Adaptive nudged elastic band approach for transition state calculation

P. Maragakis[a)]
Department of Physics and Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138

Stefan A. Andreev, Yisroel Brumer, and David R. Reichman
Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

Efthimios Kaxiras
Department of Physics and Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138

(Received 26 March 2002; accepted 31 May 2002)

We present a method for the location of transition states in complicated physical systems. Our algorithm is a variation of the well-established nudged elastic band method and leads to significant improvements in efficiency and accuracy. We assess the applicability of our method by testing it on several systems of practical interest representing a variety of physical situations. At the molecular level, we apply the method to tautomerization processes in nucleic acid bases and the double proton transfer in nucleic acid base pairs. For bulk systems, we considered the concerted exchange mechanism in Si, which is a complicated pathway for defect-free diffusion in the diamond lattice. For surface systems, we considered ad-dimer diffusion mechanisms on Si(100). We incorporated the climbing image extension of the nudged elastic band method and compared it against the original approach on two-dimensional model potential energy surfaces. Based on favorable comparisons with related methods and the general implementation of our method, we believe that this is well suited for efficient estimates of activation barriers with sophisticated electronic structure codes.
 [DOI: 10.1063/1.1495401]

I. INTRODUCTION

An important problem in condensed matter physics and theoretical chemistry is the understanding of transition processes such as chemical reactions or diffusion events. Within the Born–Oppenheimer approximation, the atoms comprising the system of interest move along a continuous path between two metastable states. The system is more likely to move along the paths with the lowest intermediate free energy maximum, and especially the one that minimizes the total action along the path. A path of practical interest, which does not correspond to the path of lowest free energy, is the path with the lowest intermediate maximum that minimizes the contour integral of the total energy; this is the so-called minimum energy path (MEP). The maximum along this path is a first-order saddle point in the potential energy surface. The location of this saddle point is an important quantity in the study of transition processes; in fact, many approximate methods only use local information around this saddle point and the two end points of the transition to predict observables such as reaction rates. Following the pioneering work of Elber and Karplus[1] and Pratt,[2] a variety of methods have been developed and are currently in use for the location of transition paths as well as a method for harvesting *dynamical* pathways.[3] Arguably, the most efficient method for finding an approximate MEP is the nudged elastic band (NEB) method,[4,5] which has been used extensively in a broad range of chemical reactions, proton transfer, and diffusion processes. The NEB method locates a transition path that minimizes the contour integral of the total energy and lies in the neighborhood of an initial guess. Although this method only locates local paths, sampling of these paths, for example through parallel tempering, provides useful information on the possible transition paths.[6] The NEB method has been used with several *ab initio* and empirical electronic structure codes. Recent enhancements of the method have improved the accuracy of the saddle point estimate[7] and the convergence of the transition path in special situations.[8]

In this paper we discuss a variation of the original NEB method and the climbing image NEB (CI-NEB),[7] that converges quickly to the proximity of the saddle point. Because of the efficiency compared to standard NEB and CI-NEB, we believe our approach will be useful in the study of complicated transition processes with more accurate and computationally demanding electronic structure methods than has been possible up to now.

We have selected representative examples to demonstrate the wide applicability of the new method. These include hydrogen transfers in small biomolecules, an exchange process in bulk Si,[9] and a diffusion process on the Si (100) surface.[10] Finally we have tested the method on typical model potentials including the Müller–Brown potential. The paper is organized as follows: We present the method in Sec. II and we demonstrate it on a typical model potential. In Sec. III we present selected applications of the method to problems that are typical of state-of-the-art applications in physical chemistry, condensed matter physics, and surface science.

