model potential in Fig. 2. We have also observed similar conditions for the atomistic systems treated by *ab initio* methods. Therefore, it is very important to avoid such situations. Our recipe to this problem is to perform a preliminary NEB run with a few movable images. Since no accurate calculation with tight convergence criteria is expected for the preliminary NEB run, one can perform such calculations with fast parameters or with low level of theory. Therefore, this additional part will be a small fraction of the total computational time. A typical problem in interpolating saddle point search methods is that the LST intermediate configurations are physically unreasonable. Because the anchor points in the splined saddle method need not be distributed uniformly along the pathway, one can distribute initial NEB nodes near the two ends and initiate the NEB optimization smoothly. Our inspections show the FIRE method performs excellently for such circumstances. Using this recipe, we can circumvent two problems by adding the preliminary NEB part. In fact, the preliminary NEB run is crucial and regarding our investigations, the splined saddle method was unable to converge without incorporating it for most of the test cases in Sec. V.

C. Maximization

As mentioned in Ref. 21, the inner loop is where this method can be improved to make it more efficient. In Ref. 21, the Brent method is used to find the maximum point along pathway by locating the root of $\frac{d\tilde{v}(t)}{dt}$. In order to provide the input interval for the Brent method, the potential is computed at $K \times N_a$ points equally spaced with respect to the pathway parameter. A reasonable value for K is proposed to be 4. In Ref. 21 it is mentioned that a typical number of anchor points for realistic problems is said to be $N_a - 1 = 4$. This means $K \times N_a = 20$ potential evaluations (no force evaluation is needed) even before starting to find the root by means of Brent method. This is obviously inefficient.

We propose a different scenario by performing the maximization using Newton method while using all information at hand. In order to provide a good initial point for a Newton method, the potential and forces at n_v points are computed. For low-dimensional model potentials like the one in Eq. (4), $n_v = 1$ is sufficient and for more realistic problems $n_v = 6$ might be necessary. If information from forces are considered as well, the data v_i and $\frac{dv_i}{dt}$ at $n_v + 2$ (including the two end points which need to be calculated once in the beginning of the simulation) are available. In fact $\tilde{v}(t)$ can be approximated by spline. In order to choose an appropriate interpolation method we should consider two points; first is to use as much information as we have; second is to invoke a spline which has one more derivative continuity than that of we plan to extract from it. For the case of a pathway, the first derivative is needed so that the second derivative should be continuous. Thus the natural cubic spline is an optimal choice to interpolate the pathway. Since during a maximization with the Newton method, the second derivative $\frac{d^2v}{dt^2}$ is needed and taking into account the fact that v_i and $\frac{dv_i}{dt}$ are available, the best choice to interpolate $\tilde{v}(t)$ seems to be a quintic spline. Using this spline, the function value and its first, second, and third derivatives are continuous (see Appendix C for detailed information about this spline). Our investigation shows that the quintic spline performs excellently for the 2D model potential, both to provide a sufficiently precise initial guess for maximization and to estimate accurately the second derivative value required by the Newton method. Unfortunately, this is true only for the model potentials or force fields in which there is no noise in the energy and force values and the force is exact derivative of the potential energy. In *ab initio* methods, noise is a ubiquitous ingredient which can be reduced at the expense of computational efficiency.

A very good initial guess for the maximization process can be obtained by finding the list of the maxima of the spline $\tilde{v}(t)$ and selecting the one with the highest potential value. For *ab inito* calculations, the natural cubic spline is more appropriate and reliable for the interpolation of $\tilde{v}(t)$ compared to the quintic spline since it is made of lower degree polynomial. In our implementation using natural cubic spline, we make a list of all maximum points of the potential along the pathway and then select the one with largest potential value as the starting point for maximization using the Newton method. The first and second derivatives in natural cubic spline are obtained globally so it is more sensitive to local rapid variations of the derivatives. Therefore, in order to obtain an estimate for the second derivative, which is needed by the Newton method, one can achieve a more reliable second derivative value using information of local points, i.e., neighboring points to the current maximum point. A quadratic function fitted to the potential values of the maximum point and its two neighboring points provides a remarkably reliable second derivative value. This approach is used for all of our atomistic test systems and it performed successfully in various conditions such as small and large curvature near the maximum point.

V. EXAMPLES

A. Two-dimensional model potentials

Previously, the model potential in Eq. (4) was used to extend our understanding of the spline saddle method. We also consider it as a test case to assess the performance of the splined saddle method. The difference between the exact saddle point and the calculated one is used as the criterion to stop the calculation, $|\mathbf{q}_{sp}^{true} - \mathbf{q}_{sp}^{calculated}| < tol$. The maximum, minimum, and average number of force calls to find the saddle point for two different tolerance values are given in Table I. Based on statistical data, Table I shows that two orders of magnitude more accurate results can be achieved by only a little increase in computational cost. Only one anchor point $N_a - 1 = 1$ is used for this simple test case. We have also compared the efficiency of the CI-NEB, original

Downloaded 16 Feb 2012 to 131.152.33.215. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

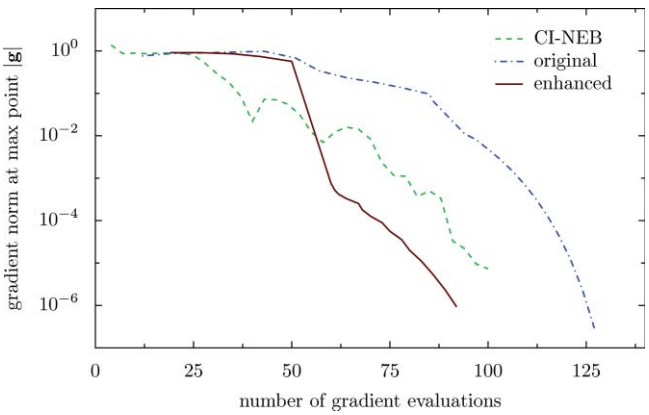

FIG. 4. Convergence of the CI-NEB, original splined saddle, and the enhanced splined saddle methods for the Müller-Brown model potential. The same optimization method, namely BFGS, with three movable images $N_a - 1 = 3$ was used in all three cases.

splined saddle, and the enhanced splined saddle methods for the Müller-Brown model potential and the results are shown in Fig. 4. The enhanced splined saddle is more efficient than the original version mainly due to the maximization part of the algorithm. Figure 4 shows that the performances of CI-NEB and enhanced splined saddle are comparable for model potentials since the hybrid approach cannot be exploited for model potentials.

