Letter to the Editor
Paths to which the Nudged Elastic Band Converges

DANIEL SHEPPARD, GRAEME HENKELMAN
Department of Chemistry and Biochemistry, and the Institute for Computational and Engineering Sciences, The University of Texas at Austin, Austin, Texas 78712-0165

Received 21 October 2010; Revised 29 November 2010; Accepted 11 December 2010
DOI 10.1002/jcc.21748
Published online 15 February 2011 in Wiley Online Library (wileyonlinelibrary.com).

Abstract: A recent letter to the editor (Quapp and Bofill, J Comput Chem 2010, 31, 2526) claims that the nudged elastic band (NEB) method can converge toward gradient extremal paths and not to steepest descent paths, as has been assumed. Here, we show that the NEB does in fact converge to steepest descent paths and that the observed tendency for the NEB to approach gradient extremal paths was a consequence of implementation errors. We also note that while the NEB finds steepest descent paths, these are not necessarily minimum energy paths in the sense of being a set of points which are minima in the potential energy surface perpendicular to the path. An example is given where segments of steepest descent paths follow potential energy ridges.

 J Comput Chem 32: 1769–1771, 2011

Key words: nudged elastic band; steepest descent paths; gradient extremal path; minimum energy path; saddle points

Steepest Descent Paths

There has been some discussion in the literature about the paths to which the nudged elastic band (NEB) converges. The NEB has been described as a method to find a minimum energy (ME) path between stable states.[1–3] A more accurate description is that the NEB finds a steepest descent (SD) path from saddle point(s) to minima, a so-called intrinsic reaction coordinate (IRC).[4] A recent paper by Quapp and Bofill,[5] claims that the NEB does not necessarily find a SD path. Without spring forces, the authors claim, the NEB can converge toward another kind of path called a gradient extremal (GE). Here, we clarify that the NEB will only converge to a SD path.

The NEB calculation discretizes a path into a set of images i, which are subjected to projected forces,

$$\mathbf{F}_i^{\mathrm{NEB}} = \mathbf{F}_i^{s\parallel} + \mathbf{F}_i^{g\perp}. \tag{1}$$

$\mathbf{F}_i^{s\parallel}$ are spring forces that act parallel to the path $\hat{\boldsymbol{\tau}}_i$,

$$\mathbf{F}_i^{s\parallel} = k(|\mathbf{R}_{i+1} - \mathbf{R}_i| - |\mathbf{R}_i - \mathbf{R}_{i-1}|)\hat{\boldsymbol{\tau}}_i \tag{2}$$

and $\mathbf{F}_i^{g\perp}$ are the forces due to the potential that act perpendicular to the path,

$$\mathbf{F}_i^{g\perp} = \mathbf{F}_i - (\mathbf{F}_i \cdot \hat{\boldsymbol{\tau}}_i)\hat{\boldsymbol{\tau}}_i. \tag{3}$$

Here, $\mathbf{F}_i = -\nabla V(\mathbf{R}_i)$ is the negative gradient of the potential at image i, $\hat{\boldsymbol{\tau}}_i$ is the local tangent to the path, and k is a spring constant which is used to keep the images equally spaced along the path.

When the NEB forces are followed by an appropriate optimizer, such as those described in ref. 6, convergence is reached when all images lie on a SD path. The SD path is precise in the limit of many images. A proof of this follows from considering the conditions when the NEB forces vanish. When the parallel force from eq. (2) is zero, the images are equally spaced,

$$|\mathbf{R}_{i+1} - \mathbf{R}_i| = |\mathbf{R}_i - \mathbf{R}_{i-1}|, \tag{4}$$

independent of the choice of spring constant k. When the perpendicular force is zero, eq. (3) reduces to

$$\mathbf{F}_i = (\mathbf{F}_i \cdot \hat{\boldsymbol{\tau}}_i)\hat{\boldsymbol{\tau}}_i, \tag{5}$$

showing that the tangent $\hat{\boldsymbol{\tau}}_i$ to the path at each image i is parallel to the force at that image, $\mathbf{F}_i$. This defines a SD path.

In ref. 5, the authors claim that the NEB does not necessarily converge to a SD path, but rather converges toward a GE. A GE is a path containing a set of points in which the derivative of the potential is parallel to a normal mode; it does not generally follow the gradient.