a)Electronic mail: maragakis@cmt.harvard.edu

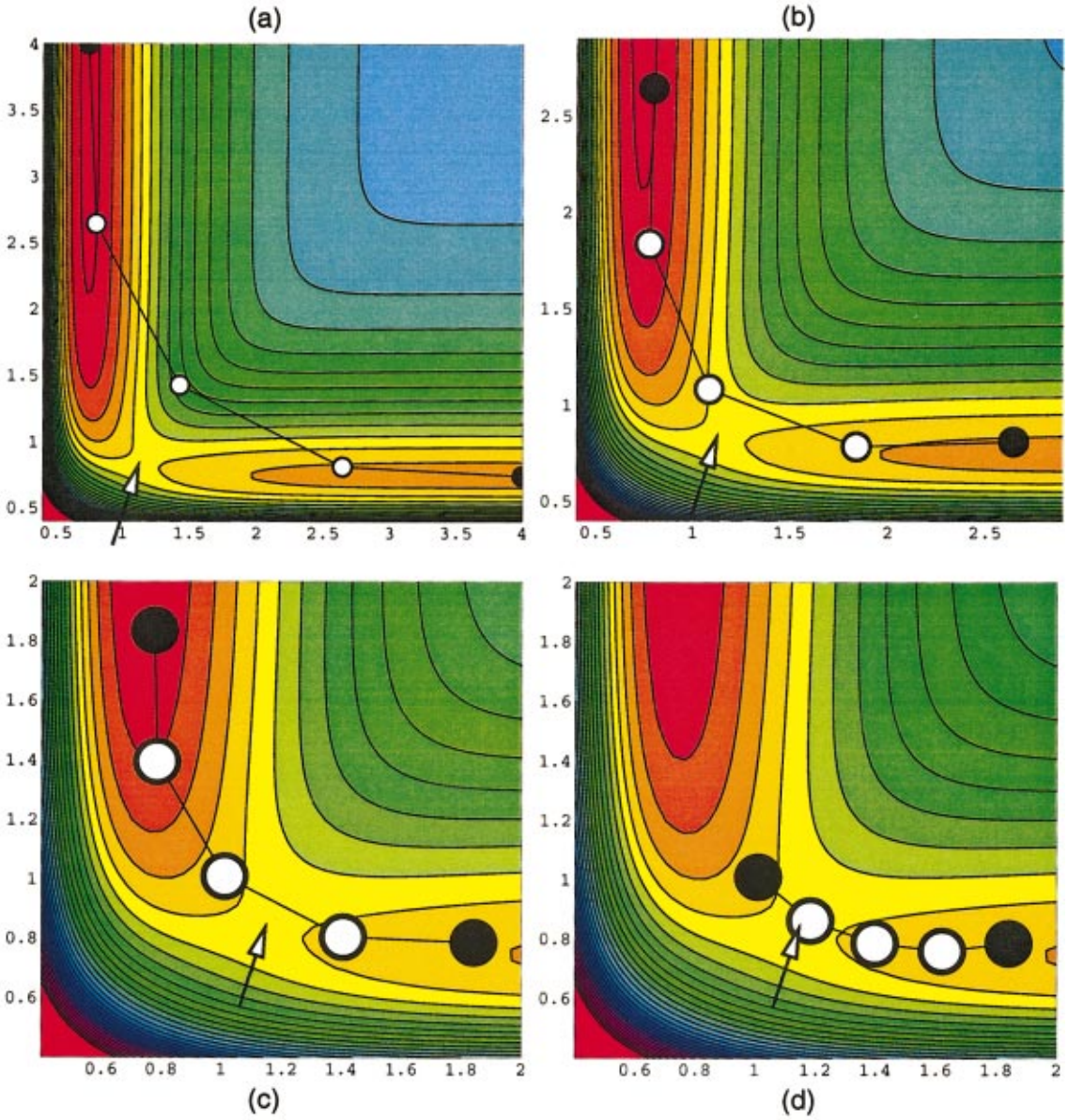

FIG. 1. (Color) Successive steps of the adaptive search of the saddle point in a simple model potential in two dimensions, for which the saddle point is exactly known. The closed circles stand for the fixed end-point images at each iteration step, and the open circles show the final location of the moving images. The arrow indicates the exact location of the saddle point.

In Sec. IV we incorporate the CI-NEB approach and compare it with our method on two model potentials.

II. METHOD

According to the classification of Jensen,[11] modern methods for transition state location can be separated into two main categories: those based on interpolating between two local minima and others that only use local information. In the former category, a specific process is usually studied, thus the end points of the transition are predetermined. Methods of the latter category usually start from a given local minimum and follow the normal modes of the system up to a saddle point. A method of this type is sometimes more efficient in determining a saddle point but there is no guarantee that the saddle point corresponds to a given transition. These methods are best suited for more thorough studies of the configuration space around a local minimum. Further, most of them depend on the Hessian or some approximation of it and thus exhibit poor scaling properties with the size of the system.

A subcategory of the interpolation methods are variations of the simple elastic band method in which the trajectory of the system is described by a set of images connecting two local minima; these methods are reviewed in Ref. 5. Each image has a mass and two springs connecting it to its neighboring images. On a flat energy landscape the springs would equipartition the images along the trajectory. On many transition processes the springs need careful tuning: if they are too soft the images slide down the energy hills failing to sample the saddle point, whereas if they are too hard they alter significantly the path that the system would choose. An important advancement in this class of methods is the