B. Ammonia

Computational efficiency of a saddle point search method becomes particularly valuable when atomic interactions are based on first-principle calculations. Hence, in order to demonstrate the performance of the splined saddle with the modifications described above in the context of *ab initio* methods, we have chosen several prototypical systems of which some have been used previously in other papers on transition state search methods. The first example is the ammonia molecule which has two degenerate energy minima. It contains 12 Cartesian degrees of freedom including translational and rotational ones. Throughout this manuscript we use the following notation $P = \{N_n - 1, N_a - 1, n_v\}$ where the three elements denote the number of nodes during NEB part, number of anchor points in the splined saddle method, number of points along the pathway to obtain an input guess point for maximization, respectively. In this example we used $P = \{2, 1, 1\}$ because the reaction pathway is not sophisticated. The most legitimate way to assess the performance of an optimization method is usually obtained on the basis of the number of energy/forces calculations required to converge the extremum point rather than the computational time which strongly depends on the *ab initio* code used for energy/forces evaluation and on the computer on which the program is executed. In the hybrid approach there are two average times for force evaluation within the more accurate and less accurate calculations. However, the more accurate is the main computational setup and the saddle point energy and structure is based on that. Therefore, we use the following notation $G = \{ng_1, ng_2, \mathbf{ng_3}\}$, where ng_1 is the number of high-accuracy energy/forces evaluations using fine parameters, ng_2 is the number of energy/forces evaluations with loose parameters, $\mathbf{ng_3}$ is ng_1 plus the ratio of average computational time for low-accuracy/high-accuracy calculations times ng_2. $\mathbf{ng_3}$ will be referred as the effective number of force calls. All the DFT calculations are performed with the BIGDFT (Ref. 29) package, a pseudopotential based[30] DFT code with a wavelet basis set. Wavelets are a systematically extendable basis set and the basis size was chosen sufficiently large that energies were converged to better than to 10^{-6} eV. All DFT calculations were performed at the Γ-point. Grid spacing of 0.16 Å is used in all calculations which contain hydrogen and 0.22 Å is used for C_{60} and NaCl systems. The low accuracy calculations were performed with nearly 25% larger grid spacing. The number of grid points determines the number of basis functions used in the calculation. While the convergence criteria for wave function optimization in the accurate calculations was 10^{-5} a.u., we used 10^{-4} a.u. for the low accuracy calculations. The BIGDFT has two levels of adaptivity,[29] a high resolution region that contains all the chemical bonds and a low resolution region further away from the atoms where the wave functions decay exponentially to zero. We used radii of approximately 3 bohrs and 11 bohrs to define the high and low resolution regions around each atom, respectively. The low accuracy calculations were performed with nearly 20% smaller values for these two radii.

The structure of ammonia molecule is pyramidal and the saddle point which connects the ammonia molecule to its mirror plane conformer is a planar structure, as shown in Fig. 5(a). The saddle point was easily found and the gradient evaluation set is $G = \{7, 44, \mathbf{15}\}$. The corresponding barrier height is very low, 0.16 eV. All the geometry optimizations, including the minimization for relaxing the products and reactants as well as finding the saddle point were performed such that the maximum force component at the minima or the saddle point is less than 0.01 eV Å^{-1}. This specification is chosen as a reasonable compromise between accuracy and computational cost. This convergence criterion is used for all of the atomistic examples.

C. Ethane

Rotation of three hydrogen atoms around C-C axis in an ethane molecule is our second example with a low barrier of 0.11 eV. Similar to the calculations in the ammonia molecule the splined saddle method was able to locate the transition state very efficiently, with $G = \{15, 95, \mathbf{33}\}$ while $P = \{2, 2, 2\}$ was used. This reaction is illustrated in Fig. 5(b).

D. Cytosine

As a next step to a more challenging test system we perform calculations for the cytosine tautomerization[28] which includes the transfer of an atom, i.e., a bond breaks and a new bond with another atom is formed. Unlike the other two test systems, the reactant and product in cytosine tautomerization process are asymmetric and the corresponding barrier height is moderately large. A larger number of nodes and anchor points were required for this system. In our calculations

Downloaded 16 Feb 2012 to 131.152.33.215. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

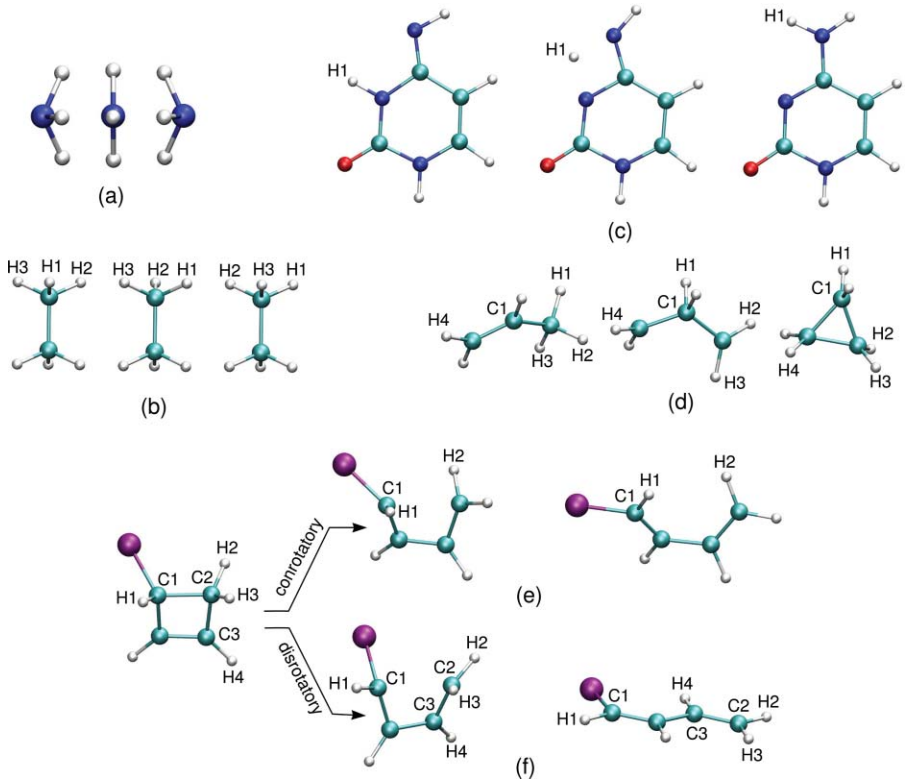

FIG. 5. The reactant, saddle point, and product configurations of all organic molecular test systems are illustrated, (a) the isomerization of the ammonia molecule, (b) the isomerization of the ethane molecule, (c) tautomerization of the cytosine molecule, (d) intermolecular rearrangement of cyclopropane to propylene, (e) conrotatory ring opening of 1-chloro-2-cyclobutene to 1-chlorobuta-1,3-diene, and (f) disrotatory opening ring of 1-chloro-2-cyclobutene to *trans*-1-chlorobuta-1,3-diene. In each inset the configuration in the middle is saddle point.

$P = \{5, 5, 5\}$ was used which lead to a fast convergence both in preliminary NEB part and the splined saddle minimization at the cost of $G = \{24, 192, \mathbf{59}\}$ energy/forces evaluations. Figure 5(c) shows the reactant, transition state, and the product of the cytosine tautomerization.