Correspondence to: G. Henkelman; e-mail: henkelman@mail.utexas.edu

Contract/grant sponsor: National Science Foundation; contract/grant number: CHE-0645497
Contract/grant sponsor: Texas Advanced Computing Center

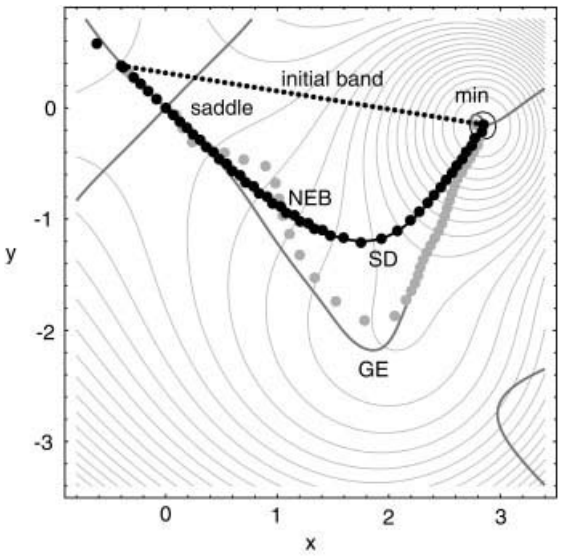

Figure 1. The grey points are the result of an NEB calculation described in ref. 5, in which an initial (linear) band converges toward a GE path. When the optimizer is corrected, the NEB is properly converges to a SD path.

The example cited in ref. 5 is an NEB calculation on the NFK potential as modified by Hirsch and Quapp.[7] Figure 1 shows SD and GE paths leading from a saddle point to a minimum on the potential energy surface. With the NEB implementation used by the authors in ref. 5, an initial linear band is seen to converge toward a GE. The reason that NEB did not converge to the SD path is due to a coding error in the damped-dynamics optimizer whereby sequential images along the NEB would accumulate the velocity of previous images in the band. Specifically, in a loop over images, i, the damped dynamics velocity was calculated as $\mathbf{V} = \mathbf{V} + \Delta t\,\mathbf{F}_i$ where $\mathbf{V}$ is the velocity, $\mathbf{F}_i$ is the force on the image, and Δt is the integrator time step. Instead, the velocity should be calculated for each image separately, $\mathbf{V}_i = \mathbf{V}_i + \Delta t\,\mathbf{F}_i$. When the code was corrected the NEB converges to the SD path. The unequal spacing of the images in the converged NEB is due to a choice of $k = 0$ in ref. 5. A more accurate path would be found using the standard implementation of the NEB described in ref. 2, with springs to keep the images equally spaced and an upwinding tangent to avoid oscillations in steep sections of the path. Two other changes would improve this calculation: fewer images are required to identify the saddle point with the climbing image NEB,[3] and second-order optimizers increase the rate of convergence.[6]

A second example, cited in ref. 5, of the NEB failing to converge to a SD path is shown in Figure 2. The authors refer to a calculation on the Wolfe-Quapp (WQ) potential, detailed in Figure 5 from ref. 8, in which the NEB was found to converge toward a ridge and pass through the maximum. Only five optimization steps were possible before this calculation diverged. Again, a review of the code used for this calculation revealed implementation errors. The primary problem was that the sign of the force was incorrect so the ridge appeared to the NEB as a valley. The perpendicular force on image i was calculated as $\mathbf{F}_i^{g\perp} = \mathbf{G}_i - (\mathbf{G}_i \cdot \hat{\boldsymbol{\tau}}_i)\hat{\boldsymbol{\tau}}_i$ where $\mathbf{G}_i = -\mathbf{F}_i$. Instead, it should be $\mathbf{F}_i^{g\perp} = \mathbf{F}_i - (\mathbf{F}_i \cdot \hat{\boldsymbol{\tau}}_i)\hat{\boldsymbol{\tau}}_i$. Figure 2 shows a properly converged NEB to a SD path passing through saddle point (c), calculated using a corrected version of the code.