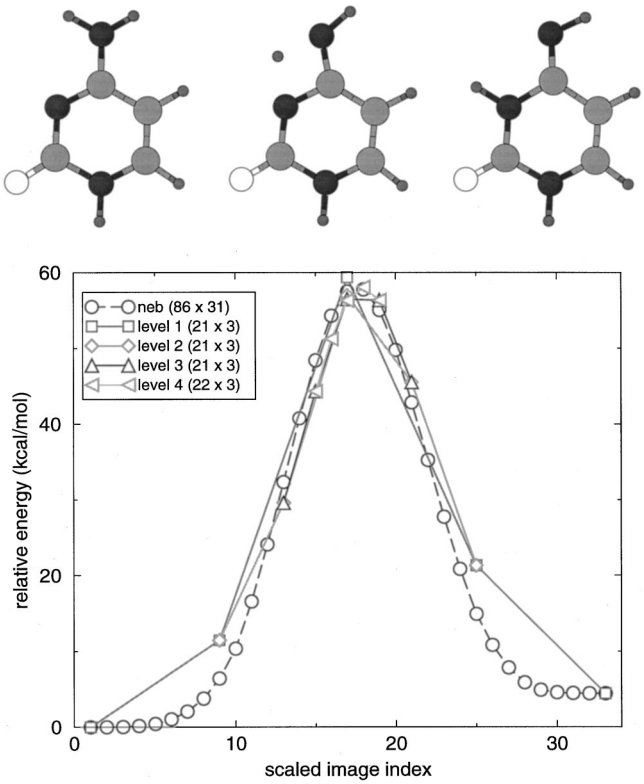

FIG. 2. Location of the saddle point in the proton transfer process in the cytosine nucleic acid base. At the top we show the initial keto state, the saddle point intermediate, and the final state. At the bottom we show the energy profile for the gradual levels of relaxation of the adaptive algorithm, together with the NEB energy path profile. We list the number of total energy and force evaluations for each step in the label of the graph.

nudged elastic band (NEB) method, which flattens the energy landscape along the current path by removing the force components due to the internal degrees of freedom and which avoids the disturbance of the energy landscape introduced due to the springs by removing the off-path spring forces. However, due to the projections involved, it is not a simple matter to determine an accurate Lagrangian for the discretized system and thus the use of powerful minimization methods is inhibited. On the other hand, a minimization based on plain Newtonian dynamics works well: The method proves stable with respect to the extra parameters introduced by the springs and in many typical cases the path gives a reasonably good estimate of the MEP with $\sim$50 steps per image along the path (this rises to a few hundreds of steps if one uses more stringent convergence criteria). In order to achieve the necessary density of images around the transition state, the initial path is sampled with many images, typically $\sim$20–30, although it is not uncommon to go up to 50 images. The total number of force evaluations depends on the system and on the required accuracy; typical cases require of the order of a few thousand force evaluations. While this is perfectly acceptable when using fast qualitative methods for describing the system, it requires large amounts of supercomputer time when used in conjunction with accurate *ab initio* methods. The climbing image nudged elastic band (CI-NEB) method[7] is an improvement of the original algorithm in which one of the images climbs up to the exact saddle point. This works by locally inverting the potential of the highest energy point along the current path estimate. This approach eliminates the need for fine sampling to obtain an accurate estimate of the saddle point and keeps the number of path relaxation steps at the same order as the NEB.

We propose a variation of the NEB method, which we call the adaptive nudged elastic band approach (ANEBA), designed to improve the efficiency and accuracy of the original NEB method. Instead of choosing a large number of images to bracket the saddle point with high accuracy, we adaptively increase the resolution in the neighborhood of the saddle point. Specifically, we start with three movable images connecting two local minima and use the NEB method as a starting point. After the method converges to some given accuracy (i.e., the forces perpendicular to the path are relatively small, but not necessarily as small as we would like them to be in the saddle point region), we consider the energy of the three movable and the two end-point images and choose the two images adjacent to the one that has the highest energy as our new starting points for the NEB method. We then repeat the above-mentioned procedure; this usually leads to a bracketing of the region that contains the saddle point. Each iteration roughly corresponds to using a full NEB method with double the number of images, thus after four iterations we have the equivalent of 33 images (31 movable images) in a NEB method, while we only needed to relax 12 images. In Fig. 1 we demonstrate the convergence properties of the method in a prototypical two-dimensional potential. This potential is model I, of Appendix A.1 in Ref. 5; it mimics a reaction involving three atoms confined to motion along a line with the coordinates being the distance between two pairs of atoms. In Fig. 1 we show the final state of each NEB iteration during the ANEBA search. If we count the number of images used, we see that this simple example hints to a gain in efficiency of about a factor of 2.6, but in practice the gain can be higher for various reasons: the accuracy for the first few iterations need not be high, while at the same time the convergence of later iterations is more uniform and efficient due to the localized nature of the problem.

Although this simple version of the method works quite well for some cases, it may fail on a complicated energy land scape. The simplest failure mode occurs when the maximum energy is one of the two end points at a given level of recursion. In this case we step back one level of recursion and obtain a finer sampling on the problematic sides of the energy landscape. Note that, despite the fact that in all the examples that we tested the method seemed to perform very well, there is no general guarantee that in its present form it will not fail. Conceivably, one could extend the adaptive scheme to the full path: One only needs to minimize the remaining sections to a level that guarantees there is no higher saddle point lying in between.