E. Propylene

Transition from propylene to its higher energy isomer cyclopropane is our first test system which consists of global conformation of the molecule. Even though the molecule is small and consist of only 9 atoms, finding the corresponding transition state is a challenging task. In fact when finding the transition state with interpolating saddle point search methods, the complexity of the reaction pathway plays a more important role than the degrees of freedom. The saddle point was found with energy/force set $G = \{39, 501, \mathbf{122}\}$. The number of energy/force evaluations with fast parameters increased significantly but this is computationally much less demanding, hence the effective number of energy/force evaluations is satisfactory. Since the reaction pathway is isomerization of cyclopropane is intricate, $P = \{5, 7, 8\}$ was used so that the maximum point can be found confidently and the pathway described by spline to be sufficiently flexible. The judgement for the choice of number of points in different parts of the method can be based on the complexity of the LST. The intricacy of the reaction pathway can be seen in Fig. 5(d).

F. 1-Chloro-2-cyclobutene

In order to demonstrate the applicability of the splined saddle method to more challenging systems, which in practical situations can occur, we applied it to isomerization of cyclobutene to *cis*-butadiene. It has also been used as a test system in Ref. 11 to evaluate the performance of the dynamical dimer method.[11] In two distinct reaction pathways, the 1-chloro-2-cyclobutene molecule either conforms to *cis*-1-chloro-1,3-diene or to *trans*-1-chloro-1,3-diene. These two products are associated to conrotatory and disrotatory isomerization mechanism, respectively. The splined saddle methods found the corresponding saddle points of the two reactions successfully in a reasonably low computational cost, with $G = \{26, 223, \mathbf{65}\}$ and $G = \{28, 346, \mathbf{86}\}$ for conrotatory and disrotatory isomerization mechanisms, respectively. $P = \{4, 4, 4\}$ was used for both of these calculations.

G. Stone-Wales transformation in C_{60} fullerene

In order to assess the performance of the enhanced splined saddle method for a larger system, we apply it to a well-known process, namely the Stone-Wales transformation in a C_{60} fullerene. This is a well studied reaction mechanism where two carbon atoms effectively rotate through 90° about the midpoint of the bond that connects them. Buckminsterfullerene (first lowest C_{60} fullerene configuration) is thus connected to the second lowest fullerene isomer, which has C_{2v} symmetry and two pairs of adjacent pentagons. Even though

Downloaded 16 Feb 2012 to 131.152.33.215. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

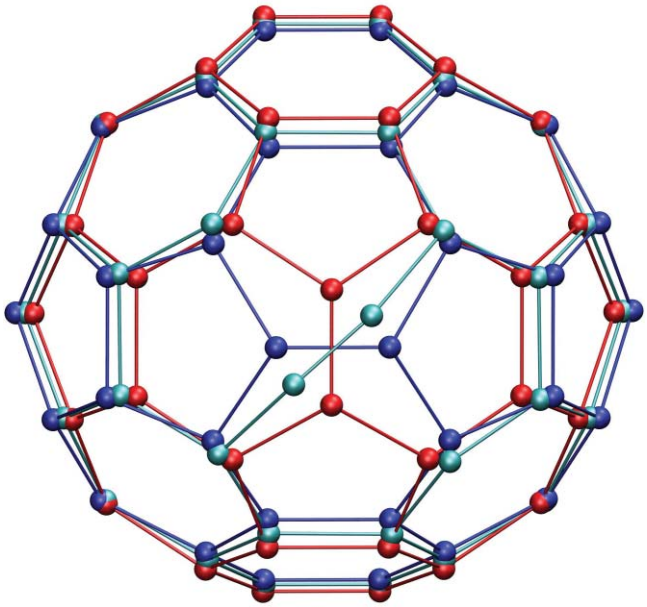

FIG. 6. Illustration of Stone-Wales transformation of C_{60} fullerene to the first excited isomer. The ground state (in which the bond at the center is vertical), saddle point (in which the bond at the center is diagonal/inclined), and the final state (in which the bond at the center is horizontal) are superimposed to stress the sizable displacement of many carbon atoms beside the 90° rotation of the two atoms. For clarity, the carbon atoms in the background are not shown.

the reaction involves the rotation of only two bonded carbon atoms, the displacement of other atoms is important as it is illustrated in Fig. 6. The uphill and downhill barrier heights, 7.14 and 5.56 eV, are slightly larger than previous studies.[31] The gradient set is $G = \{44, 333, \mathbf{152}\}$ while $P = \{4, 3, 7\}$ was used.

H. Rolling NaCl cube on a NaCl(001) surface

Finding transition states in reactions occurring on a surface is interesting and it has been also used as test cases in saddle points search methods.[17] We apply the enhanced splined saddle method to rolling of a sodium chloride $2 \times 2 \times 2$ cube on NaCl(001) surface by 90°, as shown in Fig. 7. The surface is modeled by two layers, each with 36 atoms. The atoms in the bottom layer are frozen in bulk lattice sites. The system, including the cube and surface atoms, consists of totally 80 atoms and the surface boundary conditions is applied, i.e., periodic in the lateral directions to the surface and free in the third one. This is a challenging test case since rolling a cluster on a surface involves a nontrivial movements of all cluster atoms. In fact this example can be considered as a prototypical case of rolling clusters and molecules on surfaces that has been popular in computational nano-science in recent years. We used $P = \{2, 3, 2\}$ and the splined saddle method converged to the saddle point with a reasonable efficiency, $G = \{50, 195, \mathbf{99}\}$. The starting and final states are similar and the corresponding saddle point lies exactly at the middle of the pathway. The barrier height is notably low 0.48 eV which leads the conclusion that the NaCl cube can roll in a reasonably short time about some tens of nanosecond at room temperature, according to harmonic transition state theory.