Minimum Energy Paths

The authors of ref. 5 bring up an important point about the NEB which has not been widely addressed in the literature. It is commonly assumed that the NEB finds a ME path. The term ME path can be used interchangeably with SD path, in the sense that a SD trajectory from a saddle minimizes the energy along the path. Here, we address the alternate definition of a ME path as a set of points which are energy minima perpendicular to the path, and in this sense, a ME path is not necessarily the same as a SD path. This issue has been extensively studied by Quapp, as well as related features on reaction paths such as bifurcation points and ridge-valley inflection points.[9]

To show an example where the NEB finds a SD path which is not a ME path, we have constructed a periodic potential containing two Gaussian hills,

$$\cos(\pi x) + \cos(\pi y) + \pi e^{-\pi x^2}[e^{-\pi(y-0.8)^2} + e^{-\pi(y+0.8)^2}].$$

A contour plot of this potential is shown in Figure 3. The ME path from the highest energy saddle point (b) terminates at nearby inflection points, where the SD path crosses the dashed line and the positive mode perpendicular to the path (a valley) becomes negative (a ridge). The SD path continues on to saddle points (a) and (c) and then to minima (A) and (D). In this example, the converged NEB follows the SD path along the valley, where it is also a ME path, but also along a ridge where the energy is a maximum perpendicular to the path. The NEB will follow such paths in the limit of many images, but it will become unstable when $|C| > |\mathbf{F}|/\Delta\mathbf{R}$,

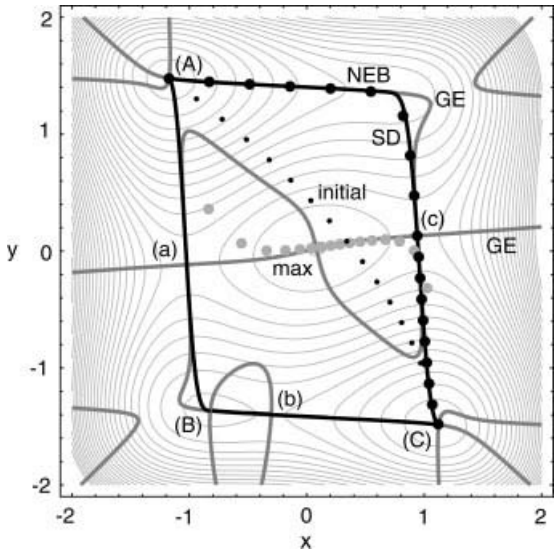

Figure 2. The NEB calculation described in ref. 8 (grey points) from minimum (A) to (C) on the WQ potential was shown to follow a ridge between saddle points (a) and (c) through a maximum along a GE path. When an error in the potential gradient is corrected (black points), the NEB properly converges to a SD path through saddle (c).

where C is the negative curvature perpendicular to the ridge, $\mathbf{F}$ is the magnitude of the force down the ridge, and $\Delta\mathbf{R}$ is the image spacing. This analysis is closely related to stability analysis of the NEB in ref. 2.

The NEB can follow ridges in higher dimensional systems as well. An example is the diffusion of an Al adatom on the Al(110) surface, which has been studied by Tiwary and Fichthorn.[10] The adatom sits in a $(1\bar{1}0)$ channel on the surface, as shown in Figure 3. Calculations of diffusion pathways were done using the Vienna *ab initio* Simulation Package using the PW91 GGA functional and the projector augmented wave method to treat core electrons.[11] Valence electrons were described using a plane-wave basis with a cut-off energy of 250 eV. The adatom was placed on a $p(4 \times 3)$ slab with 10 bilayers, freezing the bottom five to the bulk geometry. A vacuum gap of 15 Å separated the periodic images of the slabs, and a $6 \times 6 \times 1$ Monkhorst-Pack mesh of was used to sample the Brillouin zone.

Figure 4 shows the lowest energy mechanisms for cross-channel diffusion. The NEB calculation from minimum (A) to (D) converges along a ridge in a transition analogous to Figure 3. The transition path in Figure 4 climbs from minimum (A) up the ME path to saddle point (a). Continuing on to (C) would lead to a diagonal cross-channel diffusion mechanism. Instead, the SD path from the higher energy saddle (b) (which is in fact a very shallow metastable minimum separated by two saddles) is followed by an NEB connecting to (D), resulting in diffusion directly across the channel. There are two ridge segments highlighted in red, one between (a) and (b), and another between (b) and (c), where the SD path is not a ME path.

To summarize, in the limit of many images the NEB converges to a SD path. The SD path is often a ME path, but this is not necessarily the case, for example when the SD path approaches a saddle via a potential energy ridge.