The above-described algorithm has been interfaced with several total energy codes, such as CHARMM,[12] VASP,[13] and SCC-DFTB.[14] For the examples that we present in the following section we have interfaced ANEBA with the self-consistent charge density-functional based tight-binding (SCC-DFTB) method for calculating total energies and

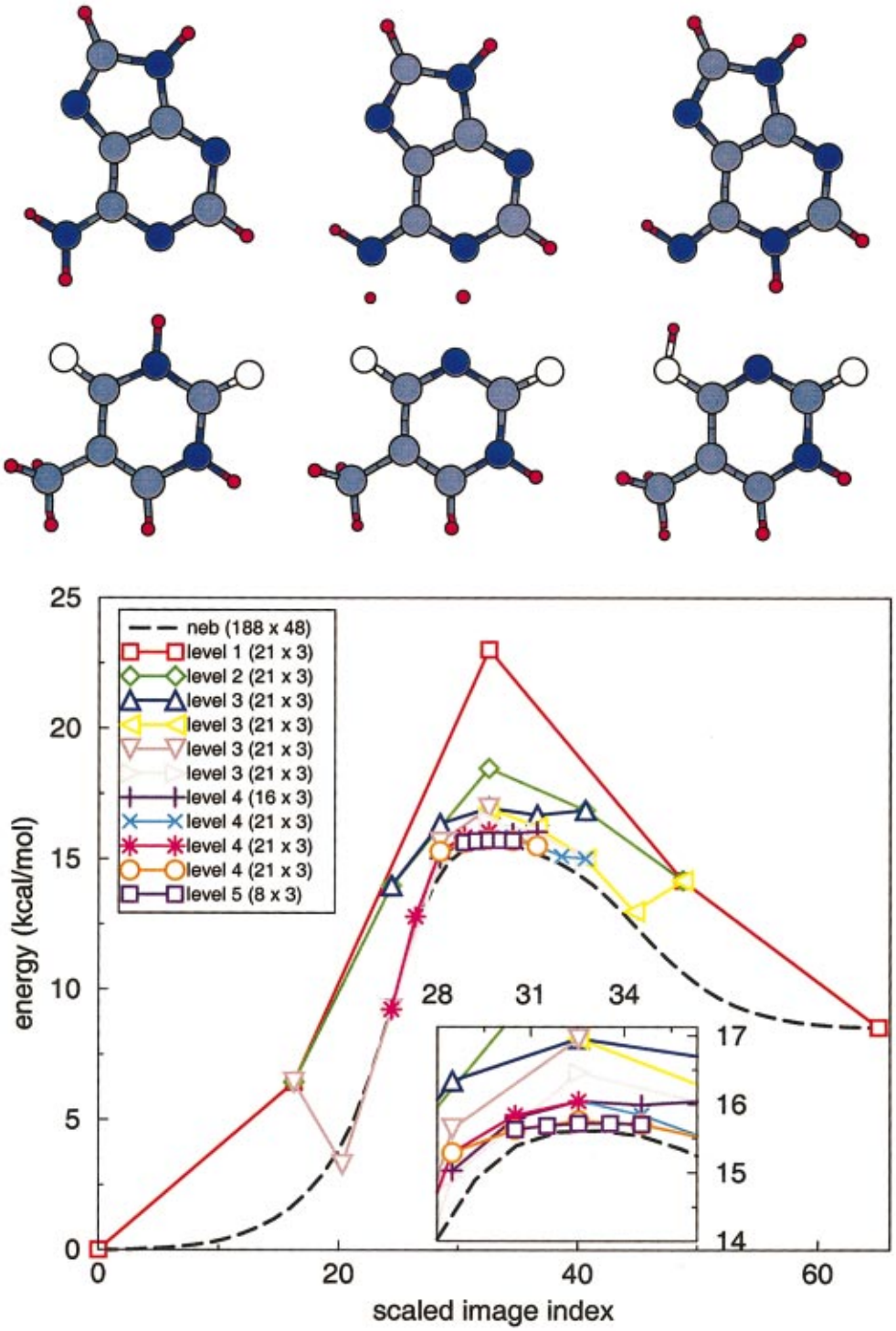

FIG. 3. (Color) Location of the saddle point in the double proton transfer process in the AT nucleic acid base pair. At the top we show the initial, intermediate saddle point and final configurations. The inset shows a detailed view of the saddle point region.

forces.[14] This method has been successfully applied to a range of physical systems spanning semiconductor materials and biologically relevant molecules and is sufficiently fast enough to be used in large-scale calculations.

III. APPLICATIONS

A. Nucleic acid bases tautomerization

We test the method in the simplest molecular setting, the study of the tautomerization process of nucleic acid bases. Varying the location of a single hydrogen leads to different stable configurations of the nucleic acid. Typically, there exist several stable configurations within 10 kcal/mol of the ground state. To obtain an estimate for the tautomerization rates we need to obtain the barrier of the transition. It turns out that for the molecules in the gas phase these barriers are very high (typically higher than 40 kcal/mol), so that the transition rates are negligibly small and the tautomers are very stable without the presence of a catalyst such as water.