VI. DISCUSSION

The splined saddle method has particular features and provides useful flexibilities for the renowned saddle point optimization problem. The primary advantage of the splined saddle method is that the saddle point optimization problem is transformed to a minimization problem which is well studied and a number of highly optimized methods are available for. The price for this transformation is the replacement of one optimization by two interdependent optimizations, namely outer loop minimization and inner loop global maximization. More precisely, the stability of the method strongly depends on the success of the inner loop. The energy landscape of the new function $\tilde{v}_{max}$, of which minima are associated with a saddle point of the main potential, is rugged and unusual. In fact, an infinite number of $\tilde{v}_{max}$ minima associated with one saddle point of the original energy landscape. In the first glance, it looks to be only useful but as we have shown in Sec. III not all these minima can be achieved with a reason-

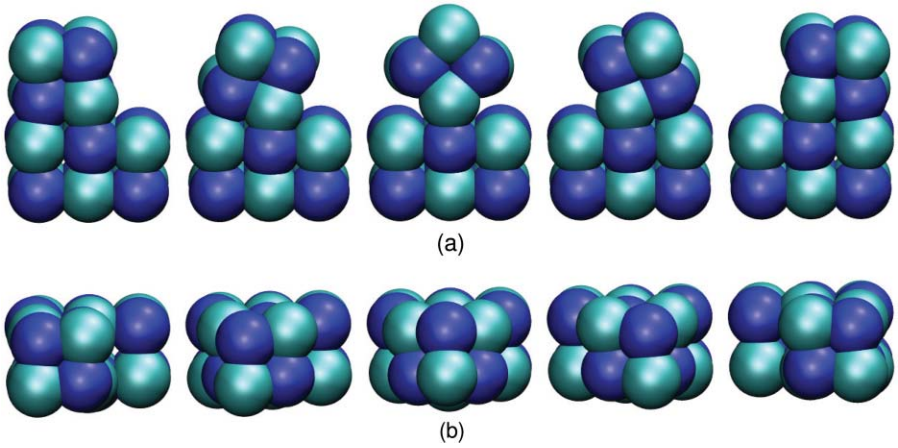

FIG. 7. Side view (a) and top view (b) of rolling an NaCl cube on NaCl(001) surface are shown. The configurations from left to right are as following: starting, an intermediate configuration of which the normalized pathway parameter is 0.25, the saddle point which is exactly at the middle of the reaction pathway, an intermediate configuration of which the normalized pathway parameter is 0.75, the ending. For clarity, only surface atoms around the NaCl cube are shown.

Downloaded 16 Feb 2012 to 131.152.33.215. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

TABLE II. The performance of the splined saddle method for all atomistic test systems. The barrier height (eV), the number image/anchor/node points in different parts of the method, and the number of force calls to quantum mechanical code with low- and high-accuracy parameters are listed below.

System	Barrier heights (eV)	Point set P =	Gradient set G =	If P={4,4,4} is used, G =	CI-NEB
NH3	0.16/0.16	{2,1,1}	{7,44,**15**}	{6,106,**29**}	72
Ethane (C2H6)	0.11/0.11	{2,2,2}	{15,95,**33**}	{18,188,**59**}	165
Cytosine	1.46/1.36	{5,5,5}	{24,192,**59**}	{24,200,**63**}	193
Propylene	3.28/3.40	{5,7,8}	{39,501,**122**}	{55,418,**133**}	327
Conrotatory ring opening[a]	1.20/1.37	{4,4,4}	{26,223,**65**}		175
Disrotatory opening ring	1.84/2.18	{4,4,4}	{28,346,**86**}		failed
SWT in C_{60}	7.14/5.56	{4,3,7}	{44,333,**152**}	{26,239,**103**}	143
Rolling NaCl cube on NaCl surface	0.48/0.48	{2,3,2}	{50,195,**99**}	{53,259,**129**}	215

[a]In Ref. 11 more than 1000 force evaluations was required using dynamical dimer method.

able computational cost. However, due to the flexibilities in the splined saddle method it is possible to avoid badly conditioned circumstances. Unlike NEB and string methods, which require the intermediate configurations to be distributed uniformly between the two ends, the splined saddle method does not restrict the distribution of the anchor points. This flexibility is very advantageous and provides the possibility of manipulating the anchor points to overcome challenges in the splined saddle method. An important trouble was found in our study of the splined saddle method for the 2D model potential, i.e., near-overlapping the high-energy portion of the pathway with an equipotential curve so that obtaining the maximum point can be inefficient and unreliable. The remedy to this problem, the preliminary NEB run, was introduced in Sec. IV B and was successfully examined for all the atomistic test cases in Sec. V. In the preliminary NEB run, we use only few number of images initially distributed adjacent to the two end points. In this way, the center of the pathway is initially kept empty of images and it will be gradually populated in the subsequent iterations provided a soft spring constant is used. This resolves the enduring problem which existed in the double-ended methods, i.e., the intermediate configurations in the LST are unreasonable such that the *ab initio* codes cannot converge typically in the wave function optimization part. This is in spirit similar to the growing string method.[3]

Another big advantage of the splined saddle method is that the movements of the anchor points depend only on the potential force at the maximum point not at the anchor points themselves. This allows to employ a hybrid scheme straightforwardly, i.e., low-accuracy calculations for finding the maximum point and high-accuracy only for the obtained maximum point. In our calculations we gained in average a factor of four better performance while using hybrid scheme. However, one can gain even more if uses different levels of theory, e.g., B3LYP as high-accuracy and LDA as low-accuracy calculations, provided the maximum point does not differ significantly in the two levels of the theory. Table II presents a detailed performance of the enhanced splined saddle method for the atomistic test systems. We used a different point set for each test system in accord with the complexity of the LST. However, we have also used a fixed point set $P = \{4, 4, 4\}$ for all of them for a fair performance evaluation of the enhanced splined saddle method. As a consequence, the method was less efficient for the simple reactions such as ammonia and ethane molecules and interestingly more efficient for Stone-Wales transformation in C_{60} fullerene. Table II shows that the performance of the enhanced splined saddle method is not very sensitive to the choice of the point set provided it has not been chosen completely unreasonable. Moreover, Table II shows that the enhanced splined saddle method is comparable to the CI-NEB method if the total number of gradient calls $ng_1 + ng_2$ is considered and it outperforms the CI-NEB method if the computational time (characterized by $\mathbf{ng_3}$) is considered.

Due to the complexity of the Jacobian in our implementation, the use of the splined saddle method is justifiable only with *ab initio* calculations. This is not a serious disadvantage of the method since the saddle point properties usually are of interests when they are treated with first principle methods. The difficulty with calculation of the Jacobian is due to our choice of natural cubic splines with parametrization based on inter-configuration distances. Though investigation on how the performance of the splined saddle method is affected by the different choices of splines and the parametrization can enhance the understanding of the splined saddle method.

VII. CONCLUSION

We used a two-dimensional model potential to enhance our understanding of the splined saddle method features. Based on the splined saddle method specifications, we could then introduce essential improvements which increase its efficiency and reliability. The enhanced splined saddle method was examined for a variety of test systems with increasing complexity ranging from simple isomerization of organic molecules to rolling an NaCl cubic cluster on an NaCl surface. The enhanced splined saddle method performance was remarkable in particular in the hybrid scheme. However the method was unsatisfactory without a preliminary NEB run that means the use of raw splined saddle method for realistic situations is not recommended. Furthermore, one problem (see Fig. 7 of Ref. 21) which was discussed in the original paper is still left without a resolution. The method in the current version is efficient, however, it needs to be improved with respect to reliability. Moreover, the effect of the spline and its parametrization needs to be investigated. The splined saddle method is complicated but since it is efficient it can be useful in the context of *ab inito* calculations. Our

Downloaded 16 Feb 2012 to 131.152.33.215. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

implementation of the splined saddle method will be available in the release version 1.5/1.6 of the BIGDFT (Ref. 29) code which is a free density functional theory package based on wavelets.