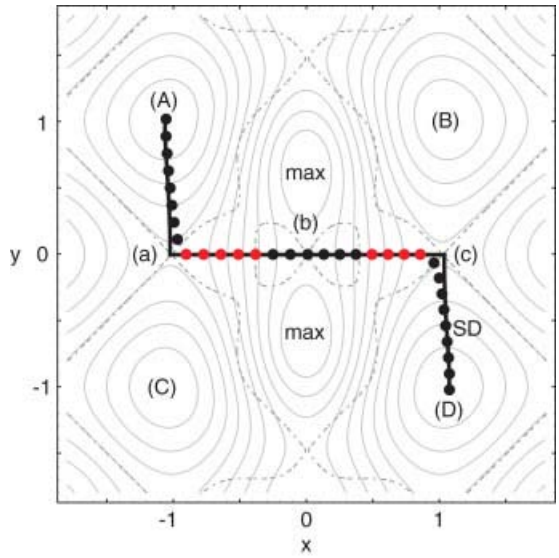

Figure 3. The SD path (solid black) from the highest energy saddle (b) terminates at minima (A) and (D) after passing through intermediate saddles (a) and (c), respectively. The images in red are where the NEB converges to a ridge along the SD path, which is not a ME path. The dashed line is the boundary between regions with one and two negative modes.

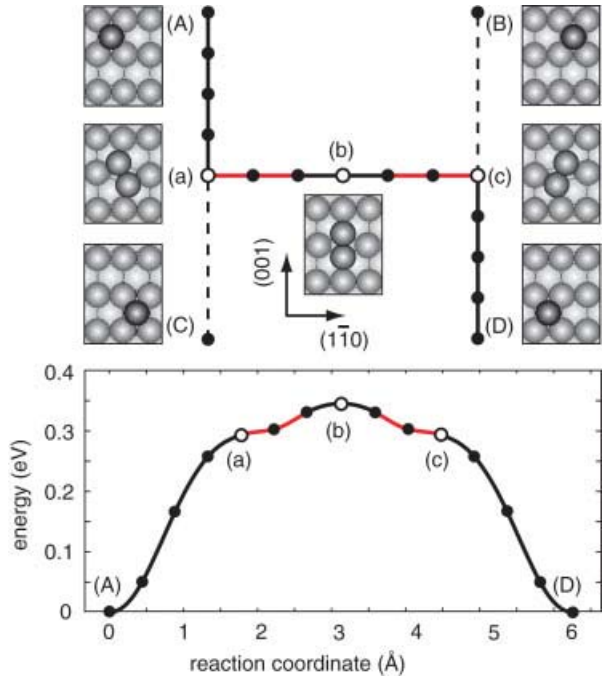

Figure 4. Al adatom cross-channel diffusion on Al(110) is shown between four minima (A)–(D). Saddles (a) and (c) correspond to diagonal exchange processes where the adatom pushes a surface row-atom into the neighboring channel along the $1\bar{1}1$ and $\bar{1}11$ directions, respectively. There is also a higher-energy cross-channel mechanism through saddle (b). The SD path found by an NEB through this saddle connecting minima (A) and (D) is shown to pass through the lower energy saddles (a) and (c) via the ridge segments, shown in red.

Acknowledgments

We thank Dr. Wolfgang Quapp and Antoni Aguilar-Mogas for supplying their NEB codes and also the anonymous reviewers for their helpful comments.

References

1. Jónsson, H.; Mills, G.; Jacobsen, K. W. In Classical and Quantum Dynamics in Condensed Phase Simulations; Berne, B. J.; Ciccotti, G.; Coker, D. F., Eds., World Scientific: Singapore, 1998, pp. 385–404.
2. Henkelman, G.; Uberuaga, B. P.; Jónsson, H. J Chem Phys 2000, 113, 9901.
3. Henkelman, G.; Jónsson, H. J Chem Phys 2000, 113, 9978.
4. Fukui, K. J Phys Chem 1970, 74, 4161.
5. Quapp, W.; Bofill, J. M. J Comput Chem 2010, 31, 2526.
6. Sheppard, D.; Terrell, R.; Henkelman, G. J Phys Chem 2008, 128, 134106.
7. Hirsch, M.; Quapp, W. Chem Phys Lett 2004, 395, 150.
8. Aguilar-Mogas, A.; Giménez, X.; Bofill, J. M. J Comput Chem 2010, 31, 2510.
9. Quapp, W. J Mol Struct 2004, 695–696, 95.
10. Tiwary, Y.; Fichthorn, K. A. Phys Rev B 2010, 81, 195421.
11. Kresse, G.; Joubert, D. Phys Rev B 1999, 59, 1758.