In Fig. 2 we present the cytosine tautomerization path convergence for ANEBA. The NEB method calculation was relaxed until the maximal force perpendicular to the path was smaller than 0.1 eV/Å; a total of 86 iterations were needed, corresponding to 2668 total-energy and forces evaluations. At that point the image with the highest energy had a maximal force of 0.64 eV/Å. This happens because the true saddle point lies between the two images with the highest energy. Out of all the nucleic acid tautomerization processes this is magnified in the cytosine case. One can obtain a better estimate of the true saddle point by interpolating the NEB path. The ANEBA very quickly converges to the correct saddle point region. We have used a cutoff of 21 moves per level for all but the last level of the algorithm. The final saddle point image has a maximal force of 0.1 eV/Å, and is located in a total of 257 total-energy and forces calculations, an improvement by an order of magnitude compared to the NEB calculation.

We next present a calculation of the adenine-thymine base pair double proton transfer as an example of a more complicated transition path. When the nucleic acid bases pair in their common Watson–Crick conformations, spontaneous tautomerization may occur by a double proton transfer between two opposite bases, a process that may lead to mutations.[15] Nowadays this mechanism is not considered to be one of the major pathways leading to errors in the genetic code.

In Fig. 3 we show the convergence characteristics of the ANEBA method. For this example this path proves to be slowly converging. The original guess of a linear interpolation path between initial and final state, which corresponds to a coherent double proton exchange, quickly changes to an almost sequential proton transfer with the thymine proton leaving first. Note that this process has a very low saddle point frequency ($\sim$200 cm^{-1}). We relax the NEB method up to the point where the maximum perpendicular force on any atom is smaller than 0.05 eV/Å. The convergence in this case takes 188 steps for a run with 50 images (48 movable ones), or a total of 9026 total-energy and force calculations. The resulting image closest to the saddle point has a maximum force 0.14 eV/Å (mainly along the direction of the transition path). ANEBA was run with a cutoff at the 21st iteration for a depth of five levels. We note that at levels three and two the algorithm needed to perform four times as many calculations in order to obtain a finer sampling on the region of interest. Nevertheless, the total number of calculations is 641, an improvement by a factor of more than 14 compared to the NEB calculation. The final saddle point approximation has a maximal force of 0.059 eV/Å and the maximum position difference compared to the NEB saddle point approximation is 0.04 Å, with an average of 0.002 Å difference per atomic position.

B. Concerted exchange in silicon

The concerted exchange mechanism in bulk Si[9] was proposed as a possible intrinsic, defect-free diffusion mechanism in silicon. This process involves a three-dimensional rotation of a dimer which can be seen as a triple Wooten–Winer–Weaire bond switch[16] performed on the same pair of atoms and leading to the exchange of the two atoms. Because of the strained configurations involved in this process, it has been used for a critical evaluation of several classical potentials for bulk Si.[17] It has also been studied using a thermodynamic integration method and Monte Carlo simulations, where it was shown that the self-diffusion coefficient corresponding to this process is in reasonable agreement with the experimental value.[18]

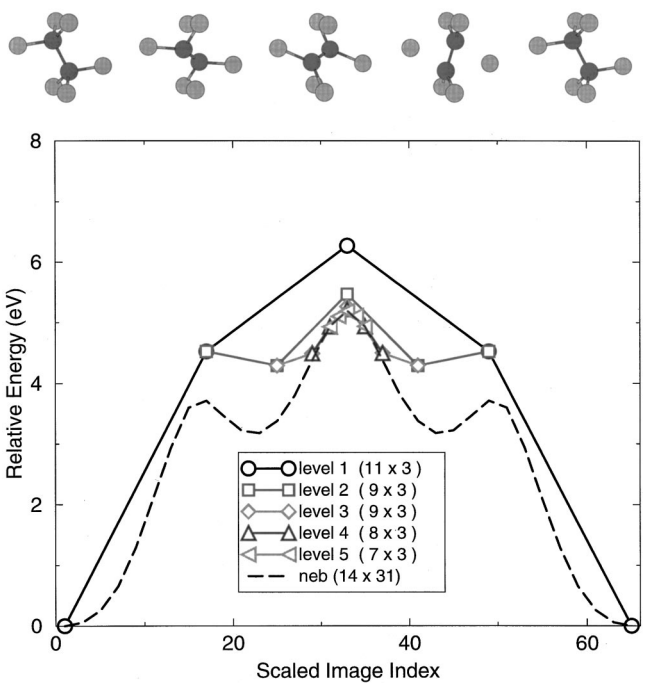

FIG. 4. Location of the saddle point in the concerted exchange mechanism for diffusion in bulk Si. At the top we show snapshots of the structure at selected points along the path. The atoms exchanging position are shown in darker shading.