ACKNOWLEDGMENTS

Financial support was provided by the Swiss National Science Foundation. The calculations were done at the Swiss National Supercomputing center CSCS in Manno.

APPENDIX A: MATHEMATICAL DERIVATION OF JACOBIAN WHEN THE PATHWAY IS INTERPOLATED BY NATURAL CUBIC SPLINE

The pathway $\tilde{q}(t)$ passes through the $N_a + 1$ anchor points $Q_\alpha, \alpha = 0, \ldots, N_a$ so that $\tilde{q}(t_\alpha) = Q_\alpha$. Indices $i, j, \ldots$ are used as superscript for different coordinates and indices $\alpha, \beta, \ldots$ are used as subscript for different anchor points. The natural cubic spline within the interval $t \in [t_{\alpha-1}, t_\alpha]$, in which the function and its first and second derivatives are continuous, passing through $N_a + 1$ points is given by

$$q^i_\alpha(t) = \left(\frac{z^i_\alpha - z^i_{\alpha-1}}{h_\alpha}\right) t^3 + \frac{3\left(z^i_{\alpha-1}t_\alpha - z^i_\alpha t_{\alpha-1}\right)}{h_\alpha} t^2 + \left[\frac{3\left(z^i_\alpha t^2_{\alpha-1} - z^i_{\alpha-1}t^2_\alpha\right)}{h_\alpha} + \left(\frac{Q^i_\alpha}{h_\alpha} - h_\alpha z^i_\alpha\right) - \left(\frac{Q^i_{\alpha-1}}{h_\alpha} - h_\alpha z^i_{\alpha-1}\right)\right] t + \left[\frac{z^i_{\alpha-1}t^3_\alpha - z^i_\alpha t^3_{\alpha-1}}{h_\alpha} - \left(\frac{Q^i_\alpha}{h_\alpha} - h_\alpha z^i_\alpha\right) t_{\alpha-1} + \left(\frac{Q^i_{\alpha-1}}{h_\alpha} - h_\alpha z^i_{\alpha-1}\right) t_\alpha\right],$$

where $\alpha = 1, \ldots, N_a$ and $h_\alpha := t_\alpha - t_{\alpha-1}$. z_i's are obtained by solving a system of linear equations derived by the continuity of the first derivative at anchor points,

$$h_\alpha z^i_{\alpha-1} + 2(h_\alpha + h_{\alpha+1}) z^i_\alpha + h_{\alpha+1} z^i_{\alpha+1} = \frac{Q^i_{\alpha+1} - Q^i_\alpha}{h_{\alpha+1}} - \frac{Q^i_\alpha - Q^i_{\alpha-1}}{h_\alpha}, \tag{A1}$$

in which $\alpha = 1, \ldots, N_a - 1$ and $z^i_0 = z^i_{N_a} = 0$. In order to obtain the derivative of the function $\tilde{v}_{max}$ with respect to the anchor points one must calculate the following Jacobian:

$$\frac{\partial \tilde{v}_{max}}{\partial Q^i_\alpha} = \sum_{j=1}^{N} \frac{\partial v}{\partial q^j} J^{j,\alpha}_i, \quad \text{where } J^{j,\alpha}_i = \frac{\partial \tilde{q}^j(t_{max})}{\partial Q^i_\alpha},$$

let us assume $t_{max} \in [t_{\beta-1}, t_\beta]$, we thus have

$$\frac{\partial \tilde{q}^j(t_{max})}{\partial Q^i_\alpha} = \frac{\partial q^j_\beta(t_{max})}{\partial Q^i_\alpha},$$

which yields

$$\begin{aligned}
\frac{\partial q^j_\beta(t_{max})}{\partial Q^i_\alpha} = {} & \left[\frac{\frac{\partial z^j_\beta}{\partial Q^i_\alpha} - \frac{\partial z^j_{\beta-1}}{\partial Q^i_\alpha}}{h_\beta} - \frac{\left(z^j_\beta - z^j_{\beta-1}\right)\frac{\partial h_\beta}{\partial Q^i_\alpha}}{h^2_\beta}\right] t^3_{max} \\
& + 3\left[\frac{\frac{\partial z^j_{\beta-1}}{\partial Q^i_\alpha} t_\beta + z^j_{\beta-1}\frac{\partial t_\beta}{\partial Q^i_\alpha} - \frac{\partial z^j_\beta}{\partial Q^i_\alpha} t_{\beta-1} - z^j_\beta \frac{\partial t_{\beta-1}}{\partial Q^i_\alpha}}{h_\beta} - \frac{\left(z^j_{\beta-1}t_\beta - z^j_\beta t_{\beta-1}\right)\frac{\partial h_\beta}{\partial Q^i_\alpha}}{h^2_\beta}\right] t^2_{max} \\
& + \left[3\left(\frac{\frac{\partial z^j_\beta}{\partial Q^i_\alpha} t^2_{\beta-1} + 2z^j_\beta t_{\beta-1}\frac{\partial t_{\beta-1}}{\partial Q^i_\alpha} - \frac{\partial z^j_{\beta-1}}{\partial Q^i_\alpha} t^2_\beta - 2z^j_{\beta-1} t_\beta \frac{\partial t_\beta}{\partial Q^i_\alpha}}{h_\beta} - \frac{\left(z^j_\beta t^2_{\beta-1} - z^j_{\beta-1} t^2_\beta\right)\frac{\partial h_\beta}{\partial Q^i_\alpha}}{h^2_\beta}\right)\right. \\
& + \left(\frac{\delta_{ij}\delta_{\alpha\beta}}{h_\beta} - \frac{Q^j_\beta \frac{\partial h_\beta}{\partial Q^i_\alpha}}{h^2_\beta} - z^j_\beta \frac{\partial h_\beta}{\partial Q^i_\alpha} - h_\beta \frac{\partial z^j_\beta}{\partial Q^i_\alpha}\right) \\
& \left. - \left(\frac{\delta_{ij}\delta_{\alpha,\beta-1}}{h_\beta} - \frac{Q^j_{\beta-1}\frac{\partial h_\beta}{\partial Q^i_\alpha}}{h^2_\beta} - z^j_{\beta-1}\frac{\partial h_\beta}{\partial Q^i_\alpha} - h_\beta \frac{\partial z^j_{\beta-1}}{\partial Q^i_\alpha}\right)\right] t_{max} \\
& + \left[\frac{\frac{\partial z^j_{\beta-1}}{\partial Q^i_\alpha} t^3_\beta + 3z^j_{\beta-1} t^2_\beta \frac{\partial t_\beta}{\partial Q^i_\alpha} - \frac{\partial z^j_\beta}{\partial Q^i_\alpha} t^3_{\beta-1} - 3z^j_\beta t^2_{\beta-1}\frac{\partial t_{\beta-1}}{\partial Q^i_\alpha}}{h_\beta} - \frac{\left(z^j_{\beta-1}t^3_\beta - z^j_\beta t^3_{\beta-1}\right)\frac{\partial h_\beta}{\partial Q^i_\alpha}}{h^2_\beta}\right.
\end{aligned}$$