In Fig. 4 we present the calculated full pathway as predicted by the NEB method and the successive recursions of ANEBA. The system was modeled with a supercell of 54 atoms in the diamond structure and the NEB calculation was started by giving five initial guesses along the path, based on the unrelaxed geometries of the pair rotation.[9] In this case a rough convergence criterion of a maximum force per atom of 0.257 eV/Å perpendicular to the MEP was used. The NEB method with 31 moving images converged to that accuracy with only 14 iterations, a total of 436 steps. The ANEBA algorithm converged smoothly, in a total of 134 steps. Note that the mirror symmetry of the path is conserved by both algorithms. The gain in performance in this case is approximately a factor of 3.5.

C. Ad-dimer diffusion on the Si(100) surface

As a final example we consider the surface diffusion of an ad-dimer on the Si(100) surface. This process has received renewed attention following the recent experimental observation by Qin *et al.*[19] of the rotational diffusion of a SiGe dimer on Si(100). The ad-dimer diffusion process has attracted increased theoretical and experimental interest following the direct measurements by Swartzentruber[10] using atom-tracking scanning tunneling microscopy (STM).

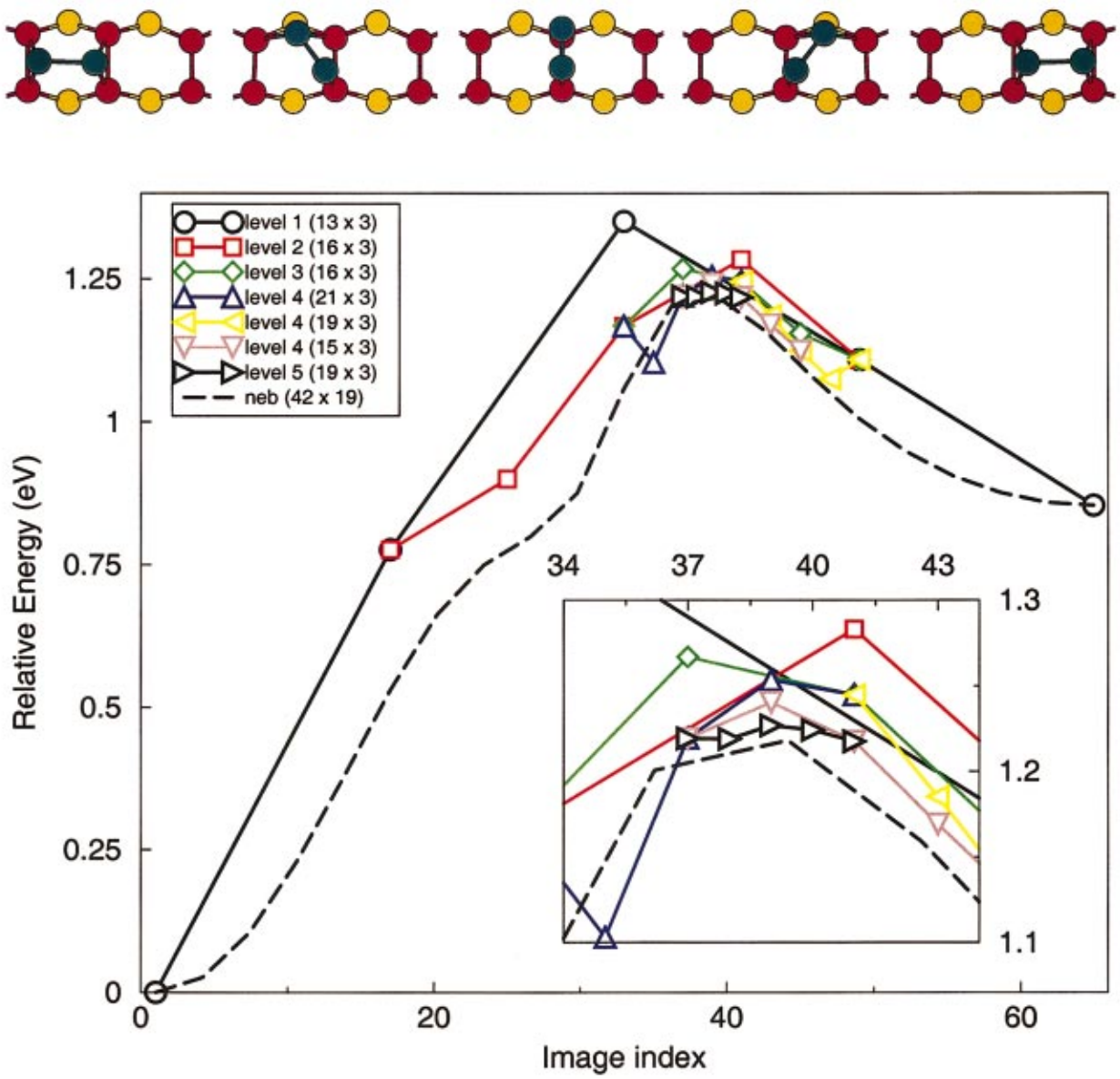

FIG. 5. (Color) Location of the saddle point in the rotational ad-dimer diffusion path. The inset is a detailed view of the saddle point region. At the top we show snapshots of the structure at selected points along the path. We use dark green, red, and gold, to color the ad-dimer, dimer row, and the atoms below the dimer row, respectively.

Whereas adatoms diffuse too rapidly at room temperature to be observed with STM, ad-dimers form more stable structures that are amenable to direct observation. The ad-dimers on the surface may form on top of the dimer row of the $c(4\times2)$ surface reconstruction or they may sit in the trough between two dimer rows. The diffusion of the ad-dimer is parallel to the dimer rows and the two distinct relaxation configurations do not easily mix.

We have calculated several possible paths of the ad-dimer on the Si(100) surface, notably the direct translational path,[20] the piecewise translation path,[21] a dimer-exchange path, and the rotational path. Within the limits of the SCC-DFTB method the lowest energy barrier is given by the rotational path, that may have two distinct forms depending on the direction of rotation. In Fig. 5 we compare the convergence characteristics of one ad-dimer rotational pathway using the NEB method and ANEBA. There is a local minimum in the middle of the path which corresponds to a symmetric configuration with the ad-dimer relaxing on top the dimer belonging to the dimer row. One of the ad-dimer atoms is located almost in the middle of the dimer and the other (the one performing the rotation) is on top of the surface dimer and slightly towards the trough. The energy of this configuration is 0.82 eV higher than the original ad-dimer configuration. The NEB method used a cutoff perpendicular force of 0.072 eV Å, and needed 42 iterations on a path of 21 images, a total of 800 force calculations. The ANEBA method, using five adaptation levels, needed a total of 359 calculations, a gain by a factor larger than 2.

IV. COMPARISON WITH AND INCORPORATION OF CLIMBING IMAGE

Recently, Jónsson and co-workers presented a method that converges at the same rate as NEB but yields the exact

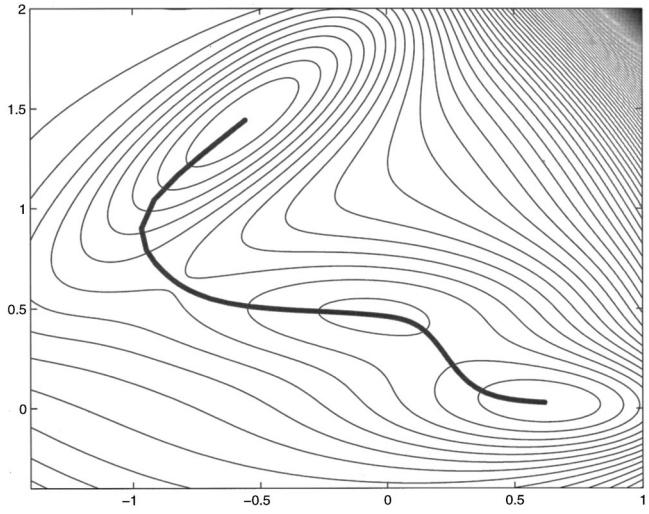

FIG. 6. The Müller–Brown potential and the MEP connecting the two distant minima.

saddle point, thus remedying one of the flaws associated with NEB.[7] The method, dubbed climbing image NEB (CI-NEB), involves freeing the highest energy state from all spring forces, and inverting the potential for that state while using only the projection along the path of the potential forces. This allows the highest energy state to search for the peak of the MEP, which is by definition the saddle point.

For the sake of comparison, we have run CI-NEB and ANEBA on a pair of two-dimensional examples, that of Fig. 2 (Ref. 5) and the Müller–Brown potential,[22] a standard example used in theoretical chemistry that is shown in Fig. 6. We also ran CI-ANEBA, wherein CI was implemented after a small number of ANEBA depth levels. This should greatly increase the efficiency of the method. ANEBA asymptotically converges to the exact saddle point, but, after only a few depth levels, locates the vicinity of the saddle with sufficient accuracy that the climbing image should rapidly find the exact saddle point. In contrast to the examples shown in the previous sections where we compared ANEBA to the original NEB method, for all the runs in this section (CI-NEB, ANEBA, and CI-ANEBA for both potentials) we used the new tangent definition[8] that was shown to sometimes improve the convergence behavior of the NEB method.