Downloaded 16 Feb 2012 to 131.152.33.215. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

$$- \left(\frac{\delta_{ij}\delta_{\alpha\beta}}{h_\beta} - \frac{Q^j_\beta \frac{\partial h_\beta}{\partial Q^i_\alpha}}{h_\beta^2} - z^j_\beta \frac{\partial h_\beta}{\partial Q^i_\alpha} - h_\beta \frac{\partial z^j_\beta}{\partial Q^i_\alpha} \right) t_{\beta-1} - \left(\frac{Q^j_\beta}{h_\beta} - h_\beta z^j_\beta \right) \frac{\partial t_{\beta-1}}{\partial Q^i_\alpha}$$

$$+ \left(\frac{\delta_{ij}\delta_{\alpha,\beta-1}}{h_\beta} - \frac{Q^j_{\beta-1} \frac{\partial h_\beta}{\partial Q^i_\alpha}}{h_\beta^2} - z^j_{\beta-1} \frac{\partial h_\beta}{\partial Q^i_\alpha} - h_\beta \frac{\partial z^j_{\beta-1}}{\partial Q^i_\alpha} \right) t_\beta + \left(\frac{Q^j_{\beta-1}}{h_\beta} - h_\beta z^j_{\beta-1} \right) \frac{\partial t_\beta}{\partial Q^i_\alpha} \Bigg]. \tag{A2}$$

Below we present the mathematical derivation of all terms in equation above in details. We used a similar parametrization as given in Ref. 8,

$$h_\beta = \sqrt{\sum_{k=1}^{N} \left(Q^k_\beta - Q^k_{\beta-1}\right)^2}.$$

Therefore,

$$\frac{\partial h_\beta}{\partial Q^i_\alpha} = \frac{\left(Q^i_\beta - Q^i_{\beta-1}\right)(\delta_{\alpha\beta} - \delta_{\alpha,\beta-1})}{h_\beta}.$$

The pathway parameter can also be rewritten in terms of h_μ's as $t_\beta = \sum_{\mu=1}^{\beta} h_\mu$, therefore its derivative with respect to the anchor points is given by

$$\frac{\partial t_\beta}{\partial Q^i_\alpha} = \sum_{\mu=1}^{\beta} \frac{\partial h_\mu}{\partial Q^i_\alpha} = \sum_{\mu=1}^{\beta} \frac{\left(Q^i_\mu - Q^i_{\mu-1}\right)(\delta_{\alpha\mu} - \delta_{\alpha,\mu-1})}{h_\mu}. \tag{A3}$$

Applying the summation in Eq. (A3), depending on α and β, it reduces to

$$\frac{\partial t_\beta}{\partial Q^i_\alpha} = \begin{cases} \dfrac{(Q^i_\alpha - Q^i_{\alpha-1})}{h_\alpha} - \dfrac{(Q^i_{\alpha+1} - Q^i_\alpha)}{h_{\alpha+1}} & \alpha < \beta \\ \dfrac{(Q^i_\alpha - Q^i_{\alpha-1})}{h_\alpha} & \alpha = \beta. \\ 0 & \alpha > \beta \end{cases}$$

In Eq. (A2) the derivatives of the z^j_β with respect to the coordinates of the anchor points, $\frac{\partial z^j_\beta}{\partial Q^i_\alpha}$ must be calculated for each α, more precisely we must solve a different system of linear equations for each α. Assuming the system of linear equations is given by

$$A\mathbf{z}^j = \mathbf{g}^j \quad \text{where} \quad \begin{cases} \mathbf{z}^j = \left(z^j_1, \ldots, z^j_{N_a-1}\right) \\ \mathbf{g}^j_\beta = \dfrac{Q^j_{\beta+1} - Q^j_\beta}{h_{\beta+1}} - \dfrac{Q^j_\beta - Q^j_{\beta-1}}{h_\beta}, \end{cases}$$

where A is a tridiagonal symmetric matrix made up with coefficients of z^i_α's in Eq. (A1). Since both z^i_α's and matrix elements of A depend on the coordinates of the anchor points, $\frac{\partial z^j_\beta}{\partial Q^i_\alpha}$ has to be calculated as

$$\frac{\partial \mathbf{z}^j}{\partial Q^i_\alpha} = A^{-1} \frac{\partial \mathbf{g}^j}{\partial Q^i_\alpha} + \left(\frac{\partial A^{-1}}{\partial Q^i_\alpha} \right) \mathbf{g}^j,$$

that is,

$$\frac{\partial z^j_\beta}{\partial Q^i_\alpha} = \sum_{\omega=1}^{N_a-1} (A^{-1})_{\beta\omega} \frac{\partial g^j_\omega}{\partial Q^i_\alpha} + \sum_{\omega=1}^{N_a-1} \left(\frac{\partial A^{-1}}{\partial Q^i_\alpha} \right)_{\beta\omega} g^j_\omega, \tag{A4}$$

$\frac{\partial g^j_\omega}{\partial Q^i_\alpha}$ is calculated straightforwardly and is given by

$$\frac{\partial g^j_\omega}{\partial Q^i_\alpha} = \Bigg[\frac{\delta_{ij}\delta_{\omega+1,\alpha}}{h_{\omega+1}} - \delta_{ij}\delta_{\omega\alpha} \left(\frac{1}{h_{\omega+1}} + \frac{1}{h_\omega} \right) + \frac{\delta_{ij}\delta_{\omega-1,\alpha}}{h_\omega} - \frac{(Q^j_{\omega+1} - Q^j_\omega) \frac{\partial h_{\omega+1}}{\partial Q^i_\alpha}}{h^2_{\omega+1}} + \frac{(Q^j_\omega - Q^j_{\omega-1}) \frac{\partial h_\omega}{\partial Q^i_\alpha}}{h^2_\omega} \Bigg].$$

The summation of the first term in Eq. (A4) can be evaluated in terms of A^{-1}, which yields

$$\sum_{\omega=1}^{N_a-1} (A^{-1})_{\beta\omega} \frac{\partial g^j_\omega}{\partial Q^i_\alpha} = (A^{-1})_{\beta,\alpha-1} \left(\frac{\delta_{ij}}{h_\alpha} - \frac{(Q^j_\alpha - Q^j_{\alpha-1})(Q^i_\alpha - Q^i_{\alpha-1})}{h^3_\alpha} \right) (1 - \delta_{\alpha,1})$$
$$- (A^{-1})_{\beta\alpha} \left(\frac{\delta_{ij}}{h_{\alpha+1}} + \frac{\delta_{ij}}{h_\alpha} - \frac{(Q^j_\alpha - Q^j_{\alpha-1})(Q^i_\alpha - Q^i_{\alpha-1})}{h^3_\alpha} - \frac{(Q^j_{\alpha+1} - Q^j_\alpha)(Q^i_{\alpha+1} - Q^i_\alpha)}{h^3_{\alpha+1}} \right)$$
$$+ (A^{-1})_{\beta,\alpha+1} \left(\frac{\delta_{ij}}{h_{\alpha+1}} - \frac{(Q^j_{\alpha+1} - Q^j_\alpha)(Q^i_{\alpha+1} - Q^i_\alpha)}{h^3_{\alpha+1}} \right) (1 - \delta_{\alpha,N_a-1}).$$

Notice that for all $\alpha, \beta = 1, \ldots, N_a - 1$ we have

$$\frac{\partial z^i_0}{\partial Q^i_\alpha} = \frac{\partial z^i_{N_a}}{\partial Q^i_\alpha} = 0.$$

In order to calculate the derivative of matrix elements of A^{-1} with respect to the Q^i_α's, we use the following identity:

$$AA^{-1} = \mathbf{I} \implies \frac{\partial A}{\partial Q^i_\alpha} A^{-1} + A \frac{\partial A^{-1}}{\partial Q^i_\alpha} = 0,$$

Downloaded 16 Feb 2012 to 131.152.33.215. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

which yields

$$\frac{\partial A^{-1}}{\partial Q^i_\alpha} = -A^{-1}\frac{\partial A}{\partial Q^i_\alpha}A^{-1}.$$

Therefore, the summation of the second term in Eq. (A4) can be calculated as

$$\left[\frac{\partial A^{-1}}{\partial Q^i_\alpha}\vec{g}^j\right]_\beta = \sum_{\omega=1}^{N_a-1}\left(\frac{\partial A^{-1}}{\partial Q^i_\alpha}\right)_{\beta\omega} g^j_\omega = -\sum_{\omega,\mu,\nu=1}^{N_a-1}(A^{-1})_{\beta\mu}\left(\frac{\partial A}{\partial Q^i_\alpha}\right)_{\mu\nu}(A^{-1})_{\nu\omega}\, g^j_\omega, \tag{A5}$$

where the derivative of matrix elements A with respect to Q^i_α's must be calculated. Using Kronecker delta, the elements of matrix A can be written as

$$A_{\mu\nu} = h_\mu\delta_{\mu-1,\nu} + 2(h_\mu + h_{\mu+1})\delta_{\mu,\nu} + h_{\mu+1}\delta_{\mu+1,\nu},$$

and its derivative is thus given by

$$\frac{\partial A_{\mu\nu}}{\partial Q^i_\alpha} = \frac{\partial h_\mu}{\partial Q^i_\alpha}\delta_{\mu-1,\nu} + 2\left(\frac{\partial h_\mu}{\partial Q^i_\alpha} + \frac{\partial h_{\mu+1}}{\partial Q^i_\alpha}\right)\delta_{\mu,\nu} + \frac{\partial h_{\mu+1}}{\partial Q^i_\alpha}\delta_{\mu+1,\nu}. \tag{A6}$$

Inserting Eq. (A6) in Eq. (A5), the matrix elements of derivative of A^{-1} with respect to coordinates of anchor points is given by

$$\begin{aligned}\left(\frac{\partial A^{-1}}{\partial Q^i_\alpha}\right)_{\beta\omega} = -\Bigg[&(A^{-1})_{\beta\alpha}\frac{(Q^i_\alpha - Q^i_{\alpha-1})}{h_\alpha}(1-\delta_{\alpha,1})(A^{-1})_{\alpha-1,\omega}\\ &-(A^{-1})_{\beta,\alpha+1}\frac{(Q^i_{\alpha+1} - Q^i_\alpha)}{h_{\alpha+1}}(1-\delta_{\alpha,N_a-1})(A^{-1})_{\alpha,\omega}\\ &+2(A^{-1})_{\beta\alpha}\frac{(Q^i_\alpha - Q^i_{\alpha-1})}{h_\alpha}(A^{-1})_{\alpha,\omega}\\ &-2(A^{-1})_{\beta,\alpha+1}\frac{(Q^i_{\alpha+1} - Q^i_\alpha)}{h_{\alpha+1}}(1-\delta_{\alpha,N_a-1})(A^{-1})_{\alpha+1,\omega}\\ &+2(A^{-1})_{\beta,\alpha-1}\frac{(Q^i_\alpha - Q^i_{\alpha-1})}{h_\alpha}(1-\delta_{\alpha,1})(A^{-1})_{\alpha-1,\omega}\\ &-2(A^{-1})_{\beta,\alpha}\frac{(Q^i_{\alpha+1} - Q^i_\alpha)}{h_{\alpha+1}}(A^{-1})_{\alpha,\omega}\\ &+(A^{-1})_{\beta,\alpha-1}\frac{(Q^i_\alpha - Q^i_{\alpha-1})}{h_\alpha}(1-\delta_{\alpha,1})(A^{-1})_{\alpha,\omega}\\ &-(A^{-1})_{\beta,\alpha}\frac{(Q^i_{\alpha+1} - Q^i_\alpha)}{h_{\alpha+1}}(1-\delta_{\alpha,N_a-1})(A^{-1})_{\alpha+1,\omega}\Bigg].\end{aligned}$$

Now all the terms in Eq. (A2) are known and the Jacobian can be calculated. Consequently, the gradient of the function $\tilde{v}_{max}(Q_1,\ldots,Q_{N_a-1})$ is known as well.

APPENDIX B: 2D MODEL POTENTIAL

The straight line $y = 0$ is also a solution of the differential equation for MEPs, while it passes the origin which is an undefined point for this model potential. For this reason, one should avoid using a linear interpolation of the two minima as the starting pathway. In our calculations with $N_a - 1 = 2$, we limited the initial random interior anchor points to be in two squares with side length equals to 1/2 one centered at $(-1/2, 1/2)$ and the other at $(1/2, 1/2)$. In the calculations with $N_a - 1 = 1$, the single anchor point is randomly chosen restricted to any of the two squares mentioned above.