All three methods were optimized to give a saddle point energy within 5×10^{-4} of the exact transition state in the fewest number of steps. The results on the potential shown in Fig. 2 (Table I) indicate that ANEBA is more efficient than CI-NEB, requiring about 60% of the force calculations to find a slightly more accurate saddle point. The ANEBA method performed even better with CI as fewer ANEBA depth levels are required and the issue of asymptotic convergence is side-stepped. For the Müller–Brown potential, a similar procedure was executed. Table II presents these results and shows a similar but more pronounced trend: ANEBA is more efficient than CI-NEB, and CI-ANEBA more efficient still.

TABLE I. Results and calculation requirements for CI-NEB, ANEBA, and CI-ANEBA for the potential shown in Fig. 1. The second column, (SE), shows the saddle point energy estimate that is the outcome of the numerical approach, and the third column, (|SE-TSE|), shows the absolute difference of SE and the true saddle point energy (TSE).

Method	SE	|SE-TSE|	Force calculations	NEB iterations
CI-NEB, 9 states	−3.167 756	4.40×10^{-4}	319	43
ANEBA	−3.168 196	1.96×10^{-7}	194	28
CI-ANEBA	−3.168 030	1.66×10^{-4}	142	34

TABLE II. Same as Table I, for the Müller–Brown potential shown in Fig. 6.

Method	SE	|SE-TSE|	Force evaluations	NEB iterations
CI-NEB, 17 states	−40.664 487	3.56×10^{-4}	1069	69
ANEBA	−40.664 751	9.27×10^{-5}	558	156
CI-ANEBA	−40.664 641	2.02×10^{-4}	390	110

V. CONCLUSIONS

We presented a straightforward extension of the NEB method and its CI-NEB variation that allows for an efficient focusing to the saddle point of a potential energy surface. We demonstrated the convergence behavior of our approach in cases of practical interest, selected out of the realm of physical chemistry, materials, and surface science. The approach was able to locate the saddle points in all the cases that we have tested it and it was more efficient than the methods on which we have based it on. We believe that it is a useful tool for locating saddle points in processes that are known in advance. Its efficiency makes it ideally suited for applications that would otherwise be limited by the available computer time.

ACKNOWLEDGMENTS

Discussions with Martin Karplus and Jose Soler are gratefully acknowledged. This work was partially supported by National Computational Science Alliance under Grant No. MCB010005N and utilized the Boston University SGI Origin 2000 computer. The authors would like to acknowledge the support of the European Commission through Grant No. HPRI-1999-CT-00026 (the TRACS Program at EPCC). Part of this work was funded by a Graduate Research Fellowship (to Y.B.) from the National Science Foundation.

[1] R. Elber and M. Karplus, Chem. Phys. Lett. **139**, 375 (1987).
[2] L. Pratt, J. Chem. Phys. **85**, 5045 (1986).
[3] C. Dellago, P. Bolhuis, and D. Chandler, J. Chem. Phys. **110**, 6617 (1999), and references therein.
[4] G. Mills and H. Jónsson, Phys. Rev. Lett. **72**, 1124 (1994).
[5] H. Jónsson, G. Mills, and K. W. Jacobsen, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne, G. Ciccoti, and D. F. Coker (World Scientific, Singapore, 1998).
[6] Y. Brumer, A. A. Golosov, Z. D. Chen, and D. R. Reichman, J. Chem. Phys. **116**, 8376 (2002).
[7] G. Henkelman, B. P. Uberuaga, and H. Jónsson, J. Chem. Phys. **113**, 9901 (2000).
[8] G. Henkelman and H. Jónsson, J. Chem. Phys. **113**, 9978 (2000).
[9] K. C. Pandey, Phys. Rev. Lett. **57**, 2287 (1986).
[10] B. S. Swartzentruber, Phys. Rev. Lett. **76**, 459 (1996).

[11] F. Jensen, in *Introduction to Computational Chemistry* (Wiley, Chichester, 1999), Chap. 14.
[12] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[13] G. Kresse and J. Furthmuller, Phys. Rev. B **54**, 11169 (1996).
[14] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, Phys. Rev. B **58**, 7260 (1998).
[15] P.-O. Löwdin, Rev. Mod. Phys. **35**, 724 (1963).
[16] F. Wooten, K. Winer, and D. Weaire, Phys. Rev. Lett. **54**, 1392 (1985).
[17] E. Kaxiras and K. C. Pandey, Phys. Rev. B **38**, R12736 (1988).
[18] A. Antonelli, S. Ismail-Beigi, E. Kaxiras, and K. C. Pandey, Phys. Rev. B **53**, 1310 (1996).
[19] X. R. Qin, B. S. Swartzentruber, and M. G. Lagally, Phys. Rev. Lett. **85**, 3660 (2000).
[20] T. Yamasaki, T. Uda, and K. Terakura, Phys. Rev. Lett. **76**, 2949 (1996).
[21] B. Borovsky, M. Krueger, and E. Ganz, Phys. Rev. B **59**, 1598 (1999).
[22] K. Müller and L. D. Brown, Theor. Chim. Acta **53**, 75 (1979).