APPENDIX C: QUINTIC SPLINE INTERPOLATION

Consider the values of a function and its derivatives at $N+1$ points, $(t_i, v_i, \dot{v}_i)$, $i = 0, 1, 2, \ldots, N$, are given. A quintic spline interpolation is a sequence piecewise fifth-order polynomials. We thus have N piecewise polynomials which totally have $6N$ unknown coefficients, so that $6N$ independent equations are required to calculate these coefficients. The $6N$ equations are obtained as following: (i) The function and its first derivative at $N+1$ points are assumed to be given, (ii) the function value and its first, second, and third derivatives at $N-1$ interior points must be continuous. The conditions (i) and (ii) sum up $6N-2$ equations, in our implementation two more equations are obtained by putting the third derivative at both end points equal to zero. However, different conditions can be used. The ith piece of the spline $v^{(i)}(t)$ associated with the interval $[t_{i-1}, t_i]$ is given by Eq. (C1) which ensures that the spline function satisfies value v_i and its first derivative $\dot{v}_i$, thereby the continuity of the spline and also its first derivative at interior points,

$$\begin{aligned}v^{(i)}(t) = &\frac{a_i(t-t_{i-1})^5 + b_i(t_i - t)^5}{h_i^3}\\ &+\frac{c_i(t-t_{i-1})^3 + d_i(t_i-t)^3}{h_i}\\ &+e_i(t-t_{i-1}) + f_i(t_i - t),\end{aligned} \tag{C1}$$

where $h_i = t_i - t_{i-1}$ and c_i, d_i, e_i, and f_i are

$$\begin{aligned}c_i &= \frac{1}{3}(2b_i - 7a_i) + \frac{\dot{v}_{i-1} + 2\dot{v}_i}{3h_i} + \frac{v_{i-1} - v_i}{h_i^2},\\ d_i &= \frac{1}{3}(2a_i - 7b_i) - \frac{2\dot{v}_{i-1} + \dot{v}_i}{3h_i} + \frac{v_i - v_{i-1}}{h_i^2},\\ e_i &= \frac{2}{3}(2a_i - b_i)h_i - \frac{1}{3}(\dot{v}_{i-1} + 2\dot{v}_i) + \frac{2v_i - v_{i-1}}{h_i},\\ f_i &= \frac{2}{3}(2b_i - a_i)h_i + \frac{1}{3}(2\dot{v}_{i-1} + \dot{v}_i) + \frac{2v_{i-1} - v_i}{h_i}.\end{aligned}$$

Downloaded 16 Feb 2012 to 131.152.33.215. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

The conditions of continuity of the second and the third derivative yield $N-1$ couple equations given by

$$3a_i + 2b_i - 2a_{i+1} - 3b_{i+1} = -\frac{2\dot{v}_i + \dot{v}_{i+1}}{h_{i+1}} - \frac{3(v_i - v_{i+1})}{h_{i+1}^2} - \frac{\dot{v}_{i-1} + 2\dot{v}_i}{h_i} - \frac{3(v_{i-1} - v_i)}{h_i^2},$$

$$7h_{i+1}a_i + 3h_{i+1}b_i + 3h_i a_{i+1} + 7h_i b_{i+1} = \frac{h_i}{h_{i+1}}(\dot{v}_i + \dot{v}_{i+1}) - \frac{h_{i+1}}{h_i}(\dot{v}_{i-1} + \dot{v}_i) + \frac{2h_i}{h_{i+1}^2}(v_i - v_{i+1}) + \frac{2h_{i+1}}{h_i^2}(v_i - v_{i-1}).$$

[1] F. Jensen, *Introduction to Computational Chemistry* (Wiley, New York, 2006).
[2] A. T. Bell, Mol. Phys. **102**, 319 (2004).
[3] B. Peters, A. Heyden, A. T. Bell, and A. Chakraborty, J. Chem. Phys. **120**, 7877 (2004).
[4] A. Heyden, A. T. Bell, and F. J. Keil, J. Chem. Phys. **123**, 224101 (2005).
[5] H. Jónsson, G. Mills, and K. W. Jacobsen, *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne, G. Ciccotti, and D. F. Coker (World Scientific, Singapore, 1998), p. 385404.
[6] G. Henkelman and H. Jónsson, J. Chem. Phys. **113**, 9978 (2000).
[7] W. E, W. Ren, and E. Vanden-Eijnden, Phys. Rev. B **66**, 052301 (2002).
[8] W. E, W. Ren, and E. Vanden-Eijnden, J. Chem. Phys. **126**, 164103 (2007).
[9] G. Henkelman and H. Jónsson, J. Chem. Phys. **111**, 7010 (1999).
[10] R. A. Olsen, G. J. Kroes, G. Henkelman, and A. Arnaldsson, J. Chem. Phys. **121**, 9776 (2004).
[11] A. Poddey and P. E. Blöchl, J. Chem. Phys. **128**, 044107 (2008).
[12] J. Kästner and P. Sherwood, J. Chem. Phys. **128**, 014106 (2008).
[13] C. Lanczos, *Applied Analysis* (Dover, New York, 1988).
[14] R. Malek and N. Mousseau, Phys. Rev. E **62**, 7723 (2000).
[15] O. F. Sankey, D. A. Drabold, and A. Gibson, Phys. Rev. B **50**, 1376 (1994).
[16] G. Henkelman, B. P. Uberuaga, and H. Jónsson, J. Chem. Phys. **113**, 9901 (2000).
[17] D. Sheppard, R. Terrell, and G. Henkelman, J. Chem. Phys. **128**, 134106 (2008).
[18] D. J. Wales, *Energy Landscapes* (Cambridge University Press, Cambridge, England, 2003).
[19] E. F. Koslover and D. J. Wales, J. Chem. Phys. **127**, 134102 (2007).
[20] J. Klimes, D. R. Bowler, and A. Michaelides, J. Phys.: Condens. Matter **22**, 074203 (2010).
[21] R. Granot and R. Baer, J. Chem. Phys. **128**, 184111 (2008).
[22] T. A. Halgren and W. N. Lipscomb, Chem. Phys. Lett. **49**, 225 (1977).
[23] E. Bitzek, P. Koskinen, F. Gahler, M. Moseler, and P. Gumbsch, Phys. Rev. Lett. **97**, 170201 (2006).
[24] J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer-Verlag, New York, 1999).
[25] A. Goodrow, A. T. Bell, and M. Head-Gordon, J. Chem. Phys. **129**, 174109 (2008).
[26] A. Goodrow, A. T. Bell, and M. Head-Gordon, J. Chem. Phys. **130**, 244108 (2009).
[27] J. C. Grossman and L. Mitas, Phys. Rev. Lett. **79**, 4353 (1997).
[28] P. Maragakis, S. A. Andreev, Y. Brumer, D. R. Reichman, and E. Kaxiras, J. Chem. Phys. **117**, 4651 (2002).
[29] L. Genovese, A. Neelov, S. Goedecker, T. Deutsch, S. A. Ghasemi, A. Willand, D. Caliste, O. Zilberberg, M. Rayson, A. Bergman, and R. Schneider, J. Chem. Phys. **129**, 014109 (2008).
[30] C. Hartwigsen, S. Goedecker, and J. Hutter, Phys. Rev. B **58**, 3641 (1998).
[31] Y. Kumeda and D. J. Wales, Chem. Phys. Lett. **374**, 125 (2003).

Downloaded 16 Feb 2012 to 131.152.33.215. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions