Development and evaluation of computational methods for studies of chemical reactions

Vilhjálmur Ásgeirsson

Faculty of Physical Sciences
University of Iceland
2021

Development and evaluation of computational methods for studies of chemical reactions

Vilhjálmur Ásgeirsson

Dissertation submitted in partial fulfillment of a
Philosophiae Doctor degree in Chemistry

Advisor
Hannes Jónsson

PhD Committee
Hannes Jónsson
Egill Skúlason
Elvar Örn Jónsson

Opponents
Baron Peters
Gísli Hólmar Jóhannesson

Faculty of Physical Sciences
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, June 2021

Development and evaluation of computational methods for studies of chemical reactions
Computational methods for theoretical chemistry
Dissertation submitted in partial fulfillment of a *Philosophiae Doctor* degree in Chemistry

Faculty of Physical Sciences
School of Engineering and Natural Sciences
University of Iceland
Dunhaga 5
107, Reykjavik
Iceland

Telephone: 525-4700

Bibliographic information:
Vilhjálmur Ásgeirsson, 2021, *Development and evaluation of computational methods for studies of chemical reactions*, PhD dissertation, Faculty of Physical Sciences, University of Iceland, 269 pp.

ISBN 978-9935-9564-9-1

Printing: Háskólaprent
Reykjavik, Iceland, June 2021

Abstract

Methods for identifying the mechanism and estimating the rate of chemical reactions are presented and evaluated for a wide range of systems, using both classical and quantum mechanical description of the atomic nuclei. In the classical case, the minimum energy path (MEP) connecting two minima, that represent states of the system, is found. Energy maxima on the path correspond to first order saddle points on the energy surface and give an estimate for the activation energy of the transition. For quantum mechanical description of the atoms, the optimal tunneling path (OTP) is found. This path is equivalent to a first order saddle point on the action surface, referred to as an instanton, and can be used to calculate the rate of thermally assisted tunneling. In order to navigate on these surfaces, the energy and force acting on the atoms needs to be evaluated. Such calculations are typically carried out using computationally intensive electronic structure methods. It is, therefore, important to develop both reliable and efficient algorithms to navigate on these surfaces in an efficient way with as few evaluations of the energy and atomic forces as possible.

The various methods presented in this work are extensions of the widely used nudged elastic band (NEB) method. In NEB, a trial path represented by a set of points is iteratively displaced towards a target path, an MEP or OTP. The path is displaced downhill on the surface along the directions perpendicular to the path and spring forces are used to keep discretization points evenly distributed along the path. To achieve this, an accurate estimate of the tangent to the path is needed in order to decompose the forces into perpendicular and parallel components. In the first part of this work, the computational efficiency of NEB calculations of MEPs for molecular reactions is addressed. There, excessive computational effort is often needed because the MEPs often include long segments with little or no change in the energy. Computational resources are therefore wasted on resolving irrelevant segments of the path. Moreover, a sparse distribution of points along the path may also yield an inaccurate estimate of the tangent. This can affect the efficiency of an NEB calculation and can even lead to non-convergence. Two NEB variants are presented to automatically focus the computational effort on the most important part of the MEP, i.e. the region around the highest energy maximum. In one of these methods, a loose convergence on the MEP is first obtained and then a new set of points is automatically distributed in the region of the energy barrier to improve the resolution around the energy maximum and hence improve the tangent estimate there. In the second method, an increased density of points is obtained in the critical region of the MEP by adaptively scaling the strength of the spring interaction according to the energy, making the springs stiffer in regions of higher energy. Experience will show which one of these two approaches will turn out to be optimal, or perhaps a combination of both. The computational effort when searching for saddle points can be reduced further by using a combination of NEB and an eigenvector-following (EF) method. In this approach, the points along the path are

first converged loosely to the MEP. Then, information obtained from the NEB path and the point of maximum energy are used to automatically start an EF search to swiftly target the saddle point. These methods are applied to various chemical reactions and to a database of 121 molecular reactions. The methods have been implemented in the ORCA quantum chemistry software which is rapidly becoming the most widely used tool for electronic structure calculations in computational chemistry.

In the second part of this work, the focus is on the quantum mechanical description of the atomic nuclei and identification of OTPs. An OTP traces out the same path on the action surface as an instanton and can therefore be used to estimate the tunneling rate. Calculations of OTPs are found to be more efficient than the typical search method used for instanton calculations. The main reason is that the distribution of system images along the OTP are controllable while the points accumulate near the endpoints in instanton calculations. Therefore, fewer images can be used to represent the path in OTP calculations compared to instanton calculations.

In the first two parts, the electronic structure computations are carried out using density functional theory (DFT) as is now commonly done in computational chemistry. However, the selection of an appropriate level of theory for calculations of molecules and chemical reactions can be difficult. In this regard, a particularly interesting and challenging diamine cation is studied in the third part of this work. In this case, the existence of both a localized and delocalized electronic state has been inferred from experimental measurements. While, standard electronic structure methods, e.g. commonly used density functionals and the coupled cluster singles-doubles-(triples), method are unable to predict the existence of a localized electronic state. To shed light on this issue, which has turned into a controversy in the literature, and determine whether the localized state truly exists, high-level multireference wavefunction calculations of the energy surface are carried out and found to establish the existence of the localized state.

Útdráttur

Aðferðir til þess að finna hvarfgang og meta hraða efnahvarfa eru þróaðar og prófaðar á fjölbreytilegum kerfum, þar sem bæði klassísk og skammtafræðileg lýsing á atómkjörnum er notuð. Fyrir klassíska lýsingu á atómunum eru fundnir lágmarksorkuferlar sem tengja tvö orkulágmörk og samsvara stöðugum ástöndum kerfisins. Á slíkum ferli samsvarar orkuhámark fyrsta stigs söðulpunkti á orkuyfirborðinu og gefur mat á virkjunarorkunni fyrir hvarfið. En, fyrir skammtafræðilega lýsingu á atómunum er besti smugferillinn fundinn. Þessi ferill samsvarar fyrsta stigs söðulpunkti á verkunaryfirborðinu og er oft kenndur við snareindir. Út frá slíkum punkti er hægt að reikna hraða á varmaörvuðu smugi. Til þess að hægt sé að kanna slík yfirborð og finna þessa ferla þarf að reikna orku og kraftinn sem verkar á atómin. Slíkir reikningar fela yfirleitt í sér þunga tölvuútreikninga. Því er afar mikilvægt að þróa aðferðir sem eru bæði nákvæmar og hagkvæmar að því leiti að sem fæsta orku og kraftareikninga þurfi til.

Hinar ýmsu aferðir sem eru þróaðar hér eru útvíkkanir á hinni vel þekktu teygjubandsaðferð, NEB. Í þessari aðferð er ferill, sem er lýst sem safni af hnitum atómanna, færður í átt að lausnarferlinum, annað hvort lágmarksorkuferli eða besta smugferli, með ítrun. Hver punktur í þessari strjálu lýsingu á ferlinum er þá færður niður eftir yfirborðinu í stefnu hornétt á ferilinn og gormkraftar notaðir til þess að viðhalda jafndreifingu á punktunum eftir ferlinum. Þetta krefst þess að hafa nógu gott mat á snertlinum eftir ferlinum til þess að framkvæma vörpun á kröftunum í þverstæða og samsíða þætti. Í fyrsta hluta verksins er hagkvæmni teygjubandsaðferðarinnar fyrir reikninga á lágmarksorkuferlum fyrir sameindahvörf metin og betrumbætt. Raunin er að oft er þörf á að nota óþarflega mikla reiknigetu fyrir teygjubandsreikninga á slíkum hvörfum, þar sem orkan á lágmarksorkuferlinum breytist oft lítið sem ekkert á köflum. Þar af leiðandi er reikniafli sóað í að lýsa lítilvægum hluta ferilsins á sama tíma og upplausn ferilsins er ekki nægjanlega góð til að fá mat á snertilnum á mikilvægum hluta hans. Þetta getur haft mikil áhrif á kostnað teygjubandsreikninga og getur jafnvel orsakað það að samleitni náist ekki. Tvær breytingar á teygjubandsaðferðinni eru settar fram þar sem áhersla er lögð á mikilvægasta hluta lágmarksorkuferilsins, þ.e. þann hluta sem inniheldur hæsta orkuhámarkið. Í annarri aðferðinni er veikri samleitni á lágmarksorkuferlinum fyrst náð og síðan er nýju punktasafni sjálfvirkt dreift á svæði orkuhólsins til þess að bæta upplausnina á ferlinum í grennd við hámarkið. Í hinni aðferðinni er auknum þéttleika punkta náð á þessu mikilvæga svæði lágmarksorkuferilsins með skölun á styrkleika gormkraftsins samkvæmt orku kerfinsins á hverjum stað, þar sem gormarnir eru gerðir stífari á háorkusvæðum á meðan ítranirnar eru framkvæmdar. Reynslan mun sýna hvor aðferðin, eða mögulega samsetning beggja, er hagkvæmari. Með því að setja saman teygjubandsaðferðina og eiginvigrarakningu er dregið enn frekar úr reiknikostnaði þegar leitað er að söðulpunktum. Í þeirri aðferð er grófri samleitni á lágmarksorkuferlinum náð. Upplýsingum er síðan safnað af ferlinum, þ.m.t. punktinum með hæstu orkuna. Þessar upplýsingar eru sjálfvirkt notaðar til þess að byrja eiginvigrarakningu sem auðveldlega

nær samleitni á söðulpunktinn sem svarar til viðkomandi efnahvarfs. Aðferðirnar eru notaðar á ýmis efnahvörf og á gagnasafn sem inniheldur 121 sameindahvörf. Aðferðirnar hafa verið innleiddar í ORCA skammtaefnafræði hugbúnaðinum sem nýtur ört vaxandi vinsælda og er á góðri leið með að verða útbreiddasti hugbúnaður í heimi fyrir rafeindastrúkturreikninga í efnafræði.

Í öðrum hluta verksins er lögð áhersla á skammtafræðilega lýsingu atómanna og leit að bestu smugferlum. Besti smugferill fylgir sömu leið og snareind á verkunaryfirborðinu og því er hægt að hann til þess að reikna út skammtafræðilegan smughraða efnahvarfa. Reikningar á bestu smugferlum eru hagkvæmari en hefðbundna leitaraðferðin sem notuð hefur verið til þess að finna snareindir. Aðal ástæðan fyrir þessari auknu hagkvæmi er sú að það er hægt að stjórna dreifingu punktanna á ferlinum í leitinni að besta smugferlinum, á meðan flestir punktarnir enda í grennd við endapunktana í snareindareikningum. Því er hægt að nota færri punkta til að ná góðri upplausn á ferlinum í reikningum á bestu smugferlum í samanburði við snareindareikninga.

Í fyrstu tveimur hlutum verksins er notast við þéttnifellafræði til að reikna rafeindastrúkturinn eins og almennt tíðkast nú í reikniefnafræði. Aftur á móti getur val á viðeigandi orkuyfirborði fyrir reikninga á sameindum og efnahvörfum oft reynst erfitt. Með þetta í huga er sérstaklega áhugaverð og erfið díamínkatjón skoðuð í þriðja hluta verksins. Tilraunamælingar á þessari sameind hafa verið túlkaðar þannig að bæði sé tilstaðar staðbundið og óstaðbundið rafeindaástand. Hefðbundnar rafeindastrúktúraðferðir sýna hins vegar ekki tilvist staðbundna rafeindnaástandsins, svo sem vinsæl rafeindaþéttleikafelli og jafnvel CCSD(T) aðferðin. Nákvæmir og flóknir fjölástands bylgjufallsreikningar (*e. multireference wavefunction calculations*) eru notaðir til að varpa ljósi á þetta misræmi á milli rafeindareikninga og tilrauna og skera úr um tilvist staðbundna rafeindaástandsins. Í samræmi við niðurstöður tilraunanna, spá þessir nákvæmari reikningar fyrir um tilvist staðbundna ástandsins.

Table of Contents

List of Figures

List of Original Articles

Publications included in the thesis

Article I: Ásgeirsson, V., Arnaldsson, A. and Jónsson, H. (2018). Efficient evaluation of atom tunneling combined with electronic structure calculations, *J. Chem. Phys.* **148**, 102334.

Article II: Ásgeirsson, V., Birgirson, B.O., and Jónsson, H. (2021). Nudged elastic band calculations with focus on the region around the highest saddle point, *To be submitted to Journal of Chemical Theory and Computation*.

Article III: Ásgeirsson, V., Birgirson, B.O., Bjornsson, R., Becker, U., Neese, F., Riplinger, C. and Jónsson, H. (2021). Nudged elastic band method for molecular reactions using energy-weighted springs combined with eigenvector following, *Accepted in Journal of Chemical Theory and Computation*.

Article IV: Galyńska, M., Ásgeirsson, V., Jónsson, H. and Bjornsson, R., (2021). Localized and delocalized states of a diamine cation: Resolution of a controversy, *J. Pys. Chem. Lett.* **12**, 1250–1255

.

Other publications not included in the thesis

Article V: Birgisson, B.O., Ásgeirsson, V., Ivanov, A., Galyńska, M., Bjornsson, R. and Jónsson, H. (2021). Localized and delocalized states of diamine cation: A qualitative failure of most density functional approximations. *To be submitted.*

Article VI: Schmerwitz, Y. L. A., Ásgeirsson, V., Engels, B. and Jónsson, H. (2021). Growing nudged elastic band method for initial reaction path generation on the image-dependent pair potential surface. *To be submitted.*

Article VII: Ásgeirsson, V. and Jónsson, H. (2020). Exploring potential energy surfaces with saddle point searches. *Handbook of Materials Modeling - Methods: Theory and Modeling*, Springer, 689-714

.

Article VIII: Koistinen, O.P., Ásgeirsson, V., Vehtari, A. and Jónsson, H. (2019). Minimum mode saddle point searches using Gaussian process regression with inverse-distance covariance function, *J. Chem. Theory Comput.* **16**(1), 499-509.

Article IX: Koistinen, O.P., Ásgeirsson, V., Vehtari, A. and Jónsson, H. (2019). Nudged elastic band calculations accelerated with Gaussian process regression based on inverse interatomic distances, *J. Chem. Theory Comput.*, **15**(12), 6378-6751

Article X: Koistinen, O.P., Dagbjartsdóttir, F.B., Ásgeirsson, V., Vehtari, A. and Jónsson, H. (2017). Nudged elastic band calculations accelerated with Gaussian process regression, *J. Chem. Phys.* **147**(15), 152720.

Article XI: Ásgeirsson, V., Bauer, C.A. and Grimme, S. (2017). Quantum chemical calculation of electron ionization mass spectra for general organic and inorganic molecules, *Chem. Sci.* **8**(7), 4879-4895.

Article XII: Ásgeirsson, V., Bauer, C.A. and Grimme, S. (2016). Unimolecular decomposition pathways of negatively charged nitriles by ab initio molecular dynamics, *Phys. Chem. Chem. Phys.* **18**(45), 31017-31026.

Abbreviations

BO:	Born-Oppenheimer approximation
PES:	Potential energy surface
SP:	Saddle point
MEP:	Minimum energy path
SP:	Saddle point
LQA:	Local quadratic approximation
NR:	Newton-Raphson
CM:	Cerjan-Miller method
RFO:	Rational function optimization
P-RFO:	Partition rational function optimization
MMF:	Minimum mode following
NEB:	Nudged elastic band
CI:	Climbing image
VPO:	Velocity projection optimization
FIRE:	Fast inertial relaxation engine
BFGS:	Broyden-Fletcher-Goldfarb-Shanno
L-BFGS:	Limited memory BFGS
TST:	Transition state theory
QTST:	Quantum transition state theory
HTST:	Harmonic transition state theory
QHTST:	Harmonic quantum transition state theory
qqHTST:	Quasi quantum harmonic transition state theory
CFP:	Closed Feynman Path
MAP:	Minimum action path
HF:	Hartree-Fock
CCSD(T):	Coupled cluster singles, doubles and perturbative triples
MRCI:	Multi-reference configuration interaction
DFT:	Density functional theory
XC:	Exchange-correlation functional
KS:	Kohn-Sham
LDA:	Local density approximation
GGA:	Generalized gradient approximation

Acknowledgments

First and foremost, I would like to express my heartfelt gratitude and appreciation to my academic supervisor and good friend Prof. Hannes Jónsson. Hannes is both a remarkable scientist and a great teacher. It has been an absolute pleasure to work with him for all these years, where I have received continual guidance, support and encouragement throughout the years. I sincerely feel indebted to him. Without him, I would most likely not be sitting here writing this dissertation.

I would like to give special thanks to Benedikt Orri Birgisson whom I have been fortunate enough to have shared office space and collaborated with on multiple interesting projects during my doctoral studies. We have had some really fun times together and I wish you the best of luck in your own doctoral studies. I would also like to extend my thanks to Dr. Ragnar Björnsson for the many interesting and fruitful discussions about our joint projects and really about the entirety of computational chemistry, sometimes over a beer or two. I believe these talks to have been very informative and fun! Also, I would like to thank Ragnar and Dr. Marta Galyńska for allowing me to be a part of the challenging investigation into the nature of the electronic structure of the diamine cation.

I want to thank Dr. Christoph Riplinger, Prof. Frank Neese, Dr. Ute Becker and Dr. Frank Wennmohs for the very warm welcome to the Max-Planck-Institut für Kohlenforschung where I was given the opportunity to work on the ORCA quantum chemistry software. Also, for allowing me to present my work on a very large platform in front of hundreds of people, i.e. the 5th ORCA users meeting in 2020. That was definitely an interesting (and stressful) experience. In our collaborative ORCA/NEB project, I worked closely with Christoph, who deserves a special recognition and gratitude.

For our collaboration on the line-integral nudged elastic band method I would like thank Dr. Andri Arnaldsson. For our joint work on the machine-learning augmented nudged elastic band and dimer methods I would like to thank Dr. Olli-Pekka Koistinen who was the central figure in the project. But, also Freyja Björk Dagbjartsdóttir and Prof. Aki Vehtari. I would also like to thank Prof. Alessandro Laio and Katrín Blöndal for the interesting, but still unfinished, development of a novel (nearly-deterministic) single-ended saddle point search method. I would like to thank Reynir Kristjánsson for our collaborative investigation into the 'curse-of-dimensionality', as well as, Yorick Schmerwitz for our collaboration on the new growing nudged elastic band method for initial path generation.

I would like to extend my thanks to Dr. Enrique Batista at the Center of Non-linear studies (CNLS) at the national laboratory of Los Alamos, New Mexico, who invited me to come work at the CNLS in 2017 and also to present my work in the 38th CNLS annual conference in Santa Fe, New Mexico, 2018. It was really a wonderful experience. I would also like to thank Prof. Baron Peters for allowing me to participate in the

summer school "Rare Evens: Applications, Computations and Theory" in Bangalore, India, 2019. It was definitely the most enjoyable and educational summer-school I have been to, so far. It was also quite intense. I was given the honor of teaching a practical class about adaptive kinetic Monte Carlo simulations and also to present my work, with both a poster- and an oral-presentation.

I would also like to thank Prof. Egill Skúlason for his friendship and support throughout the years, without him I might not have discovered computational chemistry. Furthermore, I want to extend my gratitude to all current and former members of the 'Theochem group' at the University of Iceland (the list of people here is simply too long to put in writing) for your discussions, friendship and care. To put it simply, an incredible group of people! All of you have really made my time as a doctoral candidate enjoyable, as well as, educational. At last but not least, I would like to give special thanks to my family for their unwavering patience and support over the years.

While writing this dissertation and looking back over my time as a doctoral candidate, I feel immensely grateful and fortunate for all of the opportunities I have been granted and also the alarming number of people that I have gotten to know and now consider as friends... thank you all!

1 Introduction

Computer simulations have become a valuable tool for chemists. Not only can such simulations help interpret results obtained in the laboratory but they can also, in principle, predict what chemicals and materials would be interesting to prepare and study for various purposes. The simulations are, however, not able to give precise results, there are uncertainties because of limitations in the methodology and computational power needed for highly accurate simulations. The improvements have been dramatic over the past decades, in part because of greatly increased computational power, but also and not less because of improved algorithms for the calculations. Calculations of the mechanism and rate of chemical reactions that were considered to be highly challenging just a couple of decades ago, can almost be considered to be routine today. Still, there is much work to be done in the development of simulation methods to make them more robust and to make them applicable to even more challenging systems, and that is the prime focus of this thesis.

The standard approach in computational chemistry research is based on the Born-Oppenheimer or adiabatic approximation. It divides the task of describing the atomic nuclei and electrons into two parts, first the properties of the electronic subsystem are calculated for a fixed position of the nuclei, and then in the second step the motion of the nuclei is accounted for. This is not entirely rigorous, there can be non-adiabatic effects where the two are coupled closely and need to be treated simultaneously, but in most cases this separation is a good approximation and this will be taken to be the case in the examples discussed here. In most cases the atomic nuclei are treated as classical particles since they are relatively heavy, about three orders of magnitude heavier than an electron, and usual ambient conditions correspond to relatively high temperature. However, for light atoms such as hydrogen and for lower temperature, it is necessary to use a quantum mechanical description of the nuclei and one of the articles in this thesis, article I, describes a method for finding the mechanism of reactions where quantum mechanical tunneling is faster than a classical over-the-barrier hop.

A great deal of effort has been invested in the development of algorithms for solving the first step, the electronic structure calculation. The basic equation is well known, the Schrödinger equation, but the challenge is to find approximations that make it possible to carry out calculations of large enough systems while still obtaining results that are accurate enough to describe the system under investigation. One of the articles in this thesis, article IV, addresses this issue as it deals with a molecular cation which turns out to be challenging for standard electronic structure methodology. Neither commonly used functionals in density functional theory (DFT) calculations, (Becke, 2014; Burke, 2012) nor the wave function based method that is considered to be the 'golden standard' of computational chemistry, i.e. the coupled cluster singles, doubles and perturbative triples (CCSD(T)) (Bartlett and Musiał, 2007), give results that are consistent with recent experimental measurements where it has been shown that both a localized and a

delocalized electronic state exists for this molecular cation. (Cheng et al., 2016) There has been some controversy over this in the literature (Ali et al., 2018; Cheng et al., 2018) but the article presents high level calculations, arguably state-of-the-art methodology, to resolve this controversy.

The second step in the two step process of characterizing a chemical reaction, the identification of the mechanism of the atomic rearrangement and estimation of the rate has not been addressed nearly as much as the first step, the electronic structure calculation. This second step is the main focus of the present dissertation. There is still much room for improvement in the computational methods that have so far been developed for finding the mechanism and rate of chemical reactions. As mentioned above, article I, presents an algorithm for incorporating the quantum mechanical description of the atomic nuclei in the rate calculation, but it is rather exceptional if a classical particle description does not suffice. So, a greater emphasis is placed on further development of the tools for describing rearrangements of classical particles and this is the task addressed in articles II and III. Each evaluation of the electronic structure, which gives the electronic energy of the system for a given set of atom coordinates and the atomic forces, typically involves a significant computational effort. The number of such computations that need to be carried out in order to find the mechanism and estimate the rate of a given transition (i.e. elementary reaction step) is therefore of critical concern. The goal is to reduce as much as possible the energy/force evaluation needed to reach a converged solution. The two articles are based on the so-called nudged elastic band (NEB) method (Mills et al., 1995; Jónsson et al., 1998) for finding minimum energy paths on that connect two local energy minima on an energy surface. The energy minima characterize the initial and final state of the transition and are also referred to as the reactant and product states. The minimum energy path is such that at each point on the path, no force acts on the atoms perpendicular to the path. The NEB method is widely used for such calculations in solid state chemistry, as well as, in condensed matter physics research, (Henkelman et al., 2002; Ásgeirsson and Jónsson, 2020). However, there is still room to improve the NEB method, i.e. to make it more reliable, as well as, efficient. The latter can be achieved by reducing the number of energy/force evaluations required to obtain the minimum energy path. In article II, the Z-NEB method is presented where a regular NEB calculation is stopped after a rough approximation to the minimum energy path has been obtained, the computational effort is then focused onto the most relevant region of the path, i.e. near the point of highest energy. The point of highest energy, which corresponds to a first order saddle point on the energy surface, provides an estimate of the activation energy for the transition. By focusing the calculation on the region around the point of maximum energy, the path can be represented with fewer snapshots (or images) of the system and thereby reducing the number of computationally intensive electronic structure computations. Another way to accomplish this improved resolution of the minimum energy path around the maximum is presented in article III where the strength of the springs connecting adjacent images along the path is adjusted according to the energy of the system. This method is referred to as energy-weighted NEB. (Henkelman et al., 2000) The higher the energy, the stiffer the springs. Moreover, as in Z-NEB, the calculation of the minimum energy path is stopped after a certain level of convergence has been reached. Then, the highest energy image is moved to the first order saddle point using a more rigorous saddle point

seach method that is based on approximations to the matrix of second derivatives of the energy with respect to atomic coordinates. (Baker, 1986) This combination of methods, NEB-TS, is found to be remarkably efficient for a large dataset of 121 molecular reactions. (Zimmerman, 2013a; Birkholz and Schlegel, 2015) Reactions of molecules in the gas phase are more challenging than the typical applications of the NEB method which have mostly been for transition in and on condensed phases of matter. (Ásgeirsson and Jónsson, 2020) In the gas phase, molecules or fragments of molecules are often highly flexible and can move significant distances without appreciable change in the energy. Therefore, for molecular reactions the minimum energy paths are often long and the segment of the path that corresponds to the more interesting and relevant reactive event (e.g. bond breaking and formation) is short. This can lead to poor resolution of the region around the energy maximum unless an excessive number of images and thereby computational effort is used. The Z-NEB, EW-NEB and NEB-TS methods address this issue in different ways. It remains to be seen which of these approaches, or even some combination, will end up being the optimal method. However, the optimal method may also depend on the type of system under study.

The dissertation consists of four articles and a preceding overview part that presents a more introductory discussion and a review. The following chapter reviews the basics of electronic structure calculations, especially as they relate to the work presented in the articles. Chapter 3 reviews the basic concepts of potential energy surfaces, how to navigate on these surfaces and estimate both the rate of reaction between two states on the energy surface, using both a classical and quantum mechanical description of the atomic nuclei. In Chapter 4, the methodologies and results presented in Articles I–IV are summarized.

2 Electronic structure calculations

In this chapter the first step of the Born-Oppenheimer approximation, the solution of the electronic structure, is addressed. The standard methodology in electronic structure calculations can be divided into two classes of methods, namely wave-function based methods (see section 2.1) and density functional theory (DFT) based methods (see section 2.2). DFT was used as the electronic structure method of choice for the majority of computations presented in Articles I-III. While in Article IV, the accuracy of both DFT and wave-function based calculations is evaluated and compared for a particularly challenging molecular cation. Most of the electronic structure calculations were carried out using the ORCA software. (Neese, 2018)

2.1 Wavefunction based methods

The time-independent Schrödinger equation is given by (Griffiths and Schroeter, 2018)

$$\hat{H}\psi = E\psi \tag{1}$$

where ψ is the wavefunction of a quantum mechanical system, $\hat{H}$ is the Hamiltonian operator and E is the energy of the system. For a system with N electrons and M atom nuclei, the Hamiltonian is expressed in atomic units ($\hbar = m_e = 1$) as

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \frac{1}{2}\sum_{A=1}^{M}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}\frac{1}{r_{ij}} + \sum_{A}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \tag{2}$$

where the integers i and j are the indices for the N electrons and A, B are the indices for the M nuclei. Z_A is the nuclear charge of atom A. The first two terms correspond to the kinetic energy of the electrons and nuclei, respectively. The attractive interaction between the electrons and the atom nuclei is given by the third term. The fourth term gives the challenging repulsive interaction between the electrons and the fifth term corresponds to the repulsive interaction between the atom nuclei. The nuclei are at least three orders of magnitude more massive than the electrons. Therefore, in computational chemistry, the common approximation is to regard the position of the nuclei as being fixed, while the wavefunction for the electrons is calculated, i.e. the electronic Schrödinger equation becomes a parametrized function of the atom coordinates. Then, in a subsequent step, the dynamics of the atom nuclei is addressed. This two step approach rests on the Born-Oppenheimer approximation (Born and Oppenheimer, 1927), where the electronic Hamiltonian is written as,

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}V_{\text{ext}}(\mathbf{r}_i) + \sum_{i=1}^{N}\sum_{j>i}\frac{1}{r_{ij}} + C \tag{3}$$

where C is the constant inter-nuclear repulsion term. The attractive electron-nucleus interaction in eq. 2 is now written in the form of an external potential. Nevertheless, despite the simplification using the Born-Oppenheimer picture, the Schrödinger equation remains impossible to solve exactly for systems containing several electrons. This stems from the fact that the third term in eq. 3 couples together the motion of the electrons in a complicated way. Without this term, the Hamiltonian would simply become a sum of single electron Hamiltonian operators and the N electronic wavefunction would become a product of N one-electron wavefunctions, so-called orbitals. Since electrons are Fermions, the wavefunction has to be antisymmetric with respect to exchange of indices of the electrons so a Slater determinant is constructed instead of a simple product of orbitals. In order to decouple the motion of the electrons and simplify the calculation, a mean field approximation can be made where each electron is subject only to the average influence of the other electrons. This is done by assuming a form for the wavefunction that is valid only for independent electrons, a single Slater determinant, and then carry out a variational minimization of the expectation value of the energy. This is equivalent to having each electron be subject to an effective potential $U_i(\mathbf{r}_i)$ so the Hamiltonian is reduced to

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}V_{\text{ext}}(\mathbf{r}_i) + \sum_{i=1}^{N}U_i(\mathbf{r}_i). \tag{4}$$

This approach is referred to as the Hartree-Fock (HF) approximation (Slater, 1928). Being a mean field approximation, the HF approximation neglects to a large extent correlation between the electrons. However, some correlation between electrons of same spin is included, i.e. two electrons cannot occupy the same point in space as a direct results of the antisymmetry principle. Therefore, the error in the estimate of the electronic energy of a system obtained by HF is typically defined as the correlation energy. For many systems, HF can give a good result for the structure of molecules, but typically the binding energy calculated as the difference between the energy of a molecule and the separated atoms (or molecular fragments) is typically too small and vibrational frequencies of molecules tend to be too large. Even though HF is able to account for most of the total electronic energy of a chemical system, or about 99% of the energy(Jensen, 2017), it is not sufficiently accurate to describe many chemical phenomena. Therefore, various methods have been developed that build on the HF solution to reach higher accuracy. These methods are referred to as post-HF methods and include e.g. configuration interaction (Nesbet, 1955), Møller-Plesset perturbation theory (Møller and Plesset, 1934) and coupled cluster theory (Bartlett, 1989). The higher-level of theory methods can often provide close agreement to experimental data and are, therefore, often used to obtain accurate benchmark values. The configuration interaction method is in principle exact, i.e. it can give the exact solution to the non-relativistic Schrödinger equation. However, such high accuracy calculations involve computational effort that grows rapidly as the number of electrons increases and are, therefore, limited to small systems with at most a few tens of electrons. The development of more efficient algorithms and approximations that scale better with system size is an active field of study.

2.2 Density functional theory

An alternative to the wavefunction theory based methods is a method based on the electron density. This method is referred to as density functional theory (DFT). The reader is referred to Refs. (Becke, 2014; Burke and Wagner, 2013; Perdew and Ruzsinszky, 2010; Perdew et al., 2009) for a more detailed discussion about DFT.

Thomas and Fermi used the electron density to approximate the electronic structure of atoms (Thomas, 1927; Fermi, 1927) and Dirac derived the functional form of the exchange energy as a function of the electron density of homogeneous electron gas (Dirac, 1930). This is referred to as the Thomas-Fermi model and forms the basis of orbital-free density functional theory. (Witt et al., 2018) The basic Thomas-Fermi model is crude and does not lead to an accurate description of physical systems. Since then the development of a different and formally exact way of writing the energy of an electronic system as a functional of the total electron density has been developed and is referred to as density functional theory (DFT). (Hohenberg and Kohn, 1964; Kohn and Sham, 1965) In DFT, it is possible to carry out calculations on much larger systems than can otherwise be treated with the more accurate wavefunction methods, i.e. the post HF methods. However, it is not clear what corrections need to be applied to a DFT calculation, since the approximations that are made in practice are not as well defined, as for the wavefunction approaches.

The electron density

The Hohenberg-Kohn theorem proofs that the external potential is uniquely determined by the electron density of a chemical system in its ground state, i.e. if it exists. (Hohenberg and Kohn, 1964) In other words, the electron density contains all the information about the system necessary to determine the external potential and thereby the full Hamiltonian of the system. An integral of the electron density over all space yields the total number of electrons, N, which allows, in principle, for the kinetic energy term and the electron-electron repulsion term to be determined. The only unknown term in the electronic Hamiltonian that remains is then the external potential. The location of the atom nuclei and the respective nuclear charge can be determined by cusps in the electron density. The energy due to the external potential can be written exactly in terms of the density as,

$$V_{\text{ext}}[n] = \int n(\mathbf{r}) v_{\text{ext}}(\mathbf{r}) \mathrm{d}^3\mathbf{r}. \tag{5}$$

The classical Coloumb average inter-electronic potential, often referred to as the Hartree potential, constitutes a large part of the electron-electron repulsion, and is defined in terms of the density as,

$$U = \iint \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} \mathrm{d}^3\mathbf{r} \mathrm{d}^3\mathbf{r}' \tag{6}$$

In fact, Hohenberg and Kohn showed that the ground-state energy of the system obeys a variational principle in terms of the density, i.e. any trial electron density n will give an

upper bound of the true ground state energy E_0

$$E_0[n_0] \leq T[n] + V_{\text{ext}}[n] + V_{\text{ee}}[n] \tag{7}$$

and the equality will only hold when $n = n_0$, where n_0 is true ground-state electron density. In principle, with the exact density functional, that involves the extraction of N and construction of the electron-electron repulsion, the kinetic energy and the external potential from the electron density, will allow for the solution of the Schrödinger equation, where the highly complex, correlated multi-electron wavefunction is not needed. However, as discussed below, this is only a dream and approximation have to be made.

The Kohn-Sham equations

The main disadvantage of the Thomas-Fermi model and the orbital-free DFT methods is the inaccurate representation of the electronic kinetic energy. (Jensen, 2017) The approach taken by Kohn and Sham to obtain an improved estimate of the kinetic energy that goes beyond the TF model, is to introduce a fictitious non-interacting electron reference system that has the same total electron density as the real system.(Kohn and Sham, 1965) This non-interacting electron system can be described with a product of orbitals, i.e. one electron wavefunctions and are referred to as Kohn-Sham (KS) orbitals. The KS orbitals are related to the density by the expression

$$n = \sum_{i}^{N} |\phi_{\text{s},i}|^2. \tag{8}$$

An estimate of the kinetic energy of the non-interacting system can then be obtained from the KS orbitals, T_{s}. This form of the kinetic energy is not exact. The kinetic energy of the true, interacting electron system will differ from this estimate, but the difference is typically believed to be small.

In Kohn-Sham DFT, the electron-electron interaction is decomposed into the Hartree potential (i.e. the classical Coulomb interaction between the electrons) and a remainder term that includes the contributions to the energy from correlation effects. The estimate of the classical Coulomb interaction from the total electron density, defined in eq. 6, includes self-interaction, as can be seen most easily for a system of a single electron where this estimate gives non-zero positive value. Therefore, the correction to the estimate of the kinetic energy, a correction to remove the self-interaction in the classical Coulomb interaction as well as the missing exchange and correlation energy are all represented by a remaining contribution to the functional commonly referred to as the exchange-correlation term, E_{xc}. The functional form of the exchange-correlation potential is written by,

$$v_{\text{xc}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}}{\delta n(\mathbf{r})} \tag{9}$$

and the one-electron effective KS potential v_{eff},

$$v_{\text{eff}} = V_{\text{ext}}(\mathbf{r}) + \int \frac{n(\mathbf{r})}{|\mathbf{r}' - \mathbf{r}|} \mathrm{d}^3\mathbf{r} + \frac{\delta E_{xc}}{\delta n(\mathbf{r})} \tag{10}$$

The variational minimization of the energy gives a set of coupled equations, known as the KS equations, for the orbitals,

$$\left[-\frac{1}{2}\nabla^2 + v_{\text{eff}}\right]\phi_{\text{s},i}(\mathbf{r}) = \varepsilon_i \phi_{\text{s},i}(\mathbf{r}) \tag{11}$$

where ε_i is the energy of the i-th KS orbital. The effective potential is dependent on the electron density which is constructed from the KS orbitals. Hence, the KS equations need to be solved iteratively. To begin with an initial guess of the orbitals is constructed. The effective potential is computed and solved to generate a new set of orbitals. The next iteration is then started from the new orbitals. This procedure is continued until a set of self-consistent solutions is acquired, i.e. changes to the KS orbitals or the orbital energy between two subsequent cycles are minimal. The orbitals that are obtained by solving the KS equations become for all practical purposes the same as the orbitals used to evaluate the effective potential.

In order to make an effective use of the KS approach, an expression is needed for the E_{xc} term. In modern DFT development, most of the effort is directed towards improved approximation to this term, referred to as exchange-correlation functionals. The large number of functionals that have been developed, some tailored to specific types of systems, makes this field rather bewildering. The typical strategy to categorize the plethora of available functionals is to use the metaphor of Jacob's ladder (Perdew and Schmidt, 2001). At the first rung of Jacob's ladder lies the fundamental local density approximation (LDA) (Dirac, 1930; Ceperley and Alder, 1980) and as one ascends the ladder, both the complexity and computational effort of the functionals typically increases. At the top of the ladder the exact functional is obtained and KS-DFT becomes exact. This is, of course, only a dream.

In the LDA, the E_{xc} is taken to be that of homogeneous electron gas with density equal to $n(\mathbf{r})$ and can be written as,

$$E_{\text{xc}}^{\text{LDA}} = \int \varepsilon_{\text{xc}}^{\text{hom}}[n] n(\mathbf{r}) \mathrm{d}\mathbf{r} \tag{12}$$

In practice, the exchange-correlation energy is often partitioned into two contributions (formally ignoring then the correction to kinetic energy and self-interaction correction which should also be included)

$$E_{\text{xc}} = E_{\text{x}} + E_{\text{c}} \tag{13}$$

where the form of the exchange energy E_{x} is given by the analytical expression, (Dirac, 1930)

$$E_{\text{x}}[n] = \frac{-3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \int n(\mathbf{r})^{\frac{4}{3}} \mathrm{d}\mathbf{r}. \tag{14}$$

and E_{c} is parametrized to reproduce numerical quantum Monte Carlo simulations. (Ceperley and Alder, 1980)

The LDA is at the lowest rung of Jacob's ladder and is the simplest density functional approximation. It and can provide reasonably accurate results for systems with slowly

varying electron density (such as metals). In LDA, the exchange energy is overestimated, while the correlation contribution is underestimated. This results in an accidental cancellation of errors and is the reason for the success of LDA. For most chemical systems of interest, the LDA is typically too approximate. For open shell systems the local spin density approximation (LSDA) (Vosko et al., 1980) should be used. It includes dependency on the spin densities, $n\uparrow$ and $n\downarrow$, where $n = n\uparrow + n\downarrow$. All methods discussed in the following are equally valid for the spin density, but the spin notation will be neglected for simplicity.

At the second level of Jacob's ladder lies the generalized gradient approximation (GGA). In some sense this approximation scheme can be thought of as a next step in an expansion in terms of spatial derivatives of the electron density. GGA adds to the LDA method a dependency on the gradient at each point $\mathbf{r}$, thereby making the functional semi-local. The form of the exchange-correlation energy in GGA can be expressed as

$$E_{\mathrm{xc}}^{\mathrm{GGA}} = \int \varepsilon_{\mathrm{xc}}^{\mathrm{hom}}[n(\mathbf{r}), \nabla(\mathbf{r})]n(\mathbf{r})\mathrm{d}\mathbf{r}. \tag{15}$$

The added complexity beyond LDA results in considerably better description of molecules, e.g. binding energy and energy barriers for transitions. The most widely used GGA functionals are BLYP (Becke, 1988; Lee et al., 1988; Gill et al., 1992) and PBE (Perdew et al., 1996a). The effect of the gradient is included with an exchange enhancement function. Its functional form is not known from first principles but some limits and bounds have been derived. While these kinds of calculations are often referred to as *ab initio* or *first principles*, many choices of functional form and parameter values have in fact been made based on comparison of calculated results with high level wavefunction calculations of small systems and even experimental data. Experience has shown that GGA still overestimates the strength of chemical bonds and underestimates energy barriers for transitions, although not as strongly as LDA. Moreover, the GGA functionals tend to favor delocalized electronic states over localized states.

At the third rung of Jacob's ladder lie the meta-GGA functionals. In these methods a dependency on the kinetic energy density is added to the exchange-correlation term, i.e., the sum over the Laplacian of all occupied KS orbitals. Therefore, in comparison to the aforementioned GGA functionals, the meta GGA involve one higher spatial derivative of the electron density and thereby require a more detailed representation of the orbitals. The most commonly used meta-GGA functional is probably the TPSS functional (Tao et al., 2003). However, the more recent SCAN functional (Sun et al., 2016) is becoming used more and more.

On the next rung beyond meta-GGA is the hybrid functionals. In these methods a fraction of the exact exchange contribution is included in the exchange-correlation term, analogous to Hartree-Fock. With the difference being that the Slater determinant is formed from the KS orbitals (instead of the canonical HF orbitals). The exchange-correlation energy is expressed as,

$$E_{\mathrm{xc}}^{\mathrm{hyb}} = E_{\mathrm{xc}}^{\mathrm{LDA}} + a_0(E_{\mathrm{x}}^{\mathrm{HF}} - E_{\mathrm{x}}^{\mathrm{LDA}}) + a_{\mathrm{x}}(E_{\mathrm{x}}^{\mathrm{GGA}} - E_{\mathrm{x}}^{\mathrm{LDA}}) + a_{\mathrm{c}}(E_{\mathrm{c}}^{\mathrm{GGA}} - E_{\mathrm{c}}^{\mathrm{LDA}}) \tag{16}$$

where $a_{\mathrm{x}}, a_{\mathrm{c}}$ and a_0 are parameters that are fitted to some set of data. The most commonly used hybrid functional, B3LYP functional is of this form (Stephens et al., 1994). The weight of the exact exchange term in B3LYP is $a_0 = 0.2$. Another popular hybrid

functional is the PBE0 (Perdew et al., 1996b) functional. It has a simpler form with just one adjustable parameter,

$$E_{\mathrm{xc}}^{\mathrm{hyb}} = E_{\mathrm{xc}}^{\mathrm{GGA}} + a_0(E_{\mathrm{x}}^{\mathrm{HF}} - E_{\mathrm{x}}^{\mathrm{GGA}}). \tag{17}$$

where $a_0 = 0.25$. The non-local exchange term is known to significantly improve values obtained for bond energy, structure of molecules and vibrational frequencies. However, it also increases the computational effort needed for the calculation. The constant a_0 is often treated as an adjustable parameter depending on the system and molecular property under investigation. In some cases, e.g. for calculations of charge transfers, accurate results require a value of a_0 around 0.5. While, for other cases values of a_0 that are smaller than 0.2 give good results, e.g. for thermochemical data. The HF and GGA approximations tend to have errors of opposite sign, for example HF giving too weak bonds while GGA gives too strong bonds. Also, HF favors localized electronic states while GGA favors delocalized states. It is, therefore, not surprising that some mix of the two can be made to give good results, but the optimal recipe depends on the system under study.

Unlike the wavefunction theory approaches, it is not clear what approximation actually is being made at each level of Jacob's ladder and what exactly the required correction should be in order to obtain the exact results. The advantage of DFT, however, is that it is applicable to larger systems for a given level of computational resources.

Computational considerations of density functional theory

In application of both KS-DFT and wavefunction methods, some representation of the orbitals is needed. For this purpose, the unknown KS orbitals are given as linear combinations of a set of known functions, known as a basis set. A KS orbital is then expanded in a set of K basis functions η_k,

$$\phi_i = \sum_{k=1}^{K} c_{k,i}\eta_k \tag{18}$$

where the $c_{k,i}$ are expansion coefficients. Inserting this expression in eq. 11 and manipulating the expression yields,

$$F^{\mathrm{KS}}C = SC\varepsilon \tag{19}$$

where F^{KS} is the Kohn-Sham matrix of size $K \times K$ and S is the overlap matrix of the same size. Their form can be found in standard textbooks (Koch and Holthausen, 2001). This treatment and the resulting equation is analogous to that of the Hartree-Fock-Roothaan equation (Roothaan, 1951).

In modern electronic structure software, most of the basis sets are made from either contracted Gaussian functions or plane waves. The contracted Gaussian basis sets are constructed to mimic atomic orbitals, i.e. Slater-type orbitals. A Gaussian basis function is usually expressed in the form,

$$\psi = Nx^{l_x}y^{l_y}z^{l_z}\exp(-\xi r^2) \tag{20}$$

where l_x, l_y, l_z are constants related to the angular momentum and hence determine the type of the orbital. N is a normalization factor. Many basis sets based on Gaussian functions have been developed throughout the years. The most notable ones are the Pople style basis (Pople et al., 1987), correlation-consistent basis set (Dunning, 1989) and the Ahlrichs type basis (Schäfer et al., 1992). The different basis sets vary in the number of functions used per atomic orbital and the addition of polarization and diffuse functions beyond the functions that describe filled atomic orbitals. As the space spanned by the basis set is increased, the accuracy and computational effort of an electronic structure calculation increases. The accuracy is improved by increasing the number of linearly independent basis functions, and the computational effort increases at the same time. For a large enough basis set, the calculated energy of the system will converge, i.e. a basis set limit is reached.

Shortcomings of density functional theory

One of the more severe shortcomings of (commonly used) density functionals is the lack of dispersion interaction. The dispersion interaction is a weak long range interaction that arises from correlated induced multipoles caused by fluctuations. A semi-local functional cannot describe such long range interactions where the electron clouds are barely overlapping.

A simple and computationally efficient empirical approach to include dispersion effects are the DFT-D methods, (Grimme, 2004, 2006) where the total energy is written as

$$E_{\text{tot}} = E_{\text{DFT}} + E_{\text{disp}} \tag{21}$$

and E_{disp} is calculated by adding up pair-wise atomic dispersion contributions

$$E_{\text{disp}} = -\sum_{A,B} f(r_{AB}, A, B) \frac{C_6^{AB}}{r_{AB}^6}. \tag{22}$$

C_{AB} is the dispersion coefficient for species A and B. f is a damping function used to correct the unphysical behavior caused by divergence of the inverse distance at small inter-nuclear distances. The damping function and C_6^{AB} need to be compatible with the exchange-correlation functional used. This dispersion-correction scheme is, of course, approximate since the dispersion interaction is in reality not pairwise additive.

Another shortcoming of the semi-local exchange-correlation functionals is the self interaction error. As discussed above, the Hartree energy in eq. 6 describes the interaction of the total electron density with itself. This is best illustrated with a system of only one electron where the KS-DFT Hartree energy is non-zero and positive. While the exchange-correlation term of the functional should contain a self-interaction term which gives a contribution of opposite sign and thereby cancels out the error (as is the case in HF), a semi-local approximation to the exchange-correlation term cannot provide such a cancellation since the self-interaction error in the classical Coulomb interaction is highly non-local, involving a double integral over all space. Mathematically, an elimination of the self-interaction error is not possible in a semi-local functional. As a result of this error GGA functionals have systemic problems such as an underestimation

of band gaps, underestimation of energy barriers and incorrect balance between localized and delocalized states. As mentioned above, hybrid functionals can be made to give more accurate results by tuning the fraction of exact exchange and sometimes this is described in terms of a cancellation of the self-interaction error. The self-interaction error is a limitation of the semi-local functional form commonly used for the exchange-correlation term, but a more accurate density functional of a more general form could in principle be developed without introducing self interaction (Steckel, 2009).

3 Calculations of chemical reactions

In this chapter, theory and methods used in the second step of the Born-Oppenheimer approximation, i.e. the treatment of the atomic nuclei, are reviewed. The focus is on methods used for the identification and computation of the mechanism and rate of chemical reactions. The chapter starts with a review on potential energy surfaces and important features thereof. In the following section, the well established transition state theory (TST) for estimating the classical rate of thermally activated transitions is described. In section 3.3, the, practical, harmonic approximation to TST is introduced, within this approximation the estimation of the rate of reaction requires only two points on the energy surface to be known, namely, the reactant energy minimum and a connected first order saddle point. In sections 3.4 and 3.5, methods used to navigate on the energy surface and to locate minimum energy paths and first order saddle points are reviewed. More precisely, in section 3.4, three of the most widely-used double-ended methods are reviewed, where the common objective is to identify a minimum energy path (and the corresponding highest energy, first order, saddle point) for a transition between a given reactant and product states. Then, in section 3.5, two classes of commonly used single-ended methods are reviewed, where the objective is to directly and rigorously locate a first order saddle point from a given initial configuration. Finally, in section 3.6, semi-classical rate theories to incorporate quantum mechanical description of the atomic nuclei are reviewed.

3.1 Potential energy surfaces

The concept of potential energy surfaces (PES) arises from the Born-Oppenheimer approximation, where the nuclear and electronic degrees of freedom of a molecular system are decoupled. Then, the energy of a system (for a given electronic state) can be expressed as a parametric function of the position of the N-atoms in the system, or in Cartesian coordinates as

$$E(\mathbf{r}) = E(x_1, x_2, x_3, ..., x_{3N-1}, x_{3N}) \tag{23}$$

The set of $3N$-coordinates is referred to as a configuration. The force acting on individual atom in the system is given by the negative gradient of the energy with respect to the atom coordinates,

$$\mathbf{F} = -\left[\frac{\partial E}{\partial x_1}, \frac{\partial E}{\partial x_2}, ..., \frac{\partial E}{\partial x_{3N}}\right] \tag{24}$$

The Hessian matrix, i.e. the matrix of second derivative of the energy with respect to the atom coordinates,

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 E}{\partial x_1^2} & \frac{\partial^2 E}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 E}{\partial x_1 \partial x_{3N}} \\ \frac{\partial^2 E}{\partial x_2 \partial x_1} & \frac{\partial^2 E}{\partial^2 x_2} & \cdots & \frac{\partial^2 E}{\partial x_2 \partial x_{3N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 E}{\partial x_N \partial x_1} & \frac{\partial^2 E}{\partial x_N \partial x_2} & \cdots & \frac{\partial^2 E}{\partial^2 x_{3N}} \end{bmatrix} \tag{25}$$

describes the curvature of the energy landscape and is often referred to as the force constant matrix. In principle, the energy surface can be constructed by computing the energy, E, for all possible configurations of the system, $\mathbf{r}$. Of course, in practice, this is only possible for low dimensional systems, e.g. a system of 10 dimensions on a small grid of 10 points would require 10^{10} energy computations! To illustrate some of the features of an energy surface, a model two-dimensional energy surface is shown in Fig. 3.1.

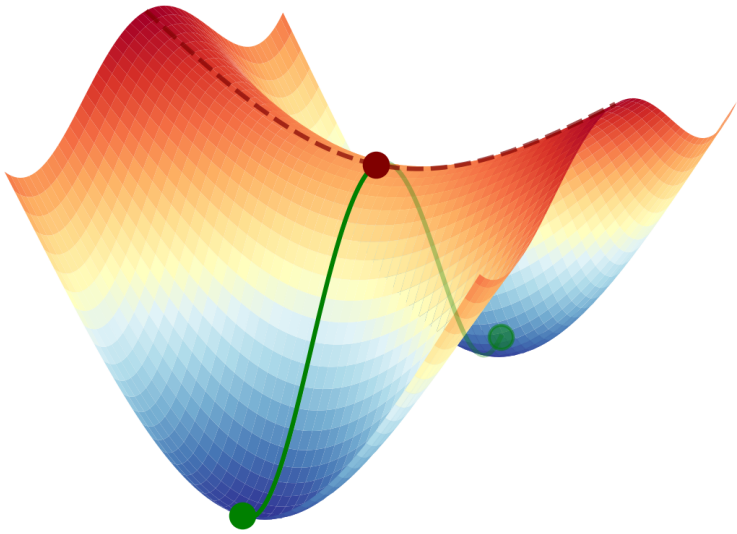


Figure 3.1. A model two-dimensional energy surface, $E(\mathbf{r}) = E(x,y)$. On the surface, there are two energy minima (shown by green circles) separated by an energy ridge (shown by a dashed red line). The minimum of energy along the energy ridge is a first order saddle point (shown by a red point). A path of minimum energy goes through the saddle point and connects the two energy minima.

An energy minimum on the PES corresponds to a (meta)-stable configuration of a molecule, often referred to as a equilibrium structure. An energy minimum is a stationary point on the energy surface (i.e., $\mathbf{F} = 0$) for which the Hessian matrix is positive semi-definite, i.e., all eigenvalues (λ_i for $i \in [1,2,...,3N]$) of the Hessian matrix are positive

or zero. For non-linear molecular systems, $3N-6$ (and $3N-5$ for linear molecules) of the eigenvectors of the mass-weighted Hessian matrix, $\frac{1}{\sqrt{m_i}\sqrt{m_j}}H_{ij}$, correspond to the normal modes of vibration about the equilibrium structure (also referred to as vibrational modes). The frequency of vibration for particular mode is proportional to the square root of the associated eigenvalue. The remaining six modes are the zero-modes (i.e. $\lambda_i = 0$ for $i \in [1,...,6]$) that correspond to the rotational and translational degrees of freedom of the system.

The energy surface expressed by eq. 23 is written as a function of the Cartesian coordinates of the system, which is a simple and unambiguous manner of defining the position of a set points (or atoms). It is, however, important to note that the energy surface can be written in terms of any complete set of coordinates, for example in the normal modes of vibration or by a set of internal coordinates. The internal coordinates are likely to be the most chemically intuitive way to represent a molecular configuration, where the relative position of the atoms is expressed entirely in terms of bonds, angles and torsions (or dihedrals). For more complex molecular configurations (e.g. cyclic geometries) more than 3N-6 coordinates are needed to obtain a complete description of the molecular configuration. This is referred to as redundant internal coordinates. Then, linear combinations of redundant internal coordinates, or delocalized non-redundant internal coordinates, may offer a more suitable representation. Baker et al. (1996) One of the main advantages of using internal coordinates is that coupling between individual coordinate components is highly reduced compared to e.g Cartesian coordinates. This serves to accelerate optimization methods and exploration of energy surfaces. However, the use of internal coordinates makes the methods and discussion more complicated. Therefore, in the remainder of the dissertation Cartesian coordinates are assumed, unless specified otherwise.

It is relatively simple to locate an energy minimum on the PES, starting from a trial configuration ($\mathbf{r}_0$), the atom force $\mathbf{F}$ can simply be followed downhill on the energy surface to the local minimum,

$$\mathbf{r}_{i+1} = \mathbf{r}_i + \alpha \mathbf{F}(\mathbf{r}_i) \tag{26}$$

where i is the current optimization step and α is some adjustable parameter used to control the size of the step. This method is referred to as the steepest descent method and is known to be very inefficient. (Nocedal and Wright, 2006) All configurations that lead to the same energy minimum when a steepest-descent direction is followed on the energy surface define the potential well surrounding the energy minimum. Collectively, the minimum and the potential well are defined together as a state of the system. The lower the energy of a state, the more likely the system is to be found in that state according to Boltzmann statistics, i.e. $P(\mathbf{r}) \propto \exp(-\beta E(\mathbf{r}))$ where $\beta = 1/k_{\mathrm{B}}T$.

Most commonly used minimization methods to locate energy minima are based on the local quadratic approximation (LQA), (Schlegel, 2003) where the energy surface around a configuration $\mathbf{r}$ is approximated by a second-order expansion of the surface,

$$E(\mathbf{r}+\Delta\mathbf{r}) \approx E(\mathbf{r}) + \mathbf{F}^T\Delta\mathbf{r} + \frac{1}{2}\Delta\mathbf{r}^T\mathbf{H}\Delta\mathbf{r} \tag{27}$$

In Fig. 3.2 an illustration of a LQA around the energy minimum of a non-quadratic (i.e. third and higher order derivatives are non-zero) one-dimensional energy surface is

shown. The step Δr to the stationary point of the LQA is given by

$$\Delta \mathbf{r} = \mathbf{H}^{-1}\mathbf{F}. \tag{28}$$

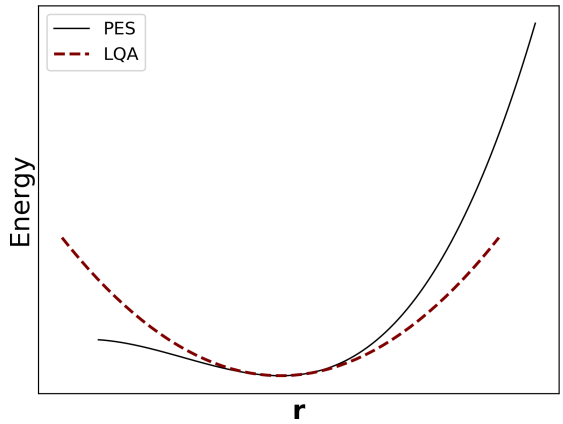


Figure 3.2. The local quadratic expansion (shown by red dashed line) around the energy minimum of a model non-quadratic, one-dimensional, energy surface, $E(\mathbf{r}) = E(x)$.

If the starting configuration, $\mathbf{r}_0$, is near a minimum, the $\mathbf{H}$ is positive-definite and the step will be downhill in energy along all degrees of freedom. As most energy surfaces are not quadratic, the quadratic expansion of the energy surface and stepping to the minimum of the expansion becomes an iterative procedure to locate an energy minimum, i.e. $\mathbf{r}_{i+1} = \mathbf{r}_i + \Delta\mathbf{r}$ where $\Delta\mathbf{r}$ is computed from eq. 28. This method is known as the Newton-Raphson method. Note that if $\mathbf{H} = \mathbf{I}$, the method reduces to the steepest-descent method discussed above. If the starting configuration of Newton-Raphson is in region on the energy surface that yields a Hessian matrix with one or more negative eigenvalues, the step will be uphill in energy along the direction of the eigenvectors corresponding to the negative eigenvalues. In this case, an energy minimum will be not located. But, instead, a different kind of stationary point on the energy surface might be identified (as discussed below). The hyper-surface on the energy surface, where the lowest eigenvalue of the Hessian matrix goes from being positive to negative ($\lambda_1 = 0$) is known as the inflation surface.

In Newton-Rapshon, the Hessian matrix is needed for each iterative step in the search for a stationary point. However, the evaluation and diagonalization of the $3N$ x $3N$ dimensional Hessian matrix is computationally demanding or even prohibitive for large systems, in particular if electronic structure methods are being used. Sometimes analytical Hessian calculations are available but still the computational cost is significantly larger than that of the energy and force computations. In this case, quasi-Newton methods can be used, where an approximate Hessian matrix is used instead. The search for a stationary point is then either started from the exact or an approximation to the Hessian matrix. After which, changes to the (inverse) Hessian matrix, $\Delta\mathbf{H}$, are

approximated in each step of the optimization,

$$\mathbf{H}_i = \mathbf{H}_{i-1} + \Delta\mathbf{H} \tag{29}$$

where the change expected to occur in the Hessian matrix needs to fulfill the following condition

$$\Delta\mathbf{F} = -\mathbf{H}_i\Delta\mathbf{r} \tag{30}$$

There are numerous way to fulfill this condition, the most popular one is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update, (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970)

$$\Delta\mathbf{H}^{\mathrm{BFGS}} = -\frac{\Delta\mathbf{F}\Delta\mathbf{F}^T}{\Delta\mathbf{F}\Delta\mathbf{r}} - \frac{\mathbf{H}_{i-1}\Delta\mathbf{r}\Delta\mathbf{r}^T\mathbf{H}_{i-1}}{\Delta\mathbf{r}^T\mathbf{H}_{i-1}\Delta\mathbf{r}}. \tag{31}$$

The BFGS is a suitable update method for approximating the Hessian matrix in energy minimizations since it ensures that the Hessian matrix is positive definite. For larger optimization problems, it is more suitable to use the limited memory version of BFGS, referred to as L-BFGS. (Nocedal, 1980) The storage of the full Hessian matrix is avoided in L-BFGS. Instead only a memory of M previous optimization steps and forces is kept.

When searching for stationary points which are not characterized by a positive-definite Hessian matrix, e.g. saddle points (discussed below), other Hessian updates need to be used. These updates should not enforce the Hessian to be positive-definite. This includes the symmetric rank one (SR1) update, (Murtagh, 1970)

$$\Delta\mathbf{H}^{\mathrm{SR1}} = -\frac{(\Delta\mathbf{F} - \mathbf{H}_{i-1}\Delta\mathbf{r})(\Delta\mathbf{F} - \mathbf{H}_{i-1}\Delta\mathbf{r})^T}{(\Delta\mathbf{F} - \mathbf{H}_{i-1}\Delta\mathbf{r})^T\Delta\mathbf{r}} \tag{32}$$

and the Powell-symmetric Broyden (PSB), (Dennis Jr and Schnabel, 1996)

$$\Delta\mathbf{H}^{\mathrm{PSB}} = -\frac{(\Delta\mathbf{F} - \mathbf{H}_{i-1}\Delta\mathbf{r})\Delta\mathbf{r}^T + \Delta\mathbf{r}(\Delta\mathbf{F} - \mathbf{H}_{i-1}\Delta\mathbf{r})^T}{\Delta\mathbf{r}^T\Delta\mathbf{r}} \tag{33}$$

$$+ \frac{\Delta\mathbf{r}^T(\Delta\mathbf{F} - \mathbf{H}_{i-1}\Delta\mathbf{r}))\Delta\mathbf{r}\Delta\mathbf{r}^T}{(\Delta\mathbf{r}^T\Delta\mathbf{r})^2}. \tag{34}$$

Bofill proposed that it was better to use a linear combination of SR1 and PSB and introduced a mixing parameter θ, given by (Bofill, 1994)

$$\Delta\mathbf{H}^{\mathrm{Bofill}} = \theta\Delta\mathbf{H}^{\mathrm{SR1}} + (1-\theta)\Delta\mathbf{H}^{\mathrm{PSB}} \tag{35}$$

$$\theta = \frac{((\Delta\mathbf{F} - \mathbf{H}_{i-1}\Delta\mathbf{r})^T\Delta\mathbf{r})^2}{|\Delta\mathbf{F} - \mathbf{H}_{i-1}\Delta\mathbf{r}|^2|\mathbf{r}|^2} \tag{36}$$

For the initial Hessian matrix, an inexpensive estimate is typically used. The simplest being a scaled identify matrix, $\mathbf{H}_0 = \alpha\mathbf{I}$. For molecular systems, the initial Hessian is often empirically constructed by accounting for the connectivity of the atoms and the flexibility of the molecule. A positive definite diagonal Hessian matrix is then computed using empirical estimates of second derivatives (or force constants) of the bonds, angles and dihedrals that involve connected atoms. This method was first proposed by Schlegel (Bernhard Schlegel, 1984) and revised later by Fischer and

Almlöf (Fischer and Almlöf, 1992). The model, empirical, Hessians are known to give satisfactory results. However, when possible it is advantageous to use more intensive (but approximate) methods to estimate the initial Hessian, e.g. by using force-fields that are fitted to the given system or semi-empirical electronic structure methods. (Jensen, 2017) It is important that the properties of the initial Hessian matrix reflect the problem at hand, e.g. be positive-definite for energy minimization.

The Newton-Raphson and quasi-Newton methods may lead to a large optimization step being taken that exceeds the boundaries of the LQA. For example, near the boundary of the inflation surface or if the starting configuration is far from the stationary point and the surface is non-quadratic. (Schlegel, 2011) This may bring the system to unreasonable regions of the energy surface and can even lead to poor convergence. It is, therefore, essential to control the step-size of Newton-Rapshon and quasi-Newton calculations. The simplest approach is to use a constant trust-radius, Δ_{trust}, where if the optimization step exceeds the trust-radius, the size of the step is simply scaled back to be equal to Δ_{trust}. In this case, if too small Δ_{trust} is used, the optimization becomes inefficient and may resemble the steepest descent method. Instead, adaptive scaling of Δ_{trust} can be used. (Schlegel, 2011; Dennis Jr and Schnabel, 1996) For minimization problems, a better alternative is to use an inexact line-search method to select the step-size such that the objective function (e.g., the energy) and the gradient along the search direction are ensured to be reduced in each step, i.e. by fulfilling the Wolfe conditions. (Nocedal and Wright, 2006) The third option is to use Lagrange multipliers to constrain the step to be equal to a given step-size, Δ. (Fletcher, 2013; Dennis Jr and Schnabel, 1996)

Adjacent states (i.e. the energy minimum and the surrounding potential well) are separated by energy ridges on the energy surface. Dips in energy along the energy ridges correspond to saddle points on the surface. For an n-th order saddle point, the energy is at a maximum along n degrees of freedom and a minimum along all other degrees of freedom. In other words, the Hessian matrix of a n-th order saddle point has n negative eigenvalues and the corresponding eigenvectors are referred to as imaginary vibrational modes.

First order saddle points are at a maximum in energy along one direction and minimum along all other $3N-1$ degrees of freedom. These points are of particular importance as they can represent dynamical bottle-necks of the transition from one state to another. In other words, in the dynamical evolution of the system, majority of the transition events would occur in vicinity to the first order saddle points. Therefore, identification of these points becomes essential for the computation of the rate of reaction, see sections 3.2 and 3.3. At a first order saddle point, the Hessian matrix has one and only one negative eigenvalue and the corresponding eigenvector is referred to as the unstable-mode. Therefore, in order to locate a first order saddle point on the energy surface, the energy needs to be maximized along one degree of freedom corresponding to the correct reaction coordinate (e.g. the unstable mode), while being minimized along all other degrees of freedom. Typically, this reaction coordinate is not known *a priori* and needs to be identified. For example, to locate the saddle point on the model surface presented in Fig. 3.1 from any given starting configuration, the energy needs to be maximized in the x-direction, while being simultaneously minimized along the y-direction. For a high-dimensional molecular energy surface, the identification of first order saddle points presents a significant computational challenge and the development

of reliable and efficient methods to locate saddle points is an active field of research.

As previously stated, if a Newton-Raphson calculation is started from an initial configuration where the Hessian matrix has a single negative eigenvalue, i.e. $\lambda_0 < 0$, the displacement of the system will be uphill in energy along the corresponding eigenmode, $\mathbf{q}_0$. In principle, the Newton-Rapshon method can then be used to identify first order saddle points, but this only applies if the trial configuration is already in very close proximity to a first order saddle point, where the eigenmode corresponding to the negative eigenvalue aligns with the unstable mode. Such a configuration is very hard (or even impossible) to obtain for a high-dimensional molecular system when guided only by chemical intuition. Moreover, even if the initial guess configuration is started from a region on the energy surface beyond the inflation surface and with a single negative eigenvalue of the Hessian matrix, the Newton-Raphson method may still not necessarily converge to a first order saddle point. (Peters, 2017)

An early and important development of saddle point search methods was the Cerjan-Miller method. (Cerjan and Miller, 1981) In this method, the eigenmode of the Hessian matrix corresponding to the lowest eigenvalue (λ_1) is followed uphill in energy from the energy minimum towards the saddle point using a modified Newton-Rapshon method. (Fletcher, 2013; Dennis Jr and Schnabel, 1996) The Cerjan-Miller method uses the LQA expansion of the energy surface, but subject to the constraint of a constant step-size, Δ. In other words, by defining the Lagrangian,

$$L(\Delta\mathbf{r},\Delta) = E(\mathbf{r}) + \mathbf{F}^T\Delta\mathbf{r} + \frac{1}{2}\Delta\mathbf{r}^T\mathbf{H}\Delta\mathbf{r} - \frac{\lambda}{2}(||\mathbf{r}|| - \Delta)^2 \tag{37}$$

the step to the stationary solution of the Lagrangian is given by,

$$\Delta\mathbf{r} = (\mathbf{H} - \lambda\mathbf{I})^{-1}\mathbf{F} \tag{38}$$

where λ is a shift parameter used to shift the Hessian matrix to ensure that it has always one negative eigenvalue, i.e. $\lambda_0 < 0$. Then, the system is always displaced, uphill in energy along $\mathbf{q}_1$ and down-hill in energy along all other modes according to eq. 38, towards a first order saddle point. The value of the shift parameter, λ, must be selected to be larger than the lowest eigenvalue, λ_1, and smaller than the second lowest eigenvalue, λ_2 of the Hessian matrix. The use of a shift parameter allows the Cerjan-Miller method to be started from a configuration within the energy minimum basin or even in regions far beyond the inflation surface where the Hessian matrix has more than one negative eigenvalue. The Cerjan-Miller method hence eliminates the need for having an unreasonably accurate initial guess configuration to identify a saddle point, as is needed in the Newton-Rapshon method. (Peters, 2017) However, the saddle point to which the Cerjan-Miller method converges is highly dependent on the initial guess configuration. The idea of maximizing the energy along a selected eigenmode sparked the development of the eigenvector-following (Baker, 1986) and minimum-mode following methods (Henkelman and Jónsson, 1999). These methods are collectively referred to as single-ended methods and are reviewed in section 3.5.

The unstable mode at a first order saddle point is aligned in the direction normal to the energy ridge, pointing towards the two energy minima located on opposite sides of the saddle point. The two connected energy minima can, therefore, be readily identified from a first order saddle point by simply following the steepest-descent path downhill,

from the saddle point along the forward and backward directions of the unstable mode. This type of a steepest descent path is referred to as the minimum energy path (MEP) and satisfies

$$(1-\hat{\tau}(s)\hat{\tau}(s))\,\mathbf{F}(\phi(\mathbf{s}))=\mathbf{F}(\phi(\mathbf{s}))^{\perp}=0 \tag{39}$$

where $\phi(s)$ describes the MEP and s is a parameter that traces out the path, dependent on the arclength of the curve. $\tau(s)$ is the tangent to the path. In other words, all components of the atom force that are orthogonal to the MEP are zero. Hence, the MEP is the lowest energy (local) path on the energy surface that connects the reactant energy minimum to a product energy minimum. This renders the MEP a natural choice to describe the transition of the system from one state to another, i.e. a reaction coordinate. (Marcus, 1966) The MEP is important for e.g. building in anharmonic effects (Henkelman et al., 2005), including quantum corrections (Truhlar et al., 1996) and defining a reaction path Hamiltonian. (Miller et al., 1980; Peters, 2017) For an illustration of the MEP and the energy curve along the MEP, see Figs. 3.1 and 3.3. If mass-weighted coordinates (i.e. $\sqrt{\mathbf{m}}\mathbf{r}$) are used in the steepest-descent calculation, the MEP is also referred to as the internal reaction coordinate. (Fukui, 1981) Methods that follow the steepest-descent path downhill from a saddle point, to trace out the MEP rigorously require a significant computational effort. (Ishida et al., 1977; Hratchian and Schlegel, 2005)

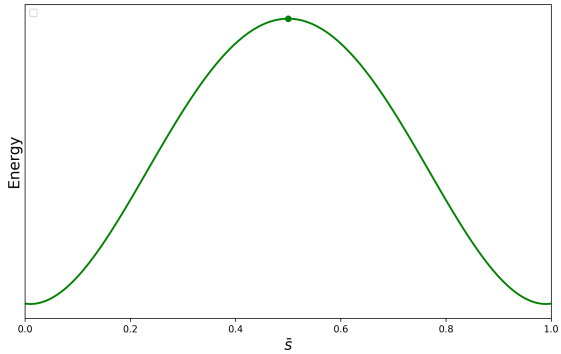


Figure 3.3. Potential energy along the minimum energy path, i.e. $E(\phi(s))$, for the model energy surface presented in Fig. 3.1. The parameter s is normalized according to the arclength of the MEP, i.e. $\bar{s}\in[0,1]$. The energy maximum along the minimum energy path (denoted by a green circle at $\bar{s}=0.5$) is the first order saddle point on the energy surface.

Alternatively, methods that search approximately for MEPs are more efficient and do not require the correct saddle point to be known, *a priori*. In fact, they are useful for identifying the correct first order saddle point that characterizes the transition between a given initial and final states. Also referred to as reactant and product states. These methods are, hence, often referred to as 'double-ended search methods'. By searching for a MEP that connects the given reactant and product states, the calculation may locate unknown intermediate extrema along the MEP, e.g. energy maxima and minima that correspond to first order saddle points and stable intermediates on the energy

surface. (McPherson et al., 2016) The highest energy maximum along the MEP is the most important point and corresponds to the first order saddle point that describes the given transition. The rise in energy along the MEP is the activation energy of the transition. Moreover, by searching for MEPs alternative or unexpected reaction mechanism for a given reactant and product state may be obtained. (Henkelman et al., 2006) In the double-ended methods, a path that lies between the reactant and product energy minima is represented by a discrete set of system configurations (or 'images' of the system), i.e. the path is defined by $\mathbf{R} = [\mathbf{r}_0, \mathbf{r}_1, ..., \mathbf{r}_{M-1}]$ using M images, where $\mathbf{r}_0$ and $\mathbf{r}_{M-1}$ are typically fixed at the reactant and product energy minima, respectively. The objective of the double ended optimization is then to iteratively shift the set of discrete images to the MEP.

In early developments of double-ended methods, Elber and Karplus, (Elber and Karplus, 1987) attempted to locate a MEP for a given reactant and product energy minima by minimizing the objective function,

$$S(\mathbf{R}) = S([\mathbf{r}_0, \mathbf{r}_1, ..., \mathbf{r}_{M-1}]) = \frac{1}{L}\sum_{j=1}^{M-2} \Delta l_j E(\mathbf{r}_j) + \lambda \sum_{j=1}^{M-1} \left(\Delta l_j - \sqrt{\frac{L^2}{M}} \right) \tag{40}$$

$$L = \sum_{j=1}^{M-1} \Delta l_j \tag{41}$$

$$\Delta l_j = ||\mathbf{r}_j - \mathbf{r}_{j-1}|| \tag{42}$$

with respect to the configuration of the $M-2$ intermediate images. For any objective function $S(\mathbf{R})$, the minimization can be done by following the steepest-descent direction, i.e. $-\frac{\partial S}{\partial \mathbf{r}_i}$, down on the objective function surface to obtain a stationary path, $\delta S = 0$. In the method of Elber and Karplus, the energy is minimized for each intermediate image, as well as, the separation between a pair of images according to the first term in eq.40. In the second term, spring forces are introduced in an attempt to control the distribution of images along the path, where the natural length of the spring interaction is the average separation between a pair of images. However, in this method, the path is not able to converge to the MEP. Instead, the intermediate images 'cut corners' in regions of high curvature along the MEP. (Olender and Elber, 1997) Moreover, the intermediate images accumulate near the energy minima and also kinks are readily formed along the path as the path loops in on itself. (Czerminski and Elber, 1990; Jónsson et al., 1998) Later developments of these methods attempted to fix the corner-cutting problem by involving second derivatives (Olender and Elber, 1997) and the aggregation problem by introducing a repulsive term acting between the images. (Czerminski and Elber, 1990). Later, Choi and Elber presented the locally updated planes method (Choi and Elber, 1991) where the images are shifted to the MEP by using an estimate of the local tangent to the path to minimize the energy of the images only in the direction normal to the path, i.e. within a $3N-1$ dimensional hyperplane. In the locally updates planes method, no measure was included to control the distribution of images along the path. This leads to an uneven distribution of images and poor estimates of the tangent to the path and could even lead to a discontinuity in the path. (Jónsson et al., 1998) The early double-ended methods mentioned here and the need to have reliable methods to locate approximate MEPs have led to the continual development of modern double ended methods which

are reviewed in section 3.4. This includes the nudged elastic band method (NEB) (Mills et al., 1995; Jónsson et al., 1998), the string method (E et al., 2002) and the growing string method (GS) (Peters et al., 2004). The development and improvement of these methods remains an active field of study and is the primary focus of this dissertation.

3.2 Transition state theory

The dynamics of a system that has reached equilibrium in a given state, referred to here as the reactant state (R), are characterized by a very large number of small fluctuations (e.g., vibrational and rotational motions of the system) back and forth in the energy basin, around the equilibrium structure. Occasionally large excursions of the system from the equilibrium structure may occur, caused by coupling to the surroundings (or heat bath). If enough energy is input along the correct degrees of freedom the system may escape the reactant state to enter a product state P. The waiting time between such reactive events, $R \rightarrow P$, is typically many orders of magnitude longer than the relaxation time of the system in the stable energy basin. This allows the system to relax and reach thermal equilibrium in the new state before the next reactive event occurs. There is, therefore, a clear separation of time-scale between the relaxation and the reactive events. This justifies the use of equilibrium statistical mechanics to compute the rate of escape from a given state. The most successful and widely used theory in chemical reaction kinetics is transition state theory (TST) (Eyring, 1935; Wigner, 1938; Evans and Polanyi, 1938; Truhlar et al., 1996; Peters, 2017).

In TST, a dividing surface in configuration space needs to be defined. It should be defined in such a way that it separates the reactant state from all possible product states, i.e. a classical trajectory going from the reactant to any product state must cross the dividing surface. This is illustrated in Fig. 3.4 using a model two-dimensional surface. The dividing surface is a $D-1$ dimensional hypersurface, where D is the number of degrees of freedom of the system. The dividing surface is given an infinitesimal width (σ) and is then referred to as the transition state, denoted by $\ddagger$ in the following discussion. In TST, it is assumed that the transition state species are populated as in an equilibrium with the reactant state. Moreover, it is also assumed that if the system makes it to the transition state and the velocity is pointing away from the reactant state, a reaction is guaranteed to occur and the system enters and then thermalizes in a product state. That is, as the system crosses the transition state towards the product state, it can not be reflected back to enter the reactant state again. Given an appropriate selection of the transition state dividing surface and a classical description of the atom nuclei, the generalized TST rate (Wigner, 1938; Evans and Polanyi, 1938) is computed as the quotient of the flux ($j^{\dagger}$) through the transition state and the equilibrium population of the reactant state $\langle R \rangle$, expressed as (Peters, 2017)

$$k^{\mathrm{TST}} = j^{\ddagger} \frac{1}{\langle R \rangle} = \nu \exp\left(-\beta \Delta F\right) \tag{43}$$

where $\Delta F = F^{\ddagger} - F^{R}$ is the free-energy barrier measured from the ensemble free energy of the transition state and the reactant state. As always, $\beta = 1/k_{\mathrm{B}}T$. ν is a kinetic prefactor accounting for the averaged absolute velocity at the transition state in the

direction of a product state, $\langle|v^{\ddagger,\perp}|\rangle$. For a more in-depth discussion and derivation of the TST rate expression in eq. 43, the reader is referred to Ref. Peters (2017).

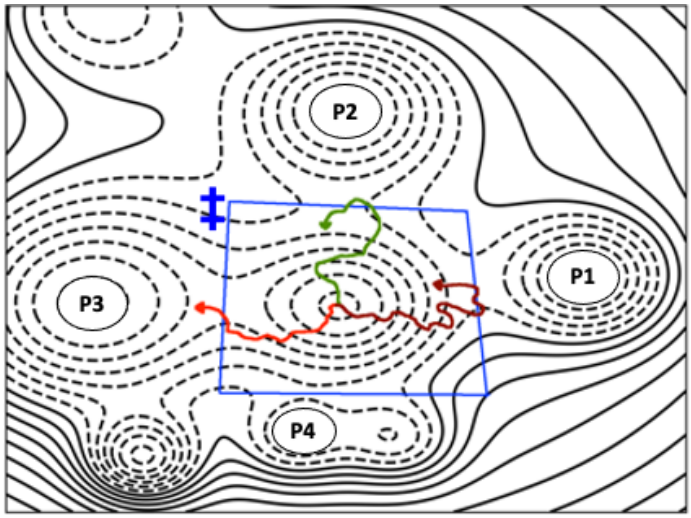


Figure 3.4. A model two-dimensional energy surface to illustrate some of the key concept of transition state theory. The surface is composed of seven minima of varying depth and size. The reactant energy minimum is selected to be at the center of the surface. Possible transitions out of the reactant state include the P_1, P_2, P_3 and P_4 states. The other energy minima are not directly connected to the reactant state and are not marked. The transition state surface (‡) is selected to be a mosaic of hyperplanes and placed along the energy ridges that separates the reactant from possible products (shown by blue), centered on the respective first order saddle points. Three hypothetical classical trajectories, originating from the center of the reactant state, are shown for illustrative purposes. In red, a reactive trajectory is shown for the transition $R \rightarrow P_3$ (red). In green, a trajectory is shown that goes towards P_2 but is then reflected back to the reactant state. The third trajectory, shown in dark red, re-crosses the transition state near P_1 two times and is then reflected back to the reactant state.

With an accurately selected transition state, generalized TST leads to a reasonably accurate estimate of the rate constant. The assumption of no re-crossing is an approximation. In reality, some dynamical trajectories will be reflected back after crossing the transition state to re-enter the reactant state. These re-crossings can occur because of e.g. large curvature of the energy surface and the coupling of the system to the heat bath. In Fig. 3.4, three hypothetical classical trajectories are shown, one reactive and two trajectories that cross the transition state one and two times before getting reflected back to the reactant state. As a result, the TST rate is guaranteed to yield an upper-bound of the rate constant leading to a variational principle for the placement of the dividing surface. In other words, the transition state that yields the lowest TST rate constant is the optimal one. However, finding the optimal transition state manually, in a high-dimensional space, is impossible. Instead, a variational optimization needs to be

carried out. The simplest approach is to use a single hyperplanar dividing surface and variationally optimize the location and orientation of the surface such that k^{TST} is minimized. (Jóhannesson and Jónsson, 2001) Note that a single dividing surface is typically not sufficient to represent the transition state, instead either a curved surface (Ciccotti et al., 1995) or a mosaic of hyperplanes needs to be used. (Bligaard and Jónsson, 2005) At present, finding the optimal transition state dividing surface is a computationally infeasible task for high dimensional chemical systems.

Given a reasonable placement of the dividing surface, the k^{TST} rate can be corrected for dynamical inaccuracies by the WKE method. (Wigner, 1938; Keck, 1967; Eyring, 1935) In this approach, short classical trajectories are started from the transition state and the number of re-crossings over the transition state are counted to obtain a dynamical correction factor (κ). (Keck, 1967; Voter and Doll, 1985) The corrected TST rate can then be expressed as,

$$k^{\mathrm{corr}} = \kappa k^{\mathrm{TST}} \tag{44}$$

In general, the more accurate the placement of the dividing surface, the smaller the number of trajectories is needed to obtain a statistically converged correction factor. The trajectories also serve to identify the possible product states connected to the reactant state.

Transition state theory is an incredibly powerful theory which can be used to approximate the canonical rate of reaction from one state to another, i.e. when the system is in thermal equilibrium with a heat bath. However, for unimolecular reacations, at low pressure, the Rice-Ramsperger-Kassel-Marcus (RRKM) theory should instead be used to estimate the micro-canonical reaction rate. (Marcus and Rice, 1951; Marcus, 1952)

3.3 Harmonic transition state theory

Computation of the rate of reaction with harmonic TST (HTST) is both straightforward and practical, even when electronic structure methods are used to describe the energy surface. For this reason, HTST is likely to be the most commonly used form of TST. In HTST, the transition state is simply chosen to be the ensemble of first order saddle points that lie on the energy ridge surrounding the given reactant state. This assumption is valid if the saddle points represent dynamical bottle-necks for transitions of the system from the reactant to the product states. There are typically many more saddle points on the energy surface than there are energy minima and the number of saddle points surrounding a single energy minimum can easily be enormous. The main challenge of HTST is, therefore, to locate all of the relevant saddle points.

For example to determine the reaction mechanism and when computing and comparing the relative HTST rates of two (or more) elementary reaction steps to determine the ratio of products formed. The correct saddle points need to be identified. These are the lowest energy saddle points along the energy ridge that correspond to a transition between the given reactant and desired product states. For this task, the correct saddle points can be identified by locating the relevant MEP for the given reactant and product states, see section 3.4. A more challenging application of HTST is to estimate the rate at which the system moves from one state to another, i.e. the rate of escape from a

given state. This is quantity of crucial importance for long timescale simulations on the evolution of a system using adaptive kinetic Monte Carlo. (Henkelman and Jónsson, 2001) The computation of the rate of escape is a challenging computational task. In principle, it involves locating all of the first order saddle points that encircle the given state. However, in practice, only the low energy saddle points on the energy ridge are relevant and need to be identified, as they describe the the most likely transitions out of the given state, according to Boltzmann statistics.

In the following two different formulations of HTST are presented, namely, that of Eyring (Eyring, 1935; Peters, 2017) and Vineyard (Vineyard, 1957). Starting from a slightly different but equivalent expression to the generalized TST rate. The k^{TST} rate constant can be written in terms of the partition functions for the transition state ($Q^{\ddagger}$) and the reactant state (Q^{R}),

$$k^{\mathrm{TST}} = \frac{1}{2}\frac{\langle|v^{\ddagger,\perp}|\rangle}{\sigma}\frac{Q^{\ddagger}}{Q^{R}} \tag{45}$$

where the quotient $\frac{Q^{\ddagger}}{Q^{R}}$ is the probability of the system making it to the transition state, i.e. to a first order saddle point in HTST. The factor $\frac{\langle|v^{\ddagger,\perp}|\rangle}{\sigma}$ is the rate of crossing the transition state. The average absolute velocity normal to the transition state (along the reaction coordinate) can simply be estimated from the Maxwell distribution of velocities,

$$\langle|v^{\ddagger,\perp}|\rangle = \sqrt{\frac{2k_{\mathrm{B}}T}{\pi\mu}} \tag{46}$$

where μ is an effective mass along the reaction coordinate. The factor $\frac{1}{2}$ in eq. 45 accounts for the fact that only half of the trajectories are moving in the direction of the product state.

3.3.1 Eyring's HTST

In the Eyring formalism, the factor corresponding to translation along the reaction coordinate is separated from $Q^{\ddagger}$ and treated as a free translation, or $\frac{\sigma}{\lambda^{\ddagger}} = \frac{\sigma\sqrt{2\pi\mu k_{\mathrm{B}}T}}{h}$. Combining this result with eqs. 45 and 46, the kinetic prefactor equates to $\frac{k_{\mathrm{B}}T}{h}$ and the expression for the rate constant becomes

$$k^{\mathrm{TST}} = \frac{k_{\mathrm{B}}T}{h}\frac{Q^{\ddagger}}{Q^{R}} \tag{47}$$

For calculations of the canonical rate of molecular reactions, the rigid-rotor harmonic oscillator approximation can be used to approximate the partition functions of the transition state and reactant. In this approximation, the translation, rotational, vibrational and electronic degrees of freedom are considered separable and the partition function Q can be written as,

$$Q = Q^{\mathrm{trans}}Q^{\mathrm{rot}}Q^{\mathrm{vib}}Q^{\mathrm{elec}} \tag{48}$$

where the typical individual partition functions (for polyatomic non-linear molecule) are given as, (Jensen, 2017)

$$Q^{\text{trans}} = \left(\frac{2\pi M k_{\text{B}} T}{h^2}\right)^{3/2} V \tag{49}$$

$$Q^{\text{rot}} = \frac{\sqrt{\pi}}{\sigma_{\text{sym}}} \left(\frac{8\pi^2 k_{\text{B}} T}{h^2}\right)^{(3/2)} \sqrt{I_1 I_2 I_3} \tag{50}$$

$$Q^{\text{vib}} = \prod_{i=1}^{3N-6(7)} \frac{\exp(-\beta\hbar\omega_i)}{1-\exp(-\beta\hbar\omega_i)} \tag{51}$$

$$Q^{\text{elec}} = \sum_{i=0}^{\infty} g_i \exp(-\beta E_i) \tag{52}$$

In this approximation, the vibrational degrees of freedom are treated as quantum mechanical harmonic oscillators (see eq. 52) and hence the name 'harmonic' TST. Then, k^{TST} becomes k^{HTST}. ω_i is the vibrational frequency of the 'i-th' vibrational mode $\mathbf{q}_i$, given by $\omega_i = \sqrt{\frac{\lambda_i}{\mu}}$. The product in Q^{vib} runs over the $3N-6$ vibrational modes in the reactant state. While, for the transition state (or saddle point) the product runs over $3N-7$ vibrational modes. Because, the unstable mode, i.e. the vibrational mode that corresponds to the imaginary vibrational frequency at the saddle point, is omitted from the product. In fact, this degree of freedom was factored out of $Q^{\ddagger}$ in the derivation of k^{TST} and is treated as a free translation. For the rotational partition function, the principle moments of inertia ($\prod_{i=1}^{3}\sqrt{I_i}$) can be different for the reactant and saddle point configurations even for unimolecular reactions. σ_{sym} is the symmetry number and accounts for the number of equivalent ways by which the transition may occur. For the electronic partition function, it is typically sufficient to include only the ground state electronic contribution to the partition function. Since the energy difference between the ground-state and higher energy excited states is often large and thus have negligible contributions. In k^{HTST}, given the relevant saddle points have been located, the rate of reaction is readily computed using only a single Hessian calculation per relevant saddle point and one Hessian calculation at the given reactant state. The possible product states can be obtained by following the steepest descent path down from the saddle points in the direction of the unstable mode.

3.3.2 Vineyard's HTST

The partition functions of eq. 45 can be written in terms of configurational integrals,

$$Q^R = Z^R = \int_R \exp(-\beta E(\mathbf{r}))\,\mathrm{d}\mathbf{r} \tag{53}$$

$$Q^{\ddagger} = Z^{\ddagger} = \sigma \int_{\ddagger} \exp(-\beta E(\mathbf{r}))\,\mathrm{d}\mathbf{r} \tag{54}$$

where the integration is carried out over the reactant region R and the transition state region $\ddagger$ of configuration space, respectively. Strictly speaking, to define the state of system in phase-space, i.e. $E^{\text{tot}}(\mathbf{r},\mathbf{v}) = E^{\text{kin}}(\mathbf{v}) + E^{\text{pot}}(\mathbf{r})$, both the coordinates $\mathbf{r}$

and velocities $\mathbf{v}$ of all of the atoms are needed. However, the distribution function of the system, $P(\mathbf{r},\mathbf{v})\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{v} = A\exp(-\beta E(\mathbf{r},\mathbf{v}))\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{v}$, can be factored into the individual distribution of velocities and coordinates, $P = P_{\mathbf{r}}P_{\mathbf{v}}$. The velocity distribution is simply the Maxwell distribution which is the same everywhere in configuration space and can thus be factored out.

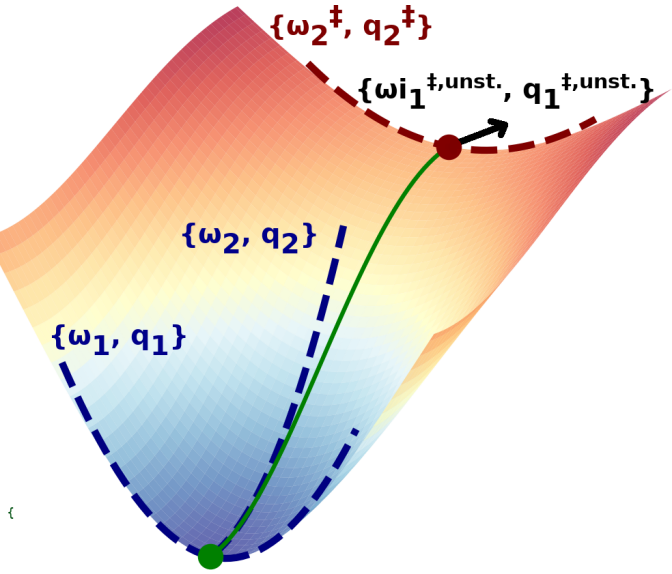


Figure 3.5. Harmonic expansion along the vibrational modes of the reactant energy minimum ($\mathbf{q}_1$,$\mathbf{q}_2$) and saddle point ($\mathbf{q}_2^{\ddagger}$) on the model energy surface from Fig. 3.1. The unstable mode, $\mathbf{q}_1^{\ddagger,unst.}$, pointing along the reaction coordinate (shown in green) is not included in the expansion. The (angular) vibrational frequency of this mode is imaginary. In Vineyard's HTST the prefactor is proportional to the ratio of the corresponding vibrational frequencies, or $\nu = \omega_1\omega_2/\omega_2^{\ddagger}$. This effective frequency is often interpreted as the attempt frequency of the system to cross the transition state.

For systems of small vibrations around an equilibrium structure, e.g. atoms or molecules reacting in and on the surfaces of solids, the theory of small oscillations becomes applicable. (Vineyard, 1957) In Vineyard HTST, the configuration integrals are evaluated by expanding the energy surface to second order in the vibrational normal modes around the reactant energy minimum and the first order saddle point, i.e.

$$E^{\ddagger}(\mathbf{q}^{\ddagger}) = E^{\mathrm{SP}} + \frac{1}{2}\sum_{i=1}^{D-1}\lambda_i^{\ddagger}\mathbf{q}_i^{\ddagger 2} \tag{55}$$

$$E^{\mathrm{R}}(\mathbf{q}^{R}) = E^{\mathrm{R}} + \frac{1}{2}\sum_{i=1}^{D}\lambda_i^{\mathrm{R}}\mathbf{q}_i^{\mathrm{R}2} \tag{56}$$

where D is the number of vibrational modes of the system. The expansion is not carried out along the degree of freedom corresponding to the unstable mode, as is shown in

Fig. 3.5. This leads to the harmonic approximation to the TST rate constant (Vineyard, 1957),

$$k^{\mathrm{HTST}} = \frac{1}{2\pi} \frac{\prod_i^D \omega_i^R}{\prod_i^{D-1} \omega_i^{\ddagger}} \exp\left(-\beta \Delta E\right) \tag{57}$$

where $\Delta E = E^{\ddagger}$-E^R is the energy difference of the saddle point and the reactant energy minimum. The prefactor is written in terms of the classical vibrational partition functions. If the quantum mechanical vibrational partition functions are used instead, the method is referred to as quasi-quantum HTST (or qq-HTST). Again, the imaginary frequency is omitted from the product over vibrational frequencies at the transition state. If the system includes translation and rotational degrees of freedom, they have to be introduced to eq.57 separately as defined in the previous section.

The form of the Vineyard HTST expression is in agreement to the empirical Arrhenius dependence of the rate constant on temperature,

$$k = A \exp\left(-E_{\mathrm{a}}/k_{\mathrm{B}}T\right) \tag{58}$$

and allows for a physical interpretation of the Arrhenius parameters. The activation energy, E_{a}, can be interpreted as the height of the potential energy barrier. The prefactor, A, is thus related to the width of the potential energy well around the minimum and the energy valley at the saddle point.

The HTST approximation can give an accurate estimate of the rate constant at ambient temperature intervals, if the energy surface is smooth and the first order saddle points are separated by regions of high enough energy for the harmonic approximation to accurately represent the energy surface in vicinity to the reactant sand saddle point stationary points. However, for systems characterized by rugged energy landscapes, e.g., reactions in enzymes and solution, the assumptions of HTST are not likely to be valid. For such reactions, TST (or WKE) will still give an accurate estimate of the rate. That is, if the assumptions of TST hold and an accurate placement of the dividing surface. However, at low enough temperature, where a classical description of the atom nuclei is not correct, the HTST (and also TST) rate become inaccurate and nuclear quantum effects need to be accounted for. This can be done within the semi-classical approximation of HTST, see section 3.6.

3.4 Double ended methods

Double ended saddle point search methods represents a class of path optimization methods used to locate and resolve minimum energy paths for a given reactant and product energy minima. Consequentially, the highest energy, first order, saddle that characterizes the reaction is also identified. The nudged elastic band method (NEB) (Mills et al., 1995; Jónsson et al., 1998) is a well-established and widely used method of this class. (Ásgeirsson and Jónsson, 2020) In this chapter, the NEB method and variants thereof are reviewed in section 3.4.1 and 3.4.2, respectively. A large portion of the body of work presented in this dissertation are developments and applications of the NEB method to (i) calculate optimal tunneling paths (see Article I) and (ii) to locate first order saddle points of molecular reactions (see Article II and III). Other commonly

used double ended methods are presented in sections 3.4.3 and 3.4.4, namely the string method (E et al., 2002) and the growing-string method (GSM) (Peters et al., 2004).

3.4.1 Climbing image nudged elastic band method

In NEB, a path between the reactant and product energy minima is discretized by creating a set of M points and arranging them in such a way as to trace out an MEP on the energy surface that connect the two minima. The discretization points are referred to as 'images' of the system. An image, denoted by i, is specified by the set of $3N$ coordinates of all atoms in the system, $\mathbf{r}_i$. The two end-points of the path are typically fixed at the reactant and product energy minimum. A point in path space is, therefore, $3N\,x\,(M-2)$ dimensional. The task is to first construct a reasonable initial path between the two endpoints and then apply the NEB combined with some optimization method to iteratively shift the images to the MEP.

The images are shifted downhill on the energy surface in a direction obtained from the atom force. However, only the component acting perpendicular to the path is used to modify the shape of the path, (Jónsson et al., 1998). Following the full atom force would lead to accumulation of images near the local energy minima. Therefore, a force projection is required based on an approximation to the local tangent to the path. While it seems more accurate to estimate the tangent at a given image, i, from the coordinates of atoms at the two adjacent images, $i-1$ and $i+1$, it turns out to be numerically more stable to use only the coordinates of the neighboring image that has the higher energy. In other words, using an (upward) finite difference instead of central finite difference approximation. (Henkelman and Jónsson, 2000) For images at an energy extremum along the path, the tangent is approximated by the energy weighted average of the two line segments connecting it to its two neighbors. (Henkelman and Jónsson, 2000) Letting the normalized tangent to the path be denoted as $\hat{\tau}_i$, the perpendicular component of the force acting on image i is

$$\mathbf{F}_i^{\perp} = \mathbf{F}_i - (\mathbf{F}_i \cdot \hat{\tau}_i)\,\hat{\tau}_i. \tag{59}$$

where $\mathbf{F}_i$ is the atom force for image i. By iteratively displacing the images in the direction of $\mathbf{F}_i^{\perp}$ until the perpendicular component of the atom force vanishes, places the images on the MEP. For each optimization step, the energy and atom forces of all movable images need to be computed. The computations of the energy and gradient are the most computationally intensive part of a NEB calculation, especially when used in conjunction with electronic structure methods. However, one of the main advantages of the NEB method is that the computations are readily carried out simultaneously, using parallel or distributed computing.

It is necessary to control the distribution of the images along the path. Otherwise, the images may slide down in energy and aggregate near the local energy minima. (Jónsson et al., 1998) This is achieved by incorporating harmonic springs between pairs of adjacent images. Several different formulations of the spring interaction have been presented, see Refs. Jónsson et al. (1998); Henkelman and Jónsson (2000); Kolsbjerg et al. (2016); Maras et al. (2017); Röder and Wales (2020). In this work, the distance-

based formulation of the springs is used, (Henkelman and Jónsson, 2000)

$$\mathbf{F}_i^{\mathrm{sp},\parallel} = (k_i|\mathbf{r}_{i+1} - \mathbf{r}_i| - k_{i-1}|\mathbf{r}_i - \mathbf{r}_{i-1}|) \cdot \hat{\tau}_i. \tag{60}$$

The spring force should only act tangential to the path, as the perpendicular component of the spring force serves to straighten out the path and induce 'corner-cutting' of the images in regions of large curvature. In other words, the images get pushed off the MEP. (Jónsson et al., 1998) The perpendicular component of the spring force has been used to aid in NEB calculations of long paths where the density of images is low compared to the local curvature of the energy surface. (Jónsson et al., 1998; Trygubenko and Wales, 2004; Sheppard et al., 2008; Kolsbjerg et al., 2016; Maras et al., 2016)

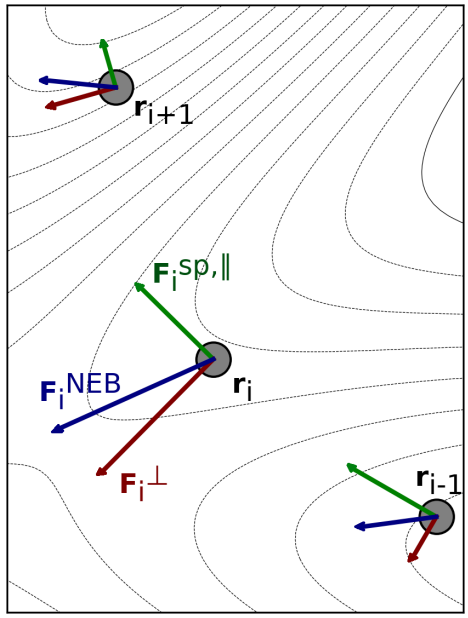


Figure 3.6. Snapshot of a NEB calculation, illustrating the different force components that are used to displace the path to the MEP. Namely, the component of the spring force acting parallel to the path $\mathbf{F}^{sp,\parallel}$ *shown in green, the component of the atom force acting normal to the path* $\mathbf{F}^{\perp}$ *shown in red and the effective NEB force* $\mathbf{F}^{NEB}$ *shown in blue for three intermediate images* $\mathbf{r}_{i+1}, \mathbf{r_i}, \mathbf{r_{i+1}}$.

The spring constants, k, are most commonly chosen to have the same value for all segments of the path resulting in an equidistant placement images along the path, i.e. an even distribution. The distribution of the images along the path can also be controlled using constraints (instead of a penalty function) based on an estimate of the total length of the path (E et al., 2002; Maras et al., 2017). In principle, different values of the spring constant can be used for different segments of the path to produce any desired distribution of images. For example, it is possible to adaptively scale the spring constants according to the energy such that stiffer springs act between images located in

higher energy regions of the path, the so-called energy-weighted (EW) spring constants. In this scheme, the spring constant for segment i is given by (Henkelman et al., 2000)

$$k_i = \begin{cases} (1-\alpha_i)k_\mathrm{u} + \alpha_i k_\mathrm{l}, & \text{if } E_i > E_\mathrm{ref} \\ k_l, & \text{otherwise} \end{cases} \tag{61}$$

$$\alpha_i = \frac{E_\mathrm{max} - E_i}{E_\mathrm{max} - E_\mathrm{ref}} \tag{62}$$

k_u and k_l are upper- and lower-bound values of the spring constant. E_max is the energy of the highest energy image (HEI) along the path, E_i is the higher energy image of the pair of images connected by line segment i and E_ref is a reference energy chosen to be equal to the higher energy state of the reactant and product states. The energy-weighted (EW) spring constants lead to a higher density of images in the higher energy regions along the path, while other less important parts of the MEP become less well resolved. (Henkelman et al., 2000) An example of the different image distributions obtained from CI-NEB and EW-CI-NEB calculations of the MEP for the ene-reaction of 1-propylene and ethylene is shown in Fig. 3.7. Even though this method was proposed some time ago, it has (to the best of the authors knowledge) not been used in NEB calculations since, nor is it available in commonly used implementations of the method. In Article III, the energy-weighted springs are found to be essential for reliable NEB calculations of molecular reactions.

The effective force acting on image i is then given by the sum of the two force components,

$$\mathbf{F}_i^\mathrm{NEB} = \mathbf{F}_i^\perp + \mathbf{F}_i^{\mathrm{sp},\parallel}, \tag{63}$$

and defines the steepest-descent direction of NEB, i.e. $\mathbf{F}_i^\mathrm{NEB} = -\frac{\partial S}{\partial \mathbf{r}_i}$ for an unknown objective function $S(\mathbf{R})$. The effective force, perpendicular atom force and parallel spring force of eq. 63 are visualized in Fig. 3.6. The force projections decouple the displacement of the images along the path and the change in the position and shape of the path towards the MEP.

The most important part of an MEP is the highest energy point along the path, i.e. the highest energy first order saddle point connecting the reactant and product states, since the rise in energy along the MEP gives an estimate for the activation energy of the transition. However, as discussed above, the images are typically distributed evenly along the path and then there is no guarantee that an image ends up being placed in the vicinity of the highest energy saddle point, leading to an uncertain underestimate of the activation energy. To improve on this, it is possible to interpolate the path using a piecewise cubic polynomial, where both the energy and the component of the atomic force acting tangential to the path are used. (Henkelman and Jónsson, 2000) Moreover, the interpolation may reveal the presence of other extrema along the path, e.g. an intermediate minimum.

In order to obtain a more accurate estimate of the activation energy for the transition, one of the images can be forced to 'climb' uphill along the path and converge on the highest energy maximum. This image is referred to as the 'climbing image'. Generally, the HEI after a few optimization steps is chosen to be the climbing image, $i = ci$. The effective force acting on the climbing image is given by, (Henkelman et al., 2000)

$$\mathbf{F}_{ci}^\mathrm{NEB} = \mathbf{F}_{ci} - 2\left(\mathbf{F}_{ci} \cdot \hat{\tau}_{ci}\right)\hat{\tau}_{ci}. \tag{64}$$

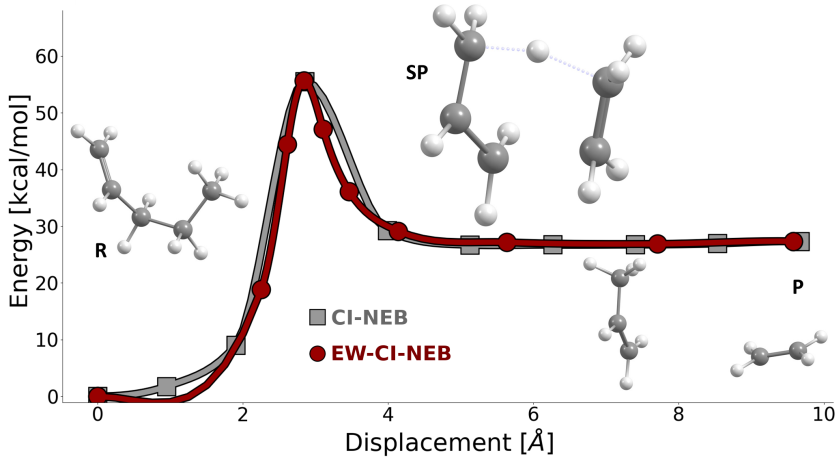


Figure 3.7. Comparison of the different image distributions obtained by CI-NEB and EW-CI-NEB calculations of the minimum energy path for the ene-reaction of 1-propylene and ethylene. The CI-NEB method uses an even distribution of images along the path, this leads to only a single image (the climbing image) being located on the actual energy barrier. In the EW-CI-NEB method, however, the images accumulate on the energy barrier. This leads to a more accurate resolution of the energy barrier in EW-CI-NEB compared to CI-NEB. But, other, irrelevant, regions of the path become more poorly resolved. The reactant (denoted by **R***), saddle point (denoted by* **SP***) and product (denoted by* **P***) configurations are shown as insets.*

Note that the CI is unaffected by the spring force and is only displaced according to the projected atom force. By inverting the force in the direction of the tangent, the force around a first order saddle point is made to mimic the force around a minimum. Therefore, ordinary minimization methods can be used to converge the climbing image on the first order saddle point. The main assumption here is that the tangent estimate at the climbing image gives an accurate estimate of the direction of the unstable mode at the saddle point. This method is referred to as the climbing-image nudged elastic band (CI-NEB) method. (Henkelman et al., 2000)

Optimization of the path

Various optimization methods can be used to zero the effective force and shift the path to the MEP, see Refs. (Sheppard et al., 2008; Herbol et al., 2017). An NEB calculation is a minimization in the $3Nx(M-2)$ dimensional path space. An illustrative CI-NEB calculation is shown for the two-dimensional Müller-Brown surface (Müller and Brown, 1979) in Fig. 3.8. The images are iteratively displaced in the direction normal to the path until they converge on the MEP, while maintaining even spacing along the path. The climbing image accurately converges to the higher energy saddle point.

Typically, a calculation is considered to be converged when the absolute maximum Cartesian component of the atom force acting perpendicular to the path, $\max(|\mathbf{F}^{\perp}|)$, is below 0.01 eV/Å for all images. However, the most important part of the MEP is the highest energy point and the estimate the tangent to the path at that point. Therefore, in practice, tighter convergence thresholds are used for CI than for the other images, e.g. $\max(|\mathbf{F}_{\mathrm{CI}}|) < 0.01$ eV/Å and $\max(|\mathbf{F}_i^{\perp}|) < 0.1$ eV/Å. An illustrative CI-NEB calculation is shown for the two-dimensional Müller-Brown surface (Müller and Brown, 1979) in Fig. 3.8.

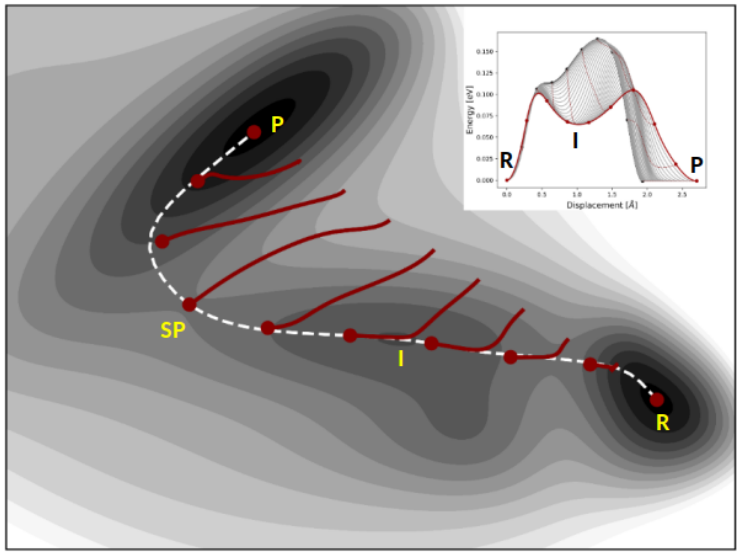


Figure 3.8. Climbing image nudged elastic band calculation on the two-dimensional Müller-Brown surface. (Müller and Brown, 1979) There are three energy minima on the surface, the reactant (denoted as **R***), intermediate (denoted as* **I***) and product (denoted as* **P***). There are also two saddle points. The higher energy saddle point is labeled by* **SP***. The calculation is started from a linear interpolation of $M = 10$ images between the reactant and product energy minima. The path is then accelerated in the normal direction, down-hill, to the minimum energy path (shown by a white dashed line) using the velocity-projection optimization method. The optimization trajectory of the images are shown by dark red curves and the final position of the images on the MEP by circles. The highest energy image converges directly to the higher energy saddle point. The optimization profile of the CI-NEB calculation is shown as an inset. In this, a cubic interpolation of the energy between the images along the MEP is plotted as a function of displacement after each optimization step.*

It is important to note that the objective function for the NEB force is not known explicitly because of the force projections. This complicates the use of optimization methods that require the use of a line-search method to determine the size of the optimization step, such as the conjugate gradient methods (Hestenes et al., 1952) and

some quasi-Newton methods (Nocedal and Wright, 2006).

In the earliest implementations of the NEB method (Mills et al., 1995; Jónsson et al., 1998), the velocity projection optimization (VPO) method was used. It is a reliable but conservative optimization method where the effective force is minimized using a quenched version of the velocity Verlet algorithm. (Andersen, 1980) The path is accelerated in the direction of the effective force by projecting the velocity along the effective force. If the velocity and force do not align (i.e. are anti-parallel) the velocity is zeroed, as this is an indication that the optimization has overshot the objective function minimum, i.e. the MEP. A parametrized extension of the VPO method has been devised and is referred to as fast inertial relaxation engine (or FIRE). (Bitzek et al., 2006) The biggest difference between the two methods, is that FIRE uses a variable time-step to accelerate the optimization more aggressively, and that only part of the velocity is projected along the effective force (i.e. first derivative), maintaining momentum in other directions.

In VPO and FIRE, only the effective force is used to displace the path to the MEP. Alternatively, more efficient optimization methods can be used, where information about the second derivative is used, as well, to bring the path to the MEP. Here, a local quadratic approximation (LQA) to the unknown NEB objective function surface is made and the system is displaced to the minimum of the quadratic approximation, using

$$\Delta\mathbf{r} = \mathbf{H}^{\mathrm{NEB}\,-1}\mathbf{F}^{\mathrm{NEB}} \tag{65}$$

where $\mathbf{H}^{\mathrm{NEB}}$ is the Hessian matrix for the path and $\mathbf{F}^{\mathrm{NEB}} = [\mathbf{F}_1^{\mathrm{NEB}}, \mathbf{F}_2^{\mathrm{NEB}}, ..., \mathbf{F}_{M-1}^{\mathrm{NEB}}]$. If no inter-image interactions were included (i.e. if $\mathbf{H}^{\mathrm{NEB}}$ is constructed solely from $\mathbf{F}$), then $\mathbf{H}^{\mathrm{NEB}}$ would be block-diagonal with the block elements equal to $\mathbf{H}_i$. However, like energy surfaces, the unknown NEB surface is not quadratic. Therefore, an iterative procedure is required to zero the atom force. Bohner et al. (Bohner et al., 2013) derived the analytical expression of $\mathbf{H}^{\mathrm{NEB}}$ and used it to achieve quadratic convergence for several simple test cases, where the analytical Hessian of the energy surface was readily available. However, the calculation and diagonalization of the second derivative of the energy function is, for most practical purposes, computationally prohibitive. To address this issue, the quasi-Newton methods can be used, where the Hessian matrix is approximated by the Hessian update methods. A good choice for NEB is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update. (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) For high-dimensional optimization problems such as NEB, the limited-memory variant of the BFGS method (Nocedal, 1980) is the more sensible choice as it scales linearly in terms of processor and memory requirements with system size. A memory of M previous steps is kept and used to construct the approximate inverse Hessian matrix. In fact, in the work of Sheppard and Henkelman, it is reported that the L-BFGS method, where $\mathbf{H}^{\mathrm{NEB}}$ includes the off-diagonal elements, is the most efficient and robust optimization method for NEB. (Sheppard et al., 2008). As previously stated, the implementation of an efficient line-search for NEB/L-BFGS is not an easily realizable task, as the objective function is unknown. This needs further investigation. Instead, the NEB/L-BFGS method is typically used with a constant trust-radius, Δ_{trust}, to prevent the optimization from taking unreasonable steps into high-energy regions on the energy surface. It has also been suggested that the L-BFGS method may reach ill-conditioned states when used with NEB. In these cases it is recommended to reset

the L-BFGS memory of previous steps. (Chill et al., 2014a,b) In Article III, a reasonable choice for Δ_{trust} is found to be of crucial importance for the reliability and efficiency of the NEB/L-BFGS method in applications to molecular reactions.

Recent developments of NEB have used machine learning approaches to reduce the number of energy and force evaluations by an order of magnitude. Both neural networks (Peterson, 2016) and Gaussian process regression (GPR) (Koistinen et al., 2017, 2019) have been applied for this purpose. In these approaches, an approximate energy surface is constructed with a machine learning model where the energy and force evaluations are used as an input, typically in tandem with an electronic structure method. Then, a CI-NEB calculation is carried out on the approximate energy surface to converge to an approximate MEP, after which electronic structure calculation are carried out for each image along the approximate MEP or possibly only for the image with the largest uncertainty estimate (in the GPR-model) of the approximate energy surface. The approximate energy surface is then refined using the new information and the CI-NEB calculation carried out again. This procedure is repeated until the MEP obtained on the approximate energy surface parallels the MEP on the actual energy surface. The GPR has the advantage (over neural networks) of having a built in error estimate that can be used to choose which image should be evaluated using the electronic structure method. The machine learning approach helps make optimal use of each electronic structure calculations, while traditional optimization methods, such as the VPO and L-BFGS method, only use the forces obtained at the current and at a few previous steps. All other force evaluations are disregarded. However, the current limitation of the GPR-NEB method is that it is only applicable to relatively small systems, as the computational resources required for such a calculation grow as the third power of the degrees of freedom and number of evaluations.

In the case of symmetric MEPs, computational resources required by the CI-NEB calculation can effectively be reduced by half. In this approach, the calculations of the energy and atom forces are only explicitly carried out for half of the images in the path and then projected onto the other half using the appropriate symmetry operations. (Mathiesen et al., 2019)

For CI-NEB calculations of finite systems, such as clusters or molecules, it is important to remove the trivial translation and rotation of the system from the available degrees of freedom during the optimization. The optimization of the path could otherwise lead to an artificial lengthening of the path involving translation and/or rotation in order to enable images to slide down from high energy regions. This can reduce the resolution of the path in regions of high energy and increase the computational effort, or even prevent the calculation from converging. A method based on quaternions has been formulated for the purpose of constraining the translation and rotation (Melander et al., 2015).

Initial path generation in NEB

The simplest method (and likely to be the most frequently) used to generate an initial path in NEB is linear interpolation in Cartesian coordinates between the reactant and product energy minima. However, the linear interpolation may lead to an initial path that is characterized by large atom forces, caused by overlapping of atoms in intermediate configurations. When the calculation of the energy and atom force is carried out using an electronic structure method, such strong overlap can slow down or even prevent the self-consistency calculation from converging. Additionally, an initial path, generated by linear interpolation, is typically located far from an MEP on the energy surface.

A better approach is to generate the initial path by taking the pairwise distances between atoms into account using the image dependent pair potential (IDPP) method (Smidstrup et al., 2014). In this method, the pairwise distances between neighboring atoms are interpolated linearly between the two energy minima and an initial path is generated to match those distances as closely as possible. Since there are many more pairwise distances than atom coordinates, the matching can only be approximate and the initial path is found by minimizing the sum of squared deviations (Smidstrup et al., 2014). In IDPP, this is achieved by defining an objective function, $S_{\kappa}^{\mathrm{IDPP}}$, and carrying out an NEB calculation on an 'effective surface' defined by the objective function,

$$S_{\kappa}^{\mathrm{IDPP}}(\mathbf{r_i}) = \sum_{A}^{M} \sum_{B>A}^{M} w(d_{AB}) \left(d_{AB}^{\kappa} - d_{AB}\right)^2 \tag{66}$$

where d_{AB} is the pairwise distance between atoms A and B for intermediate image i. d_{AB}^{κ} is the interpolated pairwise distance between atoms A and B of the same image. Here, w is a weight-function to give shorter bond distances higher weight and make unnecessary bond breaking unfavorable. The weight function is given as $w = (d_{\mathrm{AB}})^{-4}$. The NEB/IDPP calculation requires little computational effort compared to the electronic structure calculations. Typically, the resulting IDPP initial path lies closer to the MEP than a path generated by e.g., linear interpolation of Cartesian coordinates (Smidstrup et al., 2014) with much lower energy intermediate configurations. This is clear from Fig. 3.9, where the energy difference between the highest energy image and reactant energy minimum is visualized for a large benchmark set of 121 main-group molecular reactions.

Recently, a method to grow the initial path in a step-wise fashion for a NEB calculation on the effective IDPP surface has been proposed (Schmerwitz et al., 2021). This builds on the basic idea of the growing string method (Peters et al., 2004) see section 3.4.4, but is applied to the IDPP surface so as to generate an initial path for an NEB calculation. Alternatively, it is possible to avoid the problem of overlapping atoms by carrying out the linear interpolation between the energy minima in internal coordinates (Goumans et al., 2009). Another promising novel initial path generation method has been proposed (Zhu et al., 2019) where the initial path is obtained by a geodesic interpolation between the two minima.

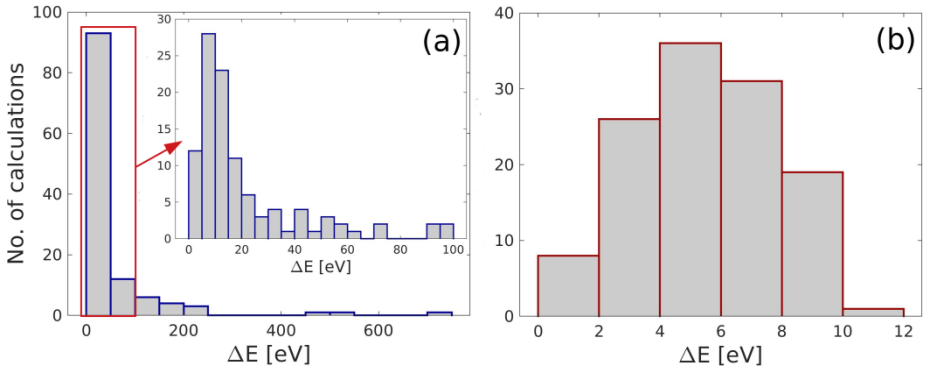


Figure 3.9. Comparison of two different initial path generation methods for NEB, namely linear interpolation in Cartesian coordinates (a) and the image-dependent pair potential (IDPP) method (b). The distribution of the maximum rise in energy (ΔE) along the initial paths generated for a large database of 121 main-group molecular reactions, is shown. Moreover, in (a), for clarity the distribution of $\Delta E \in (0, 100]$ *kcal/mol is shown as an inset. As expected, the IDPP method leads to the generation of more reasonable, lower energy, intermediate configurations compared to linear interpolation in Cartesian coordinates.*

The choice of the initial path will determine the final outcome of a NEB calculation. For instance, it can happen that two or even more MEPs exist between the same reactant and product minima, as is shown in Fig. 3.10. A NEB calculation will most likely converge to the MEP that is closest to the initial path. Then, it is important to sample different MEPs to find the one that corresponds to lowest activation energy. If a preconceived notion of the optimal MEP exists, e.g. an initial guess of the saddle point or an intermediate energy minimum, the information can be used to influence the initial path generation and bring the initial path into the vicinity of the desired MEP. This is illustrated in Fig. 3.10, where the initial paths obtained by linear interpolation converge to the closest MEP. Then, by altering the initial path with an insertion of a single image, the CI-NEB calculation can be made to converge to an alternative MEP. In this scheme, the procedure for the initial path generation is divided into two parts, i.e. interpolation from the reactant energy minimum to the inserted image and then from the inserted image to the product energy minimum. The two partial paths are then concatenated to form the full initial path which can then be used as input for the NEB optimization. When an image is inserted into the initial path generation, the position of the image along the band needs to be selected in such a way as to minimize the difference in maximum deviation of straight-line distances between all pair of images along the band. If the insertion is done properly, a desired MEP will more likely be obtained and the efficiency of the NEB calculation improved.

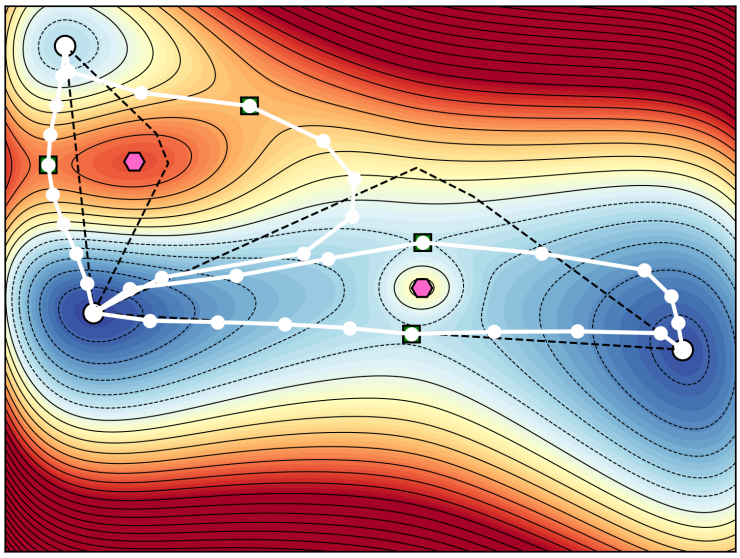


Figure 3.10. Different (discretized) minimum energy paths obtained by CI-NEB calculations on the two-dimensional LEPS potential coupled to an harmonic oscillator (and with added Gaussian functions). (Henkelman and Jónsson, 1999) The three energy minima on the surface are shown by white circles with black borders. The first order saddle points are shown by green boxes and the energy maxima are shown by magenta hexagrams. Four minimum energy paths are connected to the lower left energy minimum. Two from the upper left energy minimum and two from the right energy minimum. The initial paths of the CI-NEB calculations are shown by black dashed lines and illustrate how different initials path can be constructed to locate alternative reaction pathways.

3.4.2 Variants of the CI-NEB method

There have been numerous variants of the NEB method proposed where the objective of the method is to locate the highest energy first order saddle point for a given reaction as efficiently as possible. This is typically achieved by focusing the computational effort on the saddle point region of the path and can be particularly important for NEB calculations of clusters and molecules. For these transitions, the MEP tends to be significantly longer than for e.g. transitions occurring in and on the surfaces of solids. The MEPs are often characterized by long segments of chemically irrelevant configurational changes associated little-to-no change in the energy. The actual energy barrier then constitutes only a small portion of the reaction pathway. Distributing images over the entire reaction coordinate in an equidistant fashion is then substandard, as is highlighted in the example of Fig. 3.11. In this example, two CI-NEB calculations using different number of images are carried out for the formation reaction of 1,2-hexadiene. The reaction coordinate is very long and the actual energy barrier narrow. This results in the use of excessive number of images needed to resolve the MEP in

CI-NEB. Alternatively, the CI-NEB calculations could be confined to the region of the energy barrier and calculations of images in other regions of the path halted. In the following, several CI-NEB variants are briefly reviewed. In these methods, the main theme is to redirect the computational effort to the barrier region.

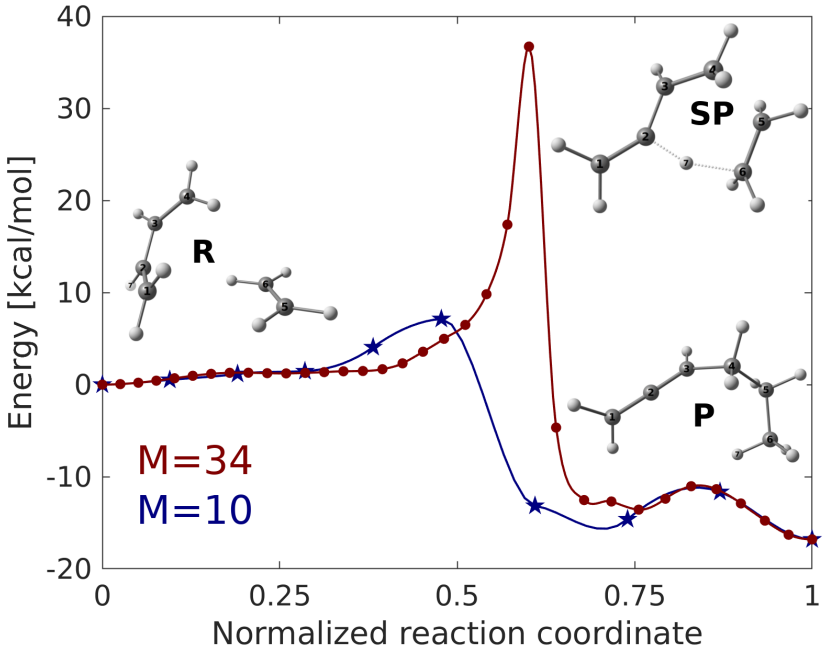


Figure 3.11. Results of CI-NEB calculations of the reaction of ethylene and 1,3-butadiene to form 1,2-hexadiene using $M = 10$ (blue) and $M = 34$ (red) images. The reaction is characterized by a long MEP and a narrow energy barrier. The CI-NEB($M = 10$) calculation is unable to converge in 500 optimization steps and the final path does not even capture the correct characteristics of the energy barrier. The CI-NEB($M = 34$) calculation converges in about 150 optimization steps, i.e. ≈ 4800 energy/gradient evaluations. Despite the large number of images being used in the CI-NEB($M = 34$), the resulting resolution of the energy barrier is, at best, moderate. The reactant (denoted by **R***), saddle point (denoted by* **SP***) and product (denoted by* **P***) configurations are visualized in (c)*

In its simplest form, the CI can be added to the path as an additional image in the gap between the two highest energy images of a regular NEB calculation. (Goumans et al., 2009) Another approach is to use two climbing images located on opposite sides of the HEI. Then, all three images move towards and bracket the saddle point. (Zarkevich and Johnson, 2015). For a more aggressive approach, the adaptive NEB (or ANEB) has been proposed. In the ANEB method, repeated NEB calculations (limited to three movable images) are carried out for a small fixed number of optimization steps. Then,

after each NEB calculation a new initial path is constructed from the position of the two images adjacent to the HEI, i.e. these images define fixed endpoints for the subsequent NEB calculation. (Maragakis et al., 2002) A more elaborate scheme for focusing the CI-NEB calculation on the saddle point region is the autoNEB method. (Kolsbjerg et al., 2016) In this method, additional images are added, on-the-fly, into the largest energy (or geometrical gap) along the path after a certain number of fixed optimization steps. The addition of images is carried out until the calculation reaches a user-defined threshold of allowed number of images. At that point, the CI is activated and the saddle point targeted. To maintain efficient parallelization, only the images surrounding the most recently added image are computed and displaced in each optimization step. All other images remain fixed during the optimization. In the ANEB and autoNEB methods, the partially converged paths are chosen to have fixed endpoints that are not local minima and are not part of the MEP. This can, however, lead to problems as, for example, kinks form on the path as neighboring images to the endpoints get pulled away from the MEP, see Article II. A better approach is to use endpoints that can move either to follow equipotential contours (Zhu et al., 2007) or the component of the atom force perpendicular to the path (Zhang et al., 2016).

The NEB method has also been extended to other application areas, e.g. for calculations of energy ridges and second order saddle points (Maronsson et al., 2012), simultaneous optimization of a periodic system and the simulation cell (Sheppard et al., 2012), temperature-corrected MEPs (Crehuet and Field, 2003), reaction paths on free-energy surfaces (Bohner et al., 2014), tunneling paths (see Article I), point-to-point ionospheric ray tracing (Nosikov et al., 2020) and magnetic transitions (Bessarab et al., 2015).

3.4.3 String method

A path connecting the reactant and product states is in principle a continuous string and such a formulation was given in one of the original NEB publications (Jónsson et al., 1998). The continuous string is represented by a function of a progress parameter, $\phi(s)$, such that $\phi(0)$ is the reactant configuration and $\phi(1)$ is the product configuration. The string is then evolved/displaced in the direction normal to the string until it converges to the MEP (i.e. $\nabla E^{\perp}(\phi) = 0$).

For numerical calculations, the string needs, however, to be discretized by introducing a finite set of images. Instead of using harmonic springs to control the image distribution as in NEB, it is also possible to reparametrize the distribution of the images after each displacement $s_i = [0, \frac{1}{M-1}, \frac{2}{M-1}, ..., 1]$ for $i \in [0, .., M-1]$, where M is again the number of images. This approach is adopted by E and coworkers (E et al., 2002). The position of image i along the string is then given by $\phi(s_i) = s_i S^{\text{arc}}$, where S^{arc} is the arclength of the string. New images can readily be introduced in the parametrization step to e.g. obtain a more accurate representation of the string. However, this will make parallelization of the method significantly more challenging. The discretized string is iteratively displaced downhill in energy along the direction normal to the path, i.e. by computing $\mathbf{F}^{\perp}$. The upwind-tangent scheme is also used in this formulation (Henkelman and Jónsson, 2000; Ren, 2003). Alternatively, the tangent can be computed by the cubic spline representation of the continuous string (Koslover and Wales, 2007).

The climbing image variant of the NEB method can readily be applied to the string method as well. (Sheppard et al., 2008). Moreover, a single-fixed-end climbing image variant of the string method has also been proposed. (Ren and Vanden-Eijnden, 2013) In this method only one of the end-points is fixed at an energy minimum and the other end-point chosen as the climbing image with the effective force transformed in an analogous manner as in the CI-NEB method. The CI is displaced uphill in energy along the tangent direction and downhill in energy along the direction normal to the string. To prevent the CI from escaping the energy basin of the minimum (which is possible for more complicated energy landscapes), a constraint is imposed on the string, where the energy along the string is required to be monotonically increasing.

Quasi-Newton optimization methods can be used in conjunction with the string method to evolve the string to the MEP (E et al., 2002; Burger and Yang, 2006, 2007; Koslover and Wales, 2007; Samanta and E, 2013). There, the energy surface is expanded using a local quadratic approximation for each image individually in the hyperplane normal to the string. However, as the images are relaxed individually, they can move at a different rate causing kinks to develop on the string. This issue has been addressed by introducing a de-kinking spring force acting perpendicular to the path (Burger and Yang, 2007) and by using a strict step-size control (Samanta and E, 2013). As could be expected, the performance and efficiency of the CI-NEB and the string method with a climbing image have been found to be comparable for benchmark calculations, for example the Pt-heptamer island diffusion on a Pt(111) surface and for small-to-medium sized clusters. (Sheppard et al., 2008; Koslover and Wales, 2007)

3.4.4 Growing string method

As discussed in section 3.4.1, the initial path used as input to NEB plays an important role as the optimization likely converges on the closest MEP and the closer the initial path is to an MEP, the fewer iterations are needed for convergence. The typical linear interpolation in Cartesian coordinates usually presents a poor choice for the initial guess that may lead to slow convergence of the calculation or even divergence in the electronic structure calculation of the images. In the growing string method (GSM) (Peters et al., 2004), the problem of strong overlap of atoms is avoided by growing separate string fragments from each of the two energy minima (i.e. the reactant and product) while simultaneously optimizing the position of the string fragments. Once, the two string fragments have merged into a single 'continuous' string, connecting the reactant and product states, the method becomes equivalent to the string method of section 3.4.3, and the string is displaced to the MEP. An illustrative GSM calculation on the two-dimensional Müller-Brown surface is shown in Fig. 3.12.

In the first step of GSM, two images are added, one from each endpoint with a fixed distance and in the direction of the other endpoint. These images are referred to as the frontier images. The 'continuous' string connecting the reactant and product minima is represented by a cubic spline interpolation of the two string fragments, from which the tangent to the path is obtained. This choice of tangent is more appropriate for GSM than the linear tangents used in NEB, because of the large unknown gap present in between the frontier images. While, the tangents are obtained from the 'continuous' string, the two string fragments, originating from each minimum, are optimized individually.

The optimization of the string fragments is carried out in an analogous manner to the string method. The images are displaced according to the component of the atom force acting normal to the string, followed by a re-parametrization step. When a frontier image is partially converged (i.e. the magnitude of the atom force drops below a prescribed tolerance) a new image is introduced to the corresponding string fragment. This image is the new frontier image. Images are added according to a fixed spacing from the frontier image along the cubic representation of the path or alternatively by using the linear synchronious transit method. (Behn et al., 2011b). Furthermore, to reduce the computational effort the innermost images can be frozen when a new image is added (Behn et al., 2011a). However, this may hinder the string from actually converging on the MEP. Similar idea, based on selecting the images with the largest atom forces, has been explored for NEB. (Goumans et al., 2009; Crehuet and Field, 2003; Galván and Field, 2008). The growing of the string fragments (i.e. adding of images) is repeated until the two disjoint strings connect to form a single 'continuous' string. Thereafter, the number of images used for the string to resolve the path remains fixed and the string is evolved to the MEP using the original string method. (E et al., 2002) By growing the path sequentially and optimizing the position of the images, high energy regions on the surface are avoided. This renders the GSM more suitable than both the NEB and string method for calculations of paths where a reasonable initial path is hard to generate. However, effective use of parallel computing is not easily realizable with the GSM and also with the development of new and improved methods beyond the linear interpolation of Cartesian coordinates, such as the IDPP initial path, the high energy regions are less likely to be encountered in NEB.

There have been significant improvements and developments of the GSM devised throughout the years. To name a few, a single ended GSM (Zimmerman, 2015) and a climbing image variant (Zimmerman, 2013b) have been used. An energy-weighted parametrization scheme has been used to distribute images along the string (Goodrow et al., 2009). Quasi-Newton optimization methods and internal coordinates have been implemented and used for GSM calculations (Goodrow et al., 2008; Zimmerman, 2013a,b). Moreover, a GSM incorporating the aforementioned improvements has also been combined in a reliable and efficient manner with a single ended method for converging on a first order saddle point (Zimmerman, 2013a), analogous to the NEB-TS method proposed in Article III.

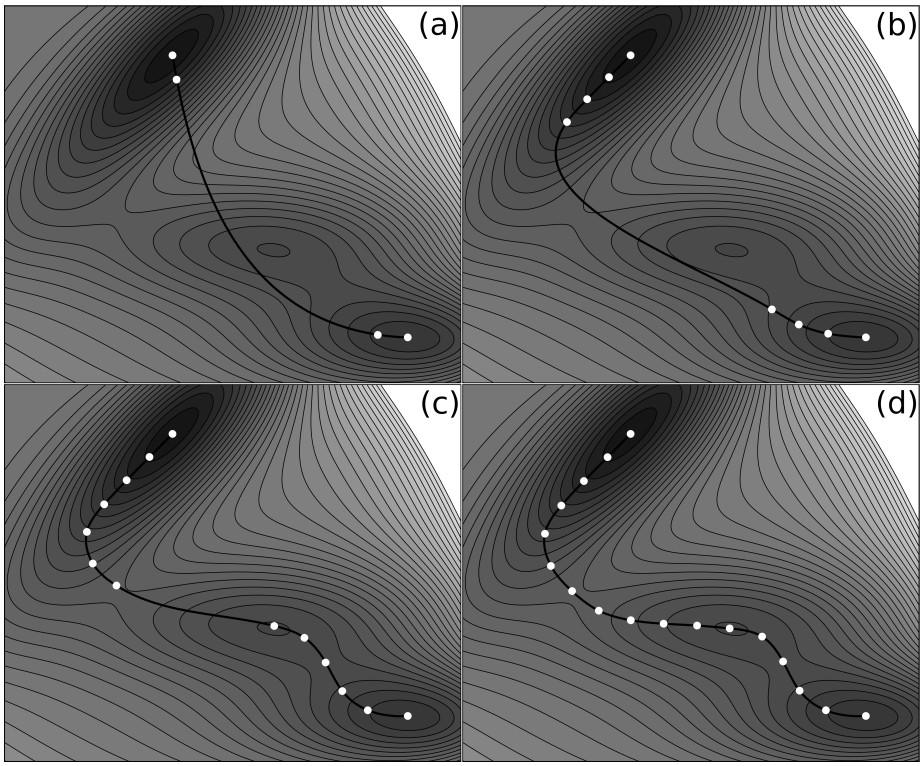


Figure 3.12. Growing string method calculation on the two-dimensional Müller-Brown surface (Müller and Brown, 1979) using a cubic spline representation of the continuous path. In (a), two images are placed adjacent to the reactant and product energy minima and in the direction of the opposite energy minimum. Then the images are optimized according the component of the force acting perpendicular to the continuous path. The final position of these images (white circles) and the continuous representation of the path (black solid line) is shown in (a). After convergence has been reached for the frontier images, other images are added separately in the large unknown gap between the two string fragments. In this manner, the string fragments continue to grow as they are evolved towards the MEP. In (b) and (c), intermediate growth steps of calculations are shown, where $M = 8$ *and* 13*. In (d), the two string fragments have connected to form a single path, composed of* $M = 17$*. This final converged GSM path is accurately aligned with the MEP.*

3.5 Single ended methods

Single ended methods is a class of saddle point search methods where the product of the transition is not specified, i.e. the saddle point search is initiated from a single configuration (e.g. the reactant energy minimum) and moved uphill on the energy surface to a first order saddle point, using both first and (approximate) second derivatives of the energy function. The first order saddle points identified by such methods are highly dependent on the initial guess configuration and the way the second derivative information is obtained and used. In this chapter, a brief description of two single ended methods, i.e. eigenvector-following and minimum-mode following, is given in sections 3.5.1 and 3.5.2, respectively.

3.5.1 Eigenvector-following

The eigenvector-following (EF) partition rational function optimization (P-RFO) method (Banerjee et al., 1985; Baker, 1986) is probably among the most commonly used single ended methods for identifying first order saddle points. In P-RFO, the energy surface is locally expanded by a rational function approximation (RFA), (Banerjee et al., 1985)

$$E(\mathbf{r}+\Delta\mathbf{r}) \approx E(\mathbf{r}) + \frac{\mathbf{g}^T\Delta\mathbf{r} + \Delta\mathbf{r}^T\mathbf{H}\Delta\mathbf{r}}{1+\Delta\mathbf{r}^T\mathbf{S}\Delta\mathbf{r}} \tag{67}$$

instead of LQA. $\mathbf{S}$ is chosen as the identity matrix times a constant γ. The RFA is a more suitable model than LQA for energy surfaces of chemical reactions as it remains bounded for large $\Delta\mathbf{r}$ (Peters, 2017; Banerjee et al., 1985). The displacement towards the minimum of the RFA expansion is then given by

$$r_i = -\frac{g_i}{\lambda_i - \gamma} \tag{68}$$

for $i = [1,..,D]$, where λ_i denotes the i-th eigenvalue of the Hessian matrix. g is the gradient of energy w.r.t atom displacements in the eigenvector basis. γ is a shifting parameter, analogous to the one in the Cerjan-Miller method. The difference between the Cerjan-Miller method and RFO is that the shift parameter naturally arises from the RFO expansion, while in Cerjan-Miller the shift parameter originates from the use of Lagrangian to constrain the size of the displacement. In RFO, the shift parameter can be used to control both the length of the displacement as well as to shift the eigenvalues of the Hessian matrix to make them all positive when an energy minimization is carried out. Or, for saddle point searches, the Hessian is shifted to have usually at least one negative eigenvalue. Banerjee et al. proposed to partition the saddle point search into two individual (but simultaneous) RFO optimizations, one for energy maximization and the other for minimization. (Banerjee et al., 1985) This warrants the use two shift parameters, $\gamma_{\max}$ and $\gamma_{\min}$. A mode, say $\mathbf{q}_k$ (with eigenvalue λ_k) is selected to represent the reaction coordinate and the shift parameter $\gamma_{\max}$ is then used to ensure that each displacement is uphill in energy along $\mathbf{q}_k$. The other shift parameter, $\gamma_{\min}$, is determined to ensure that the energy is minimized along all other modes, i.e. $\mathbf{q}_i$ where $i \neq k$. The

shift parameters are determined from the roots of,

$$\gamma_{\max} = \frac{g_k^2}{\gamma_{\max} - \lambda_k} \tag{69}$$

$$\gamma_{\min} = \sum_{i \neq k}^{3N} \frac{g_i^2}{\gamma_{\min} - \lambda_i} \tag{70}$$

where $\gamma_{\max}$ is the larger root of eq. 69 and $\gamma_{\min}$ is the smallest root of eq. 70. For a more detailed discussion about the two shift parameters, see Refs. Peters (2017) and Schlegel (2011). The position of $\mathbf{q}_k$ can shift in the eigenspectrum during the P-RFO calculation. Therefore, the mode chosen as the reaction coordinate needs to be re-identified after every displacement of the system, hence the name eigenvector-following. This is done computing the maximum overlap of $\mathbf{q}_k$ from the previous step to all modes at the current step. (Baker, 1986)

Because of the use of two shift parameters and the tracking of the reaction coordinate, P-RFO is highly reliable and can, in most cases, be started from anywhere on the energy surface, e.g. within an energy well (convex region) or in regions characterized by multiple negative eigenvalues, and still converge accurately to a first order saddle point. However, it is important to stress that most often there are multiple saddle points connected to the same energy minimum and single ended methods will converge to different first order saddle points for different initial guess configurations and different choices of the reaction coordinate. This is illustrated in the example presented in Fig. 3.13, where nine P-RFO calculations, started both within and outside of the inflation surface, are able to identify three of the four connected saddle points. This also brings us to an important point that some saddle points connected to a given energy minimum can be extremely hard to locate with single ended methods alone. Such saddle points often require the initial guess configuration to be already in close proximity to the saddle point, which can lie far outside of the inflation surface. In P-RFO, it is also possible to identify different saddle points using the same initial configuration by following a different reaction coordinate, e.g., $\mathbf{q}_k = \mathbf{q}_2$. However, if the search is started close enough to the correct saddle point, then the lowest mode of the Hessian will be the correct choice for the reaction coordinate, as it aligns with the unstable mode. It has been reported that P-RFO may sometimes encounter problems when attempting to leave the convex region (Olsen et al., 2004) or when dealing with bimolecular and unimolecular reactions characterized by multiple low frequency eigenmodes of the Hessian matrix. (Peters et al., 2004)

As for Newton-Raphson and quasi-Newton methods, the Hessian matrix (or an approximation thereof) is needed to evaluate the displacement in EF P-RFO. With an accurate Hessian and in close proximity to the saddle point, EF P-RFO may achieve quadratic convergence. However, the Hessian matrix is typically approximated using the Bofill update scheme (Schlegel, 1982; Bofill, 1994). For challenging systems with multiple (e.g., near-degenerate eigenmodes), the approximation introduced to the Hessian matrix by the Bofill method may not be sufficiently accurate. This can distort the eigenspectrum and cause the eigenvector-following to lose track of $\mathbf{q}_k$. (Peters, 2017) In these cases, it may be optimal to construct a more accurate initial guess configuration and/or rely on the computation of the exact Hessian matrix. The Hessian

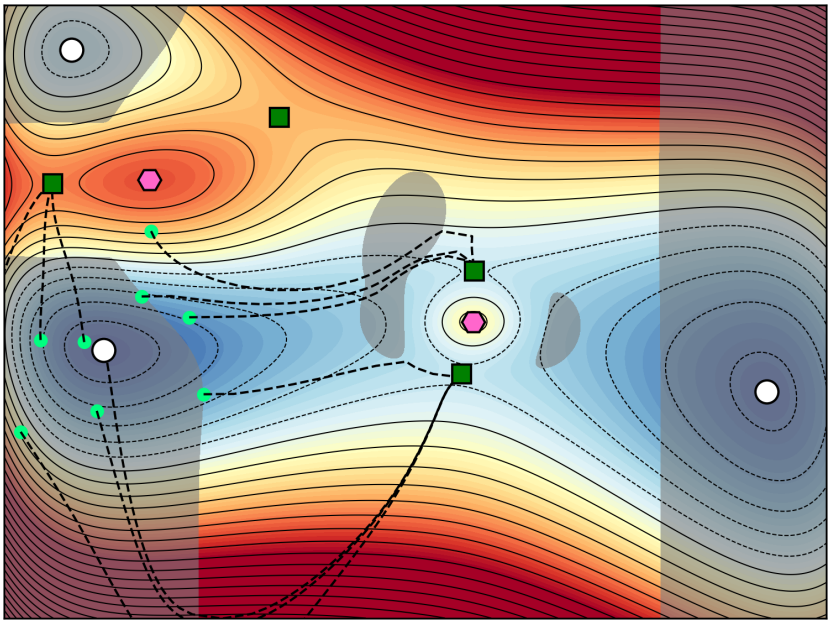


Figure 3.13. Optimization trajectories for several partitioned rational function eigenvector-following calculations on the two-dimensional LEPS potential coupled to an harmonic oscillator (with added Gaussian functions). (Henkelman and Jónsson, 1999) The three energy minima on the surface are shown by white circles with black borders. The first order saddle points are shown by green boxes and the energy maxima are shown by magenta hexagrams. The shaded part of the surface shows where $\lambda_{\min} > 0$*. In all calculations, the eigenmode,* q_{min}*, corresponding to the lowest eigenvalue is followed. The calculations are started around the lower left energy minimum, both inside and outside of the inflation surface. The starting points are given by small green circles. The optimization trajectories are shown by dashed black lines.*

matrix at the starting point of EF P-RFO calculations needs to be of the correct form and with $\mathbf{q}_k$ that resembles the actual reaction coordinate. Therefore, the initial Hessian matrix is often computed analytically or numerically at the beginning of the calculation. But, it may also be possible to use lower-level of theory electronic structure methods (e.g. semi-empirical methods) or as in Article III, construct an empirical Hessian matrix (Fischer and Almlöf, 1992) and modify it accordingly.

3.5.2 Minimum-mode following

In the minimum-mode following (MMF) method, the system is iteratively displaced to a first order saddle point using only the atom force and the eigenvector corresponding to the lowest eigenvalue of the Hessian matrix, $\mathbf{q}_{\min}$, or an approximation thereof. This eigenvector is referred to as the minimum mode.

In a similar fashion to the P-RFO method, the minimum mode is used to partition the displacements of the atoms into two parts. That is, an energy maximization along the minimum mode and simultaneous minimization in all directions orthogonal to the minimum mode, (Henkelman and Jónsson, 1999),

$$\mathbf{F}^{\mathrm{MMF}} = \mathbf{F} - 2\left(\mathbf{F}\cdot\mathbf{q}_{\mathrm{min}}\right)\mathbf{q}_{\mathrm{min}} \tag{71}$$

Note that the projection of the atom force carried out in eq. 71 is the same as for climbing image in CI-NEB, except that $\mathbf{q}_{\mathrm{min}}$ is used here instead of the approximate path tangent at the climbing image.

In MMF, the first step is to bring the system over the inflation surface, $\lambda_{\mathrm{min}} < 0$, e.g. by following gradient-ascent or stepping along the minimum mode $-\left(\mathbf{F}\cdot\mathbf{q}_{\mathrm{min}}\right)\mathbf{q}_{\mathrm{min}}$. (Henkelman and Jónsson, 1999; Olsen et al., 2004) Then, the system is iteratively displaced according to F^{MMF} until a first order saddle point is located. The displacement of the system can be carried out using any number of gradient based optimization methods, as discussed in section secs: 3.1 and 3.4.

There are many different ways to estimate the minimum mode, the most common one being the dimer method. In the dimer method, two system replicas, held at a finite, fixed distance (Δ) apart from a center image are used to compute the lowest mode of the Hessian matrix, i.e. $\mathbf{r_1} = \mathbf{r}_{\mathrm{mid}} + \frac{\Delta\hat{n}}{2}$ and $\mathbf{r_2} = \mathbf{r}_{\mathrm{mid}} - \frac{\Delta\hat{n}}{2}$, where $\hat{n}$ is pointed along the axis of the dimer. The energy surface of the dimer is then approximated by a second order expansion and simplifies to (Peters, 2017)

$$E(\mathbf{r}_1,\mathbf{r}_2) \approx 2E(\mathbf{r}_{\mathrm{mid}}) + \frac{\Delta^2}{4}\hat{n}^T\mathbf{H}\hat{n} \tag{72}$$

the lowest energy orientation of the dimer is obtained when $\hat{n}$ aligns with the direction of the minimum mode, $\mathbf{q}_{\mathrm{min}}$. In this way, the lowest mode can be computed by iterative minimization of the dimer energy in the space spanned by $\hat{n}$ using only the atom forces(Henkelman and Jónsson, 1999; Olsen et al., 2004; Heyden et al., 2005; Kästner and Sherwood, 2008), i.e. no second derivatives are needed. Alternative methods to find the minimum mode are directional minimization of a finite difference approximation of the second derivative (Voter, 1997; Munro and Wales, 1999), the iterative Lanczos method (Malek and Mousseau, 2000; Olsen et al., 2004) or the Davidson method (Plasencia Gutiérrez et al., 2017). Therefore, in the MMF methods several evaluations of the energy/atom force are required in each step to explicitly compute the minimum mode. However, since the minimum mode often change only slightly between steps, computational resources can be saved by evaluating the minimum mode only every few iterations. In comparison to EF P-RFO methods that use approximate Hessian (i.e. only one energy/gradient evaluation per displacement) MMF methods are likely to require more computational effort. However, for challenging systems where the Hessian matrix might need to be computed every few steps, MMF methods will become more efficient than EF P-RFO. Also, the MMF methods are more suitable for large systems as they do not require the storage (or diagonalization) of the $3Nx3N$ dimensional Hessian matrix.

It is important to stress that single ended saddle point searches require an initial configuration (MMF and EF P-RFO) and a qualitatively accurate initial estimate of the Hessian matrix (EF P-RFO)to locate a saddle point. When attempting to locate saddle

points on the edge of an energy basin to compute the escape rate from a given state, both EF and MMF methods suffer from the problem of re-discovering already known saddle points for different initial guesses. This can be partially remedied by maximizing the distance between initial guess configurations on a hypersphere. (Plasencia Gutiérrez et al., 2017)

If, however, the objective of a single ended calculation is to locate a saddle point for a desired transition, the initial configuration of the calculation needs to be close enough to the desired saddle point and/or the correct reaction coordinate needs to be chosen. This is near impossible to do by chemical intuition in high-dimensional systems. Therefore, to generate such an initial guess, considerable human (and computational) effort is often spent. Typical strategies involve starting from the highest energy configuration along a linear interpolation in Cartesian, internal or distance matrix coordinates (Halgren and Lipscomb, 1977). Alternatively, the synchronious transit-guided quasi Newton method (Peng and Bernhard Schlegel, 1993) or the IDPP method (Smidstrup et al., 2014) can be used (see Article III). It is also common to carry out one- or two-dimensional constrained energy minimization where a coordinate is selected and dragged from the reactant to the product state. However, all of these methods are not without faults and are often not be reliable enough for complex reactions. It is therefore advantageous to combine single and double ended methods, (Henkelman and Jónsson, 2000; Peters et al., 2004) to find first order saddle points for a given transition, see Refs. Zimmerman (2013a) and Heyden et al. (2005), as well as, Article III.

3.6 Tunneling and quantum harmonic transition state theory

Transition state theory, as described in sections 3.2 and 3.3, is a very powerful theory used to estimate the classical rate of reactions, i.e. over-the-barrier hopping mechanism. However, the theory is inherently classical. While, quantum effects like the zero point energy can readily be incorporated into the HTST rate expression by using the appropriate partition functions in the rate expression. However, the effect of quantum mechanical tunneling is entirely missing from TST. (Wigner, 1938)

A tunneling process is where the system passes through regions of the energy surface where the potential energy is higher than the total energy of the system (Meisner and Kästner, 2016). In other words, the system passes through a classically insurmountable energy barrier to enter a new state. The tunneling probability is a function of the mass of the tunneling atoms, the shape of the PES in vicinity to the reactant and product states and the total energy of the system. In other words, the lighter the participating atoms and the shorter the tunneling path, the stronger the influence of tunneling. The mass-dependence of tunneling leads to a strong kinetic isotope effect. Therefore, by substituting the expected tunneling atoms by their respective isotopes, the importance of tunneling can be probed.

Tunneling is predominantly observed for reactions involving light species such as hydrogen atoms, e.g., hydrogen atoms hopping between sites on surfaces. (Lauhon and Ho, 2000; Ásgeirsson et al., 2017) However, there are also studies that indicate the importance of heavy-atom tunneling in e.g. organic reactions. (Castro and Karney, 2020; Doubleday et al., 2017). Moreover, if the temperature is low enough, any reaction

will, in principle, be dominated by tunneling (Kästner, 2014). For example, reactions occuring under atmospheric and astrochemical conditions. (Shannon et al., 2013; Fang et al., 2016; Ásgeirsson et al., 2017). There are, also, cases where tunneling is observed to be the preferred mechanism at and above room temperature, e.g. in biochemical system (Sutcliffe and Scrutton, 2002; Masgrau, 2006). In Article I, tunneling is found to be the dominant reaction mechanism for the dissociation of molecular hydrogen from ammoniumborane at room temperature. For an excellent review on the different aspects of atom tunneling in chemistry, the reader is referred to Ref. Meisner and Kästner (2016)

The temperature for the crossover from the over-the-barrier mechanism to tunneling can be roughly estimated from the first order saddle point that characterizes the classical transition mechanism, (Gillan, 1987)

$$T_{\mathrm{c}} = \frac{\hbar|\omega^{\ddagger}|}{2\pi k_{\mathrm{B}}} \tag{73}$$

where $\omega^{\ddagger}$ is the vibrational frequency of the unstable mode at the saddle point. The temperature, T_{c}, is referred to as the crossover temperature. The more narrow the energy barrier, the larger $\omega^{\ddagger}$ is, the higher the T_{c}. At and below the cross-over temperature, tunneling will become the dominant transition mechanism. This results in a break-down of the Arrhenius behavior (i.e. the linear relationship between $ln(k)$ and $1/T$) and manifests itself as an effective reduction and eventually vanishing of the activation energy, i.e. the overall rate becomes temperature independent.

3.6.1 Computation of a tunneling correction factor

In its most simple form, tunneling can be accounted for by computing a correction factor, κ^{tun}, and applying it to the classical rate, i.e. $k^{\mathrm{corr.}} = \kappa^{\mathrm{tun}} k^{\mathrm{classical}}$. The assumption here is that the reaction coordinate is chosen accurately and is separable from all the other modes of the system. Hence, the correction factor can be computed as the ratio of the quantum mechanical and classical transmission coefficient, respectively, through the reaction coordinate,

$$\kappa^{\mathrm{tun}} = \frac{\int_0^{\infty} P^{\mathrm{QM}}(E)\exp(-\beta E)\mathrm{d}E}{\int_0^{\infty} \mathscr{H}(E-E^{\ddagger})\exp(-\beta E)\mathrm{d}E} \tag{74}$$

where $\mathscr{H}$ is the Heaviside function. The quantum mechanical transmission probability, P^{QM}, is given by the Jeffreys-Wentzel-Kramer-Brillouin (JWKB) approximation (Dunham, 1932)

$$P^{\mathrm{QM}}(E^{\mathrm{sys}}) = \frac{1}{1+\exp(\frac{2}{\hbar}\theta(E^{\mathrm{sys}}))} \tag{75}$$

where $\theta(E^{\mathrm{sys}})$ is the barrier-penetration-integral (also known as the imaginary action integral) for system of fixed energy, E^{sys},

$$\theta(E^{\mathrm{sys}}) = \int_{\mathbf{r}(s_1)}^{\mathbf{r}(s_2)} \sqrt{2\mu(s)\left[E(\mathbf{r}(s)) - E^{\mathrm{sys}}\right]}\mathrm{d}s \tag{76}$$

where $\mathbf{r}(s_1)$ and $\mathbf{r}(s_2)$ are the classical turning points along the reaction coordinate, s that fulfill $E(\mathbf{r}(s_1)) = E(\mathbf{r}(s_2)) = E^{\mathrm{sys}}$. The integration is carried out over the, classically

forbidden segment of the chosen reaction path. The dimension of the integral is action and the smaller the value θ, the larger the tunneling probability, P^{QM}. Therefore, a path on the energy surface, between $\mathbf{r}(s_1)$ and $\mathbf{r}(s_2)$, that minimizes the action integral of eq. 76, defines a path of maximum tunneling probability. This path can differ significantly from the MEP depending on temperature. The optimal tunneling path will climb up along the walls on the concave side of the reaction channel (i.e. the path will 'cut corners') to higher energy regions of the energy surface, in order to effectively shorten the tunneling path, which in turn enhances the tunneling rate. (Marcus and Coltrin, 1977) This effect of corner-cutting is more pronounced as the temperature is lowered. (Richardson, 2018)

The most important parameters of the selected reaction coordinate, e.g. $E^{\ddagger}$ and $\omega^{\ddagger}$, are often used to fit prototypical problems to energy barriers for which analytical (or approximate) JWKB solutions to the quantum mechanical transmission probability are readily available. One such example is the infinite downward parabola. (Wigner, 1932) The analytical solution of the infinite parabola gives the famous Bell correction, (Bell, 1959)

$$\kappa^{\text{tun}} = \frac{\frac{1}{2}\beta\hbar\omega^{\ddagger}}{\sin(\frac{1}{2}\beta\hbar\omega^{\ddagger})} \tag{77}$$

It was derived for and is only valid for temperature, $T > T_c$. Another correction factor (obtained from a truncated series expansion) approximates the solution to the infinite downward parabola and is referred to as the Wigner tunneling correction,

$$\kappa^{\text{tun}} \approx 1 + [\beta\hbar\omega^{\ddagger}]^2/24 \tag{78}$$

Even though, Wigner's correction factor is applicable for lower temperature than T_c, it is only accurate at relatively high temperature, where tunneling is seemingly unimportant. Usually, it is more accurate to fit a one-dimensional (asymmetric) Eckart barrier to the reaction coordinate, since the parabolic fit is too simple and narrow at the base. For the Eckart barrier, analytical solutions of κ^{tun} are also available. (Johnston and Heicklen, 1962) An Eckart barrier can be fitted to a reaction coordinate using the imaginary vibrational frequency at the classical transition state, $\omega^{\ddagger}$, as well as the, zero-point energy corrected potential energy barrier height for the forward and backward reactions. The Eckart model can yield reasonably accurate correction factors for temperature around T_c. (Kästner, 2014) However, the tunneling probability is highly dependent on the shape of the energy barrier and it becomes more important as the temperature is lowered.

A more elaborate and accurate set of methods are the curvature-tunneling methods. In these approaches, the JWKB approximation is used to compute the quantum mechanical transmission coefficient through a reaction coordinate, defined *a priori*. In the zero-curvature tunneling (ZCT) method, the reaction path is chosen the MEP. (Truhlar and Kuppermann, 1971) In the widely-used small-curvature tunneling (SCT) method, (Skodje et al., 1981) the effect of corner-cutting on the tunneling rate is realized. (Marcus and Coltrin, 1977) Therefore, by carrying out an harmonic expansion orthogonal to the MEP and introducing an effective mass along the coordinate (dependent on the reaction curvature), the tunneling probability through a path of least action (within the harmonic expansion) can be computed. Therefore, in SCT, the calculation

and diagonalization of the Hessian matrix along the MEP is required. The SCT method provides accurate results for tunneling transitions that are characterized by small corner-cutting effects, or at a temperature around and somewhat below T_c. This is realized in the large-curvature tunneling (LCT) method, where the reaction path is selected as a straight-line between the reactant and product energy minima. (Garrett et al., 1983) The small- and large curvature tunneling methods are two extremes and are accurate for limiting cases. (Kästner, 2014) Therefore, intermediate methods that combine the two have been devised and are applicable over a larger range of temperature, e.g., the optimized multidimensional tunneling method, OMT, which selects the larger tunneling correction of SCT and LCT. (Garrett et al., 1985)

3.6.2 Direct computation of the canonical tunneling rate

Multidimensional tunneling (and other quantum nuclear effects) are better described using a statistical description of the quantum dynamical behavior of the atom nuclei. The canonical quantum mechanical rate constant is then expressed in terms of the quantum dynamical partition functions, analogous to TST for the computation of the canonical rate constant. The following section is largely based on Refs. Arnaldsson (2007); Kästner (2014); Richardson (2018). Within, the Feynman path integral formalism, (Feynman et al., 2010) the canonical partition function Q of a quantum mechanical ensemble can be expressed as

$$Q = \int \exp\left(-\frac{1}{\hbar} S_{\mathrm{E}}(\tilde{\mathbf{r}}(\tau))\right) \mathscr{D}_{\tilde{\mathbf{r}}(0)=\tilde{\mathbf{r}}(\hbar\beta)} \tag{79}$$

$$S_{\mathrm{E}} = \int_0^{\hbar\beta} \left[\frac{\mu}{2}\left(\frac{\mathrm{d}\tilde{\mathbf{r}}(\tau))}{\mathrm{d}\tau}\right)^2 + E(\tilde{\mathbf{r}}(\tau))\right] \mathrm{d}\tau \tag{80}$$

where $\tilde{\mathbf{r}}(\tau) = \mathbf{r}(s(\tau))$ is a path in configuration space dependent on the time-parameter τ. The operator $\mathscr{D}_{\tilde{\mathbf{r}}(0)=\tilde{\mathbf{r}}(\hbar\beta)}$ signifies that the integration in eq. 79 is carried out over all closed Feynman paths (CFP). A closed Feynman path has a period of $\hbar\beta$, i.e. a CFP is a path that fulfills the boundary condition $\tilde{\mathbf{r}}(0) = \tilde{\mathbf{r}}(\beta\hbar)$. Each CFP is weighted by the exponent of the Euclidean action, given in eq. 80.

In analogy to the energy surface and the classical MEP, a path of largest statistical weight on the Euclidean action surface is a minimum action path (MAPs) and the highest action first order saddle point along the MAP represents a dynamical bottlenecks for the transition from the reactant state CFP to the product state CFP. (Mills et al., 1997, 1998) Or more precisely, a MAP is a path on the Euclidean action surface where each point along the path is a CFP. The derivative of the Euclidean action for each CFP is tangential to the MAP (i.e. the derivative is zero in all modes normal to the path). The CFP of highest action, along the MAP, is a first order saddle point on the Euclidean action surface and is referred to as an 'instanton'. (Arnaldsson, 2007) An alternative interpretation of instantons (referred to as the semiclassical approximation) is that instantons are a periodic classical orbit on the inverted PES that goes through the classically forbidden region of the (inverted) energy barrier, with a period of $\beta\hbar$. An instanton is a multidimensional path on the PES that minimizes θ of eq. 76, i.e. the optimal tunneling path for a given temperature. (Richardson, 2018)

In practice, for the computation of the Euclidean action, the CFP needs to be discretized by a set of system configurations (or images) and the canonical, quantum mechanical, partition function of eq. 79 needs to be approximated. The discrete representation of the Euclidean action in the space of CFPs is given by,

$$S_{\mathrm{E}}(\mathbf{y},T) = \sum_{i=1}^{P} \left(\frac{1}{2}\frac{P}{\hbar\beta}|\mathbf{y}_{k+1} - \mathbf{y}_k|^2 + \frac{\hbar\beta}{P}E(\mathbf{y}_k) \right) \tag{81}$$

where the path is discretized into P images of the system distributed in equidistant time steps along the path, i.e., $\tilde{\mathbf{r}}(\tau_k)$ for $\tau_k = k\frac{\beta\hbar}{P-1}$ with $k \in [1,2,...,P]$. This distribution of the images in configuration space represent the quantum delocalization of the atom nuclei Moreover, to conveniently account for the effective mass, μ, in eq. 80, mass-weighted coordinates, $\mathbf{y}_k = \sqrt{m_i}\tilde{\mathbf{r}}(\tau_k)_i$ (and m_i is the mass of the i-th coordinate component), are used. Therefore, a discretized CFP is a $3N\ x\ P$ dimensional point in path space and is referred to as a ring-polymer. A ring-polymer configuration of P images is denoted as $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_P]$.

For clarity, following the discussion of Ref. Arnaldsson (2007), the discretized Euclidean action can be written in terms of an effective energy surface (V^{eff}) where each image is connected by temperature-dependent springs,

$$S_{\mathrm{E}}(\mathbf{y},T) = \hbar\beta V^{\mathrm{eff}}(\mathbf{y},T) \tag{82}$$

$$V^{\mathrm{eff}}(\mathbf{y},T) = \sum_{k=1}^{P} \frac{1}{2}k^{\mathrm{sp}}(T)|\mathbf{y}_{k+1} - \mathbf{y}_k|^2 + \frac{E(\mathbf{y}_k)}{P} \tag{83}$$

where $k^{\mathrm{sp}}(T)$ is given by,

$$k^{\mathrm{sp}}(T) = P\left(\frac{k_{\mathrm{B}}T}{\hbar}\right)^2 \tag{84}$$

hence each image along the ring-polymer feels only an average potential, $\frac{1}{P}E(\mathbf{y})$, and is connected to its two neighboring images by a temperature-dependent spring interaction. The shape of the instanton is temperature dependent. For temperatures above T_{c}, the spring interaction overcomes the curvature of the underlying PES and the ring-polymer is collapsed, i.e. $\mathbf{y}_1 = \mathbf{y}_2 = ... = \mathbf{y}_P$. As the temperature is lowered below T_{c}, the spring interaction becomes weaker, the curvature of the energy surface allows the CFP to open up and begin to spread over the energy barrier toward the reactant and product states. The amount of spreading increases with decreasing temperature. At very low temperatures, where only the lowest vibrational ground state of the reactant state is occupied, the instanton can simplistically be thought of as stretching over the base of the classical energy barrier and essentially describes a 'deep-tunneling' mechanism with a temperature independent overall reaction rate. In reality, however, the instanton is a multidimensional curved, corner-cutting, path on the energy surface that can significantly differ from the classical reaction coordinate (i.e. the optimal tunneling path for a given temperature).

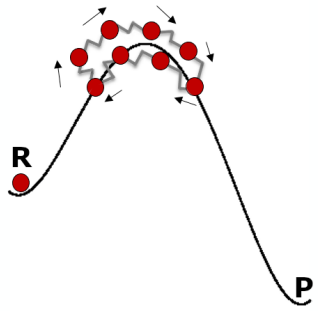


Figure 3.14. Illustration of two ring-polymers: an instanton spreading over the classical energy barrier at a temperature slightly below the cross-over temperature and the collapsed ring-polymer in the reactant energy minimum.

Armed with the definition of a ring-polymer, the quantum dynamical partition function of eq. 79 becomes

$$Q = \int \dots \int \exp\left(-\frac{1}{\hbar} S_{\mathrm{E}}(\mathbf{y})\right) \mathrm{d}^{3NP}\mathbf{y} \tag{85}$$

Hence, the quantum statistical mechanics of the system become mathematically equivalent to the classical statistical mechanics of a ring polymer of P images of the system connected by temperature dependent springs. (Barker, 1979; Chandler and Wolynes, 1981) Rigorous thermal averages are obtained as P$\rightarrow \infty$. In analogy to the Vineyard's formulation of HTST (Vineyard, 1957) (discussed in section 3.3.2). The partition function of eq. 85 for a given discretized CFP can be approximated by expanding the Euclidean action up to second order around a stationary CFP. Or, for computation of the canonical tunneling rate, the expansion is carried out for the reactant energy minimum CFP and the instanton CFP. (Langer, 1969; Miller, 1975; Coleman, 1977; Callan and Coleman, 1977; Benderskii et al., 2009; Messina et al., 1995; Richardson, 2016) This method is referred to as the harmonic quantum transition state theory (HQTST). The resulting partition function for the reactant energy minimum is,

$$Q^{\mathrm{R}} = \sqrt{\frac{(\pi\hbar)^{3NP}}{\prod_i^D \omega_i^{\mathrm{R}}}} \exp\left(-\frac{1}{\hbar} S_{\mathrm{E}}(\mathbf{y}^R)\right) = \sqrt{\frac{(\pi\hbar)^{3NP}}{\prod_{i=1}^D \omega_i^{\mathrm{R}}}} \exp\left(-\beta E(\mathbf{r}^R)\right) \tag{86}$$

where the only possible stationary (discretized) CFP for the reactant state is collapsed in the energy minimum ($\mathbf{r}^R$). Therefore, as is evident from eqs. 82 and 83, $\frac{1}{\hbar} S_{\mathrm{E}} = \beta E(\mathbf{r}^R)$ where $\mathbf{r}^R$ is the reactant energy minimum configuration on the PES. The ω_i are the angular vibrational frequencies of the vibrational modes obtained by the diagonalization of the Hessian matrix of the ring-polymer i.e. $\mathbf{H}^{\mathrm{CFP}} = [\mathbf{H}_1, \mathbf{H}_2, ..., \mathbf{H}_P]$ and D is the number of vibrational modes. For the ring-polymer of the reactant state CFP, all the images are collapsed to the same point on the N-dimensional PES and hence only a single Hessian calculation is required to evaluate the partition function of eq. 86. For molecular systems, there are six degrees of freedom that correspond to rotational and

translational motion of the ring-polymer. These degrees of freedom need to be treated separately, as discussed in section 3.3.1, and introduced into the partition functions, Q^R and $Q^{\ddagger}$ (see below). For the rotational partition function, the ring-polymer can be taken to be a 'super-molecule' and the rotation treated classically, where each image only accounts for $1/P$-th of the total mass. However, at a low temperature, this approximation needs to be validated. Also, as usual, the translational contributions can be skipped, as they cancel out for the reactant and instanton ring polymers in the final expression for the rate constant. For a more in-depth discussion for the treatment of molecular systems, the reader is referred to Ref. (Beyer et al., 2016)

The partition function for the instanton CFP (i.e. at the first order saddle point on the Euclidean action surface) is given as,

$$Q^{\ddagger} = \sqrt{\frac{\beta\hbar P S_0}{2}} \sqrt{\frac{(\pi\hbar)^{3NP-1}}{\prod_{i=2}^{D} |\omega_i^{\ddagger}|}} \exp\left(-\frac{1}{\hbar} S_{\mathrm{E}}(\mathbf{y}^{\ddagger})\right) \tag{87}$$

$$S_0 = \frac{P}{\hbar\beta} \sum_{k=1}^{P} |\mathbf{y}_{k+1} - \mathbf{y}_k|^2 \tag{88}$$

where the S_0 term has been factored out, as this degree of freedom corresponds to the permutational translation of images along the ring polymer. The motion of images along the CFP involves no change in the Euclidean action. Instead, it contributes only an entropic term which is proportional to the length of the ring polymer. In contrast to Vineyard's HTST, the frequency corresponding to the unstable mode at the instanton is not factored out. Instead, only the magnitude of the frequency is relevant. (Kästner, 2014)

The HQTST, canonical, rate constant for a tunneling transition from the reactant energy minimum to a product energy minimum, along the instanton, can be expressed in terms of the quantum dynamical partition functions,

$$k^{\text{inst.}} = \frac{2k_{\mathrm{B}}T}{\hbar} \frac{Q^{\ddagger}}{Q^R} \tag{89}$$

where $Q^{\ddagger}$ and Q^R are given by eqs. 87 and 86, respectively. (Kästner, 2014) In classical HTST, the rate constant is depends exponentially on the activation energy (i.e. difference between the energy of the saddle point and the reactant energy minumum), the canonical rate constant is dependent on the rise in Euclidean action along the MAP, i.e. $\frac{Q^{\ddagger}}{Q^R} \propto \exp\left(\frac{1}{\hbar}[S_{\mathrm{E}}(\mathbf{y}^{\ddagger}) - S_{\mathrm{E}}(\mathbf{y}^R])\right)$.

To evaluate the HQTST tunneling rate from any given reactant energy minimum, the connected instanton needs to be located. As discussed above, instantons are first order saddle points on the Euclidean action surface. Therefore, the eigenvector-following (Baker, 1986) and minimum-mode following (Henkelman and Jónsson, 1999) saddle point methods, as discussed in section 3.5, can be applied to shift the high-dimensional ring-polymer to the instanton. (Arnaldsson, 2007; Rommel et al., 2011) This is achieved by using only the negative gradient of the Euclidean action with respect to the atom coordinates, i.e. $\frac{\partial S_{\mathrm{E}}}{\partial x_i}$. In practice, only half of the P images are needed in the optimization as the images tend to pair up, i.e. $\mathbf{y}_k = \mathbf{y}_{P-k+1}$ for $k \in [1, 2, ..., P/2]$.

Still, in comparison to the classical analog, the identification of an instanton and the computation of the HQTST rate calculations is considerably more demanding. Each optimization step requires a factor of roughly $P/2$ more energy/gradient evaluations than is required in the classical case. The number of energy/force evaluations required per image can be on the order of hundreds to thousands of evaluations. Also for the evaluation of the rate, $P/2$ Hessian matrices along the ring-polymer at the instanton need to be computed. Depending on temperature, the value of P can range from 20 to 100 images. Also, to obtain converged eigenvalues of the instanton Hessian matrix, very tight convergence thresholds are usually needed in. Therefore, locating an instanton for a given temperature and evaluating the tunneling rate can be a computationally challenging task if the energy and atom forces are obtained from electronic structure calculations.

To evaluate the tunneling rate for various values of the temperature, as is customary, the instanton needs to be identified for various decreasing values of the temperature. In other words, for each temperature, the Euclidean action surface changes and new instantons (first order saddle points) emerge. Typically, for HQTST calculations over a given temperature interval, a sequential cooling procedure is used. (Rommel et al., 2011) From any given initial ring-polymer, an instanton search is carried out. Then, the temperature is lowered and the converged instanton from the previous calculation is used as the initial ring polymer for the a new instanton search. This procedure is then repeated. As the temperature is lowered, more and more images are needed to resolve the path, since the path becomes longer and the images tend to aggregate near the reactant and product energy minima. The reason for this can easily be seen by using the periodic orbit interpretation of instanton theory, i.e. the semi-classical approximation. The images along the instanton ring-polymer are distributed in equidistant time-steps along the inverted energy barrier. The system will spend most of its time near the end-points of the orbit (which corresponds to regions in vicinity to the reactant and product energy minima) with approximately zero velocity, as the system moves down-hill along the inverted energy-barrier, the potential energy is converted into kinetic energy and the velocity increases. With increased velocity in and around the center region of the inverted energy barrier, the distance traveled per time-step becomes greater.

The HQTST approximation to the tunneling rate is found to provide accurate rate constant estimate for a variety of chemical reactions. (Rommel and Kästner, 2011; Rommel et al., 2012; D.M. Einarsdóttir and Jónsson, 2012; Jonsson, 2011; Ásgeirsson et al., 2017; Meisner and Kästner, 2016) A noteworthy success of HQTST is the mechanistic study of hydrogen atom reaction with methane to form molecular hydrogen and a methyl radical. (Andersson et al., 2009) In this study, the rate estimated by HQTST was found to be in excellent agreement to multiconfigurational time-dependent Hartree calculations, (Meyer et al., 1990; Manthe et al., 1992) where the quantum dynamics of the system are propagated directly in imaginary time, from the reactant energy minimum ensemble, on a fitted PES.

4 Summary of Articles I-IV

The focus of this dissertation is the development and evaluation of tools and methods that can be used to characterize chemical reactions, i.e. the identification of the mechanism of atomic rearrangements and the estimation of the rate of reaction, where both classical and quantum mechanical description of the atom nuclei is used.

Within the commonly used harmonic approximation to transition state theory (HTST), the calculation of the classical rate of escape from a given reactant state is transformed into the problem of locating low-lying saddle points on the energy ridge surrounding the state. Or, for a given reaction (i.e. both reactant and product states are known *a priori*) by the highest energy maximum along the minimum energy path (MEP). The energy maximum along the MEP characterizes the transition of the system as it moves from the reactant to the product state. Furthermore, within the framework of harmonic quantum transition state theory (HQTST), the calculation of the tunneling rate is cast into the problem of locating a saddle point on an extended quantum mechanical energy surface (referred to as an 'instanton') or equivalently by locating the optimal tunneling path (OTP) by minimizing the action of a trial path. Therefore, the methods and tools used by computational chemist to navigate and identify classical and quantum mechanical saddle points, as well as, stationary paths on energy surfaces play a pivotal role in studies of chemical reactivity. Typically, the energy surfaces are computed using electronic structure methods and the computational cost of the calculations can readily become large and for some systems even computationally prohibitive. It is, therefore, very important to make saddle point search and path optimization methods as reliable and robust as possible for calculations on a large variety of systems, while also reducing the number of electronic structure calculations needed to obtain the desired, converged, solution.

4.1 Applications of NEB to molecular reactions

Molecular reactions are often characterized by a large degree of flexibility, i.e. degrees of freedom where there is little change in energy as the system traverses along these degrees of freedom. This results in long and flat reaction paths where the reactive event, that gives rise to the actual energy barrier, contributes only a small segment of the path. Hence, in standard CI-NEB calculations of molecular reactions, where the discretization points used to represent the path are evenly distributed along the path, significant computational effort is wasted on resolving irrelevant parts of the MEP. More importantly, the resolution of the path near the climbing image (and hence saddle point) may be too small for the tangent estimate to be accurate enough to converge to the energy maximum along the path. By simply increasing the number of discretization points, the resolution of the full path can be improved. However, this may not necessarily

improve the resolution of the energy barrier itself to a sufficient extent. It is neither practical nor computationally feasible to continually increase the number of images until a sufficiently accurate resolution of the energy barrier is achieved, especially when computationally intensive methods are used to obtain the energy and forces. The use of an excessive number of images can simply lead to other problems. Instead, more elaborate solutions are needed.

In Article II, the Z-NEB method is presented to resolve the resolution issue that often occurs in CI-NEB calculations of molecular reactions, without introducing an excessive number of images. In Z-NEB, the computational effort of a CI-NEB calculation is simply focused into the region of the highest energy maximum along the MEP using a two step procedure.

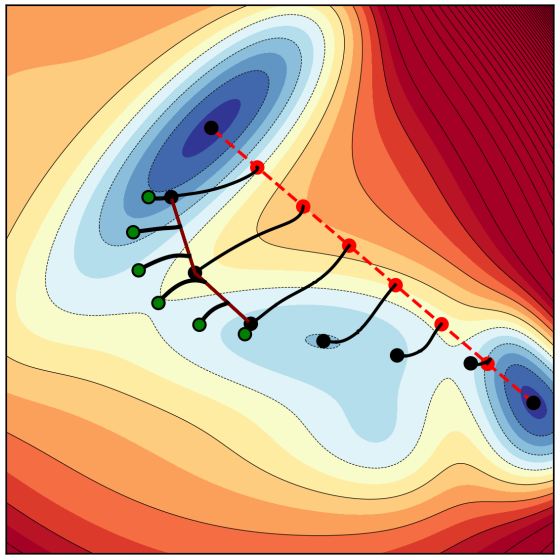


Figure 4.15. Illustration of the Z-NEB method on the two-dimensional Müller-Brown surface. The initial, linear interpolation path is shown by a red dashed line and the individual images by red circles. The displacement of the images during the iterative optimization is shown by black solid lines. The final position of the images converged to a segment of the minimum energy path is shown by green circles. In the first part of Z-NEB, a loosely converged CI-NEB calculation is carried out and the partially converged position of the images is then shown by the black circles. The solid dark-red line indicates the region selected for introduction of additional images in the second part of Z-NEB. It lies between the climbing image and its two adjacent images. New set of images are introduced in this region by linear interpolation and images outside this region are not treated further. In the second part of Z-NEB, the images are further shifted to the minimum energy path, including the endpoint images which move in the direction of the atom force acting normal to the current path.

In the first part of Z-NEB, a regular CI-NEB calculation is carried out until a rough estimate of the MEP emerges. A piecewise-cubic polynomial using both the energy and

tangential atom force is used to interpolate the energy along the approximate MEP. In the second part of Z-NEB, the interpolated path is used to identify the region surrounding the highest energy peak along the approximate MEP. A new set of images is distributed in this region of the path and optimized using CI-NEB to converge on the segment of the MEP that contains the highest energy, first order, saddle point. In the second CI-NEB calculation, the images that lie closest to the boundaries of the high-energy region are selected to be the new end-points. These points are not not necessarily located on the MEP and need to be optimized simultaneously to the MEP, along with the other images. Otherwise, the end-points may serve to pull the neighboring images of the MEP causing kinks to form along the path. The end-points are iteratively shifted to the MEP according to the component of the atom force acting normal to the path, i.e. they are not subject to the spring forces. Because of this and to maintain the near perfect parallelization property of the NEB method, Z-NEB uses two fewer images in the second CI-NEB calculation. The Z-NEB method is illustrated in Fig. 4.15 on a two-dimensional Müller-Brown surface (Müller and Brown, 1979).

In Article II, the Z-NEB method is successfully applied to two molecular reactions, hydrolysis of ethyl acetate and rearrangement of 1,5-hexadiene. Both reactions are characterized by relatively long pathways and a complicated pericyclic, 6-membered, saddle point geometry. For the rearrangement reaction of 1,5-hexadiene, a CI-NEB calculation using 8 images (6 movable) is unable to converge to the saddle point. The non-convergence is attributed to large fluctuations of the CI around the energy maximum. By increasing the number of images to 14, the oscillations of the CI are partially quenched and the CI-NEB calculation converges in 300 optimization steps or 3986 energy/force evaluations. A Z-NEB calculation, using 8 images, for the same reaction is found to converge in 140 step using only 902 energy/force evaluations, i.e. only 23% of the computational effort needed by CI-NEB($M = 14$). This shows that by focusing the computational power to the most important region of the reaction path significant computational savings can be achieved.

An alternative method of directing the computational resources of CI-NEB calculations to the relevant, high-energy, regions of the path is used in Article III. This method is referred to as energy-weighted (EW) CI-NEB. (Henkelman et al., 2000) The stiffness of the spring forces in EW-CI-NEB is determined by the relative energy of the connected images, the higher the energy, the stiffer the springs. This forces the images to climb up-hill (according to the tangential spring force) and accumulate in the chemically relevant regions of the MEP, i.e. around the energy maximum. A comparison of the image distribution offered by CI-NEB and EW-CI-NEB for an example molecular reaction is presented in Fig. 4.16. The EW-CI-NEB method, using $M = 10$, offers similar resolution of the energy barrier as the CI-NEB method using $M = 34$. The price to pay is that other less important parts of the path become more poorly resolved. CI-NEB calculation using the same number of images as the EW-CI-NEB calculation (i.e. $M = 10$) is found to be non-convergent. The final, not fully converged, path of that calculation was shown in Fig. 3.11.

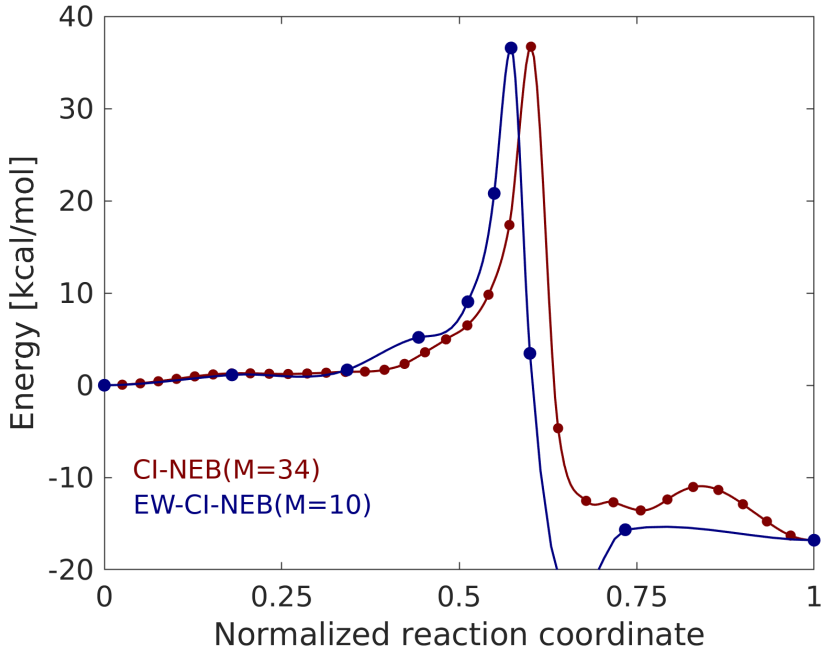


Figure 4.16. Results of CI-NEB and EW-CI-NEB calculations on the reaction of ethylene and 1,3-butadiene to form 1,2-hexadiene using $M = 34$ (red) and $M = 10$ (blue) images, respectively. This is the same reaction included in Fig. 3.11.

In Article III, the performance of EW-CI-NEB method is assessed and compared to CI-NEB for a large benchmark set of 121 main-group molecular reactions. The benchmark set combines two previously published benchmark sets (Zimmerman, 2013a; Birkholz and Schlegel, 2015) and offers a wide variety of reaction types characterized by a plethora of different features of the energy surfaces. The objective of the CI-NEB calculation is to approximately locate the highest energy first order saddle point that characterizes a given reaction, as efficiently as possible. The results are summarized in Fig. 4.17.

As shown in Fig. 4.17, the convergence ratio of CI-NEB using $M = 7$ images is found to be 74%. By increasing the number of images, the convergence ratio increases. More precisely, CI-NEB calculations using $M = 10$ and 14 images are found to be convergent for 82% and 93% of the reactions, respectively. The average energy/force evaluations needed by the CI-NEB calculations increases from 874 to 1824 as the number of images is increased from 7 to 14. In contrast, EW-CI-NEB calculations using 7 images, exhibit an impressive convergence ratio of 98% with an average of 589 energy/force evaluations. This amounts to a 33% reduction in computational effort compared to CI-NEB using same number of images. Moreover, by increasing the number of images to $M = 10$, the convergence ratio of EW-CI-NEB calculation is found to be 100%. It is, therefore, clear that the image distribution offered by EW-CI-NEB is more advantageous for calculations of molecular reactions than the even distribution of

images used in standard CI-NEB calculations. This is because, as the images accumulate in vicinity of the energy maximum along the MEP, the tangent estimate at the climbing image becomes more accurate, which in turn leads to a more stable and efficient CI-NEB calculation. It remains to be seen which of the two approaches, EW-CI-NEB or Z-NEB, or even some combination of the two, will end up being the optimal NEB method for calculations of approximate first order saddle points.

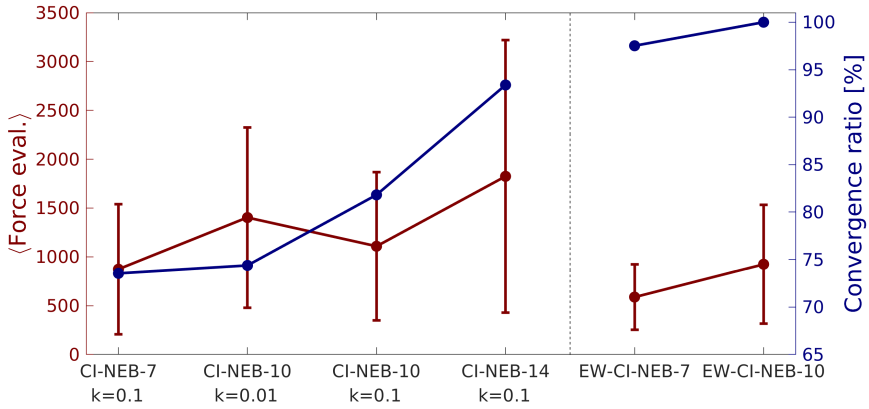


Figure 4.17. Performance of CI-NEB and EW-CI-NEB calculations on the benchmark set of 121 main-group molecular reactions. The notation (EW-)CI-NEB-X is used to denote the method used and X specifies the number of images used in the calculations, where $X = 7,\ 10$ *and* 14. *For the CI-NEB calculations, the spring constant* k^{sp} *used in the CI-NEB calculations is specified on the x-axis and is in units of Ha/Bohr*^{2}. *For the EW-CI-NEB calculation, the spring constant is scaled from* $k^{sp} = 0.01$ *to* 0.1 *Ha/Bohr*^{2}. *The average number of energy/force evaluations in convergent CI-NEB calculations is given by the left y-axis. The standard deviation of energy/force evaluations is shown as an error bar. The convergence ratio for a set of calculations is given by the right y-axis. Note that non-convergent calculations are omitted from the calculation of averages and standard deviation.*

The combination of double-ended and single-ended methods is likely to be the most efficient way of obtaining first order saddle point for a given reaction. (Heyden et al., 2005) In this scheme, the double-ended method is used to obtain an approximate saddle point configuration which is then used as the starting point for a subsequent single-ended saddle point search. The single ended method is then used to obtain a rigorous convergence to the first order saddle point. The computational gain of switching from a double-ended method to a single-ended method arises from the disparity in computational cost per optimization step, i.e. double-ended calculations are carried out in discrete path space while single-ended calculations are carried out in configuration space. Also, single ended methods are much more reliable to obtain rigorous convergence to a saddle point than double ended methods. While, single ended methods are highly dependent on the initial guess configuration (and the selected reaction coordinate) and in order to identify the correct saddle point, double ended

methods are needed to bring the system in vicinity of the desired saddle point.

In Article III, the IDPP-TS and NEB-TS methods are presented and evaluated on the same benchmark set of 121 main-group molecular reactions. In NEB-TS, the EW-CI-NEB and EF P-RFO methods (referred to as a TS search) are combined to locate, first order, saddle points for a given reaction. The EW-CI-NEB method is used in the first part of NEB-TS to obtain an approximate reaction path that connects the given reactant and product state. The climbing image, from the partially converged MEP, is used as the initial guess configuration for the TS search. The initial Hessian matrix used at the beginning of the TS search can either be computed analytically (if available by the given electronic structure method) or by an empirical model Hessian using the Almlöf scheme (Fischer and Almlöf, 1992). The path tangent at the CI is then used to select the correct eigenvector to be followed. If the EW-CI-NEB is converged to a high enough degree, the path tangent should already provide an excellent approximation to the unstable mode at the saddle point, i.e. the reaction coordinate. Also, if the Almlöf model Hessian is used, then the eigenspectrum needs to be modified accordingly, the eigenvalue corresponding to the selected eigenvector is shifted according to a finite difference estimate of the curvature of the path. The TS search method then uses the Bofill update method to approximate the Hessian in subsequent steps. Therefore, if the empirical estimate of the Hessian matrix is used to begin with, the exact computation of the Hessian matrix can be entirely avoided in NEB-TS. The IDPP-TS method is similar to NEB-TS, where the EW-CI-NEB part is skipped. Instead the TS search is instead initiated from the highest energy image along an IDPP initial path. This method is similar to the typical combination of single-ended methods and interpolation methods, e.g. the linear synchronous transit method of Halgren and Lipscomb. (Halgren and Lipscomb, 1977)

The performance of NEB-TS and IDPP-TS on the benchmark set of main-group molecular reactions is summarized in Figs.4.18 and 4.19. The NEB-TS method, using $\max(|\mathbf{F}_{\mathrm{CI}}|) < 0.51$ eV/Å, is found to be remarkably efficient and reliable, with an average of 305 energy/force evaluations per reaction and a 100% convergence ratio. Majority of the computational effort of NEB-TS calculations is used in the initial EW-CI-NEB calculation to obtain an accurate initial guess for the TS search, or an average of 260 energy/force evaluations. By converging the climbing image to $\max(|\mathbf{F}_{\mathrm{CI}}|) < 0.1$ eV/Å, the computational effort the EW-CI-NEB is roughly doubled (or 501 energy/force evaluations), while the computational effort of the TS search is decreases from 45 to 33 cycles. Surprisingly, the inclusion of the analytical Hessian matrix, only marginally improves the efficiency of NEB-TS with an average reduction of about 5-8 energy/force evaluations. As expected, IDPP-TS is found to be even more efficient than NEB-TS with an average of 120 energy/force evaluations per reaction and also a surprisingly high convergence ratio of 97%. The high convergence ratio of IDPP-TS is attributed to an interplay of both the quality of the initial path generated by IDPP and also to the robustness of the TS search method, even when using an approximate second derivative. For comparison, TS searches started from the highest energy image along an initial path constructed by linear interpolation in Cartesian coordinates yields only 69% convergence ratio. If the exact Hessian is used at the beginning of the TS search in IDPP-TS, the efficiency is reduced by about 11 energy/force evaluations.

The quality of the initial guess configuration of NEB-TS and IDPP-TS determines

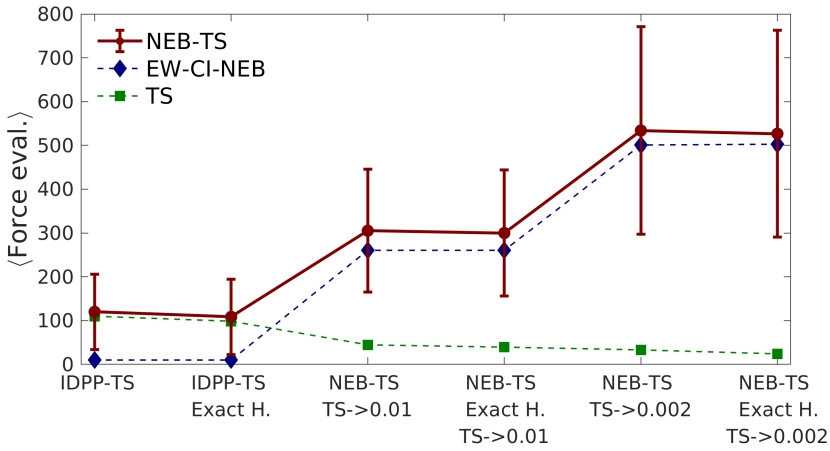


Figure 4.18. The computational efficiency of NEB-TS and IDPP-TS calculations (using $M = 10$*) on the benchmark set of main-group molecular reactions. The TS search is started when* $\max(|\mathbf{F}_{CI}|) < 0.51$ *or 0.1 eV/Å. The TS search is started either using an analytical Hessian matrix or the empirical Almlöf scheme. The average number of energy/force evaluations required by convergent NEB-TS (and IDPP-TS) calculations to locate saddle points is shown with red solid line along with the standard deviation shown as an error bar. The dashed blue and green dashed lines show how the NEB-TS calculations breaks down into the initial EW-CI-NEB phase and the subsequent TS search.*

whether the calculations are able to identify the correct, connected, saddle point. Therefore, to assess the quality of the saddle points obtained by NEB-TS, the energy deviation from a reference set of saddle points is briefly examined and shown in Fig. 4.19. Large energy differences are caused by the TS search converging to different saddle points than the ones included in the reference set. These different saddle points can for example belong to a different MEP that connects the same reactant and product state (i.e. an alternative, but valid, reaction pathway). The methods can also converge to a different, lower-energy, saddle point along a multistep MEP. The third and most severe possibility is that a saddle point that does not connect to the given product state is obtained, i.e. corresponds to a wrong reaction pathway.

The SPs obtained by IDPP-TS can exhibit large energy deviations from the reference set by up to 3.5 eV. In IDPP-TS, 30% of the SPs show a larger energy deviation than 0.02 eV (0.5 kcal/mol) and 14% of the SPs a larger deviation than 0.17 eV (4.5 kcal/mol). If NEB-TS is used, with $\max(\mathbf{F}_{\mathrm{CI}} < 0.51$ eV/Å, the fraction of SPs that deviate is found to decrease to 17% and 6%, respectively. If the exact Hessian matrix is used initialize the TS calculations (of IDPP-TS or NEB-TS), the ratios are further decreased by 1.5–2.5%. The best agreement to the reference set of SPs is obtained for NEB-TS using $\max(\mathbf{F}_{\mathrm{CI}} < 0.1$ eV/Åand an exact initial Hessian matrix, then about 15% and 2% of the SPs show a larger energy deviation than 0.02 and 0.17 eV, compared to the reference set.

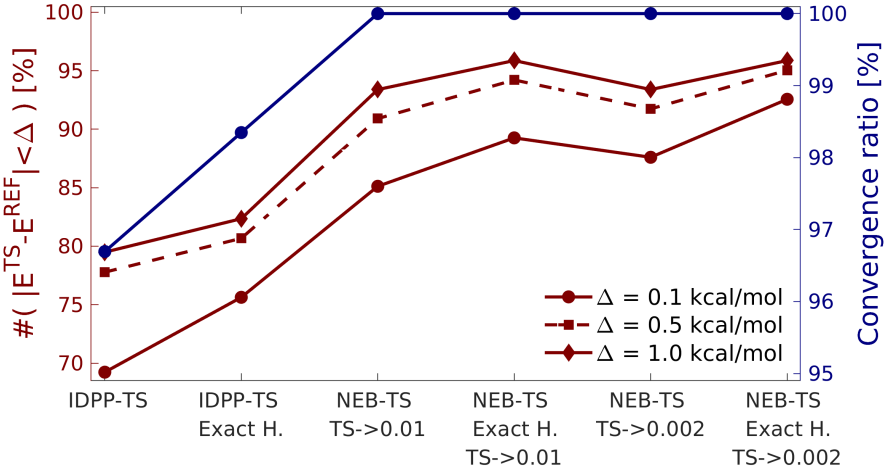


Figure 4.19. Convergence ratio and accuracy of saddle points obtained by NEB-TS and IDPP-TS calculations (using $M = 10$) on the benchmark set of main-group molecular reactions. The TS search is started when $\max(|\mathbf{F}_{CI}|) < 0.51$ *or 0.1 eV/Å. The TS search is started either using an analytical Hessian matrix or the empirical Almlöf scheme. The ratio of convergence is given by the blue line (right vertical axis). The ratio of NEB-TS and IDPP-TS calculations that yield a saddle point with an energy difference from the reference set of saddle points, that is smaller than a given* Δ *is shown in red (left vertical axis). The ratio is computed for three different values of* Δ*, namely* $0.1, 0.5$ *and* 1.0 *kcal/mol.*

The SPs obtained by IDPP-TS can deviate in energy from the reference SPs by up to roughly 3.5 eV. Around 31% of the SPs obtained by IDPP-TS show a larger energy deviation than 0.004 eV and 22% of the SPs a larger deviation than 0.02 eV. In NEB-TS, where the CI of EW-CI-NEB is converged to $\varepsilon_{\max}^{\mathrm{TS}} = 0.5$ eV/Åbefore starting the TS search, the distribution of energy differences is found to span a range of 0 to 1.3 eV. The fraction of SPs that deviate by more than 0.004 and 0.02 eV is found to decrease to 15% and 9%, respectively. If the exact Hessian matrix is used to initialize the TS calculations (of IDPP-TS or NEB-TS), instead of the model Almlöf Hessian, the ratios can be further decreased by roughly 4–7%. The best agreement to the reference set of SPs is obtained for NEB-TS using $\varepsilon_{\max}^{\mathrm{TS}} = 0.002$ eV/Åand an exact initial Hessian matrix. For this set of calculations, about 7% and 6% of the SPs show a larger energy deviation than 0.004 and 0.02 eV compared to the reference set, respectively.

These results show the IDPP-TS and NEB-TS are both efficient and reliable choices for identifications of saddle points. However, there is a possibility that the TS search may converge to an incorrect SP. This effect can be reduced by improving the initial guess configuration, i.e. using EW-CI-NEB and converging it to a tighter threshold and/or using the exact Hessian at the beginning of the TS search for the selection of the reaction coordinate.

In comparison to NEB-TS, a state of the art, integrated, combination of the climb-

ing growing string (GS) method and TS search has recently been developed and published. (Zimmerman, 2013a) The GS-TS method was found to have a 100% convergence ratio with an average of 500 energy/force evaluations using $M = 11$ images. The calculations were carried out for a slightly smaller benchmark set than used in this dissertation, or 105 main-group molecular reactions. It is, therefore, clear that the performance offered by NEB-TS is comparable to that of the GS-TS method. However, an additional advantage of the EW-CI-NEB method is the electronic structure computations carried out in NEB calculations are readily parallelizable. Interestingly, on a subset of 72 selected reactions, the GS-TS method with $M = 7$ images is found to have a convergence ratio of 81%. (Zimmerman, 2013a) While, the NEB-TS method, using 7 images on the larger set of 121 main-group molecular reactions is found to have a 98% convergence ratio, signifying that the NEB-TS method is likely to be more robust than the GS-TS method and less sensitive to the choice of the number of images.

4.2 Optimal tunneling paths and HQTST rate computation

The identification of first order saddle points on the energy surface allows for the computation of the thermally averaged classical rate constant within the framework of HTST. At low temperature, the classical description of the atom nuclei may not be sufficiently accurate. Instead, both zero-point energy and quantum tunneling need to be taken into account. In fact, at temperature below the reaction-dependent crossover temperature, quantum mechanical tunneling is found to be the dominant reaction mechanism. In this case, the computation of the thermally averaged tunneling rate constant with approximate tunneling contributions can be carried out using the HQTST method.

In HQTST, the thermally assisted tunneling rate is estimated from a second order expansion around the point of highest Euclidean action along the minimum action path (MAP) that connects the reactant and product states. Each point on the MAP is a closed Feynman path (CFP) of $Px3N$ dimensions, where P is the number of images used in the discrete representation of the CFP and N is the number of atoms. The point of highest action along the MAP corresponds to an 'instanton', or a first order saddle point on an extended quantum mechanical energy surface. Hence, both double-ended and single-ended methods may be used to locate the full MAP or the instanton, respectively. Typically, the minimum-mode following method has been used to locate instantons. These type of saddle point calculations, however, turn out to require enormous computational effort and are therefore not easily combined with electronic structure methods. First, a large number of images may be needed to resolve the instanton (e.g., in comparison to NEB calculations), since the images tend to accumulate near the reactant and product states. The number of images needed for converged results depends on the curvature of the energy surface and also on the temperature. Furthermore, due to the coupling of images in CFPs, very tight convergence thresholds are needed to determine the saddle point along the MAP to a first order, i.e., the $(Px3N)\,x\,(Px3N)$ dimensional Hessian matrix needs to have one and only one negative eigenvalue.

In Article I, the line-integral NEB (LI-NEB) method is presented to significantly reduce the computational effort needed for computation of the thermally assisted tunnel-

ing rate constant using HQTST. In this approach, instead of searching for the instanton, the optimal tunneling path is located using the LI-NEB method. The optimal tunneling path is the path of maximum tunneling probability for a fixed energy of the system. It turns out that the OTP traces out the same path on the energy surface as an instanton for a fixed temperature. Therefore, once an OTP for a given energy has been obtained, the corresponding temperature can be computed and the corresponding instanton constructed, allowing for the computation of the HQTST rate. A preliminary report of the development of this method was previously presented in a conference proceeding. (D.M. Einarsdóttir and Jónsson, 2012)

In LI-NEB, the OTP for a system with energy E^{sys} is obtained by minimizing the barrier penetration integral (θ) for a path $\mathbf{r} = \mathbf{r}(s)$ where $s \in [0,1]$,

$$\theta = \frac{1}{\hbar} \int_{\mathbf{r}(0)}^{\mathbf{r}(1)} \sqrt{2\mu(E(\mathbf{r}) - E^{\text{sys}})} \mathrm{d}r \tag{90}$$

the points $\mathbf{r}(0)$ and $\mathbf{r}(1)$ are the classical turning points with fixed energy $E(\mathbf{r}(0)) = E(\mathbf{r}(1)) = E_{\text{system}}$. In order to minimize θ, the integral is approximated as

$$\tilde{\theta}(\mathbf{r}_1, ..., \mathbf{r}_{M-1}) = \frac{1}{2\hbar} \sum_{i=1}^{M} \left[(2\mu(E(\mathbf{r}_i) - E^{\text{sys}}))^{1/2} + (2\mu(E(\mathbf{r}_{i-1}) - E^{\text{sys}}))^{1/2} \right] |r_i - r_{i-1}| \tag{91}$$

using a discrete representation of the path, i.e. $[\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_{M-1}]$. The images are then iteratively displaced towards the OTP using the regular NEB method, where the atom force ($\mathbf{F}$) is replaced by the negative gradient of $\tilde{\theta}$ with respect to the atom coordinates, i.e. $-\nabla\tilde{\theta}$. In other words, $\tilde{\theta}$ is minimized in all direction acting normal to the path and fictitious spring forces are introduced to control the distribution of images along the path.

In LI-NEB, however, the end-point images correspond to the classical turning points, $\mathbf{r}(0) = \mathbf{r}_1$ and $\mathbf{r}(1) = \mathbf{r}_{M-1}$. These images need to have a fixed energy, E^{sys}, and finding these points *a priori* is not practical or even feasible. Therefore, in LI-NEB, the end-points are optimized under the constraint to follow the iso-contour corresponding to E_{system}. Because of the simultaneous optimization of the end-point images, a LI-NEB calculation can be initiated from (almost) an arbitrary initial path, where the iterative procedure brings the two images to the correct iso-contour and then displaces them along this energy contour so as to minimize $\tilde{\theta}$. The main condition for the initial placement of the endpoint images is that they are placed on opposite sides of the energy barrier. In practice, the initial path of an OTP calculation should be taken along the MEP or by using previously converged OTP for higher E^{sys}.

Illustrative LI-NEB calculations of OTPs are shown in Fig. 4.20 for the two-dimensional Müller-Brown surface. In this case, there are two different types of tunneling pathways depending on the energy of the system, where lower energy of the system corresponds to lower onset temperature of tunneling. If the energy of the system is larger than the energy of the intermediate energy minimum, the system will tunnel from the reactant energy minimum to the intermediate minimum and then from the intermediate minimum to the product energy minimum, showing only slight corner cutting effect in the latter transit, i.e. deviation from the classical MEP. If the energy of the system, however, is smaller than the energy of the intermediate energy minimum,

the system will not go through the intermediate energy minimum, rather the tunneling occurs directly from the reactant to the product energy minimum, i.e. '*deep-tunneling*'.

The path traced out by an OTP for a fixed energy is equivalent to an instanton for a given temperature. This path corresponds to a classical trajectory on the inverted potential with the period of oscillation given by $\tau = \hbar/k_B T$. Therefore, in order to construct an instanton from an OTP, the corresponding temperature is calculated using the period of the oscillation, τ, which is obtained by performing a dynamical simulation along the one-dimensional path. Once the temperature is known, the instanton corresponding to the OTP is constructed by distributing a new set images along the path in equidistant time steps. In this manner, the number of images used in the calculation of the OTP can be significantly smaller than the number of images used for the instanton, i.e. $P >> M$. Now, armed with the instanton ring-polymer, the thermally assisted tunneling rate can readily be computed using HQTST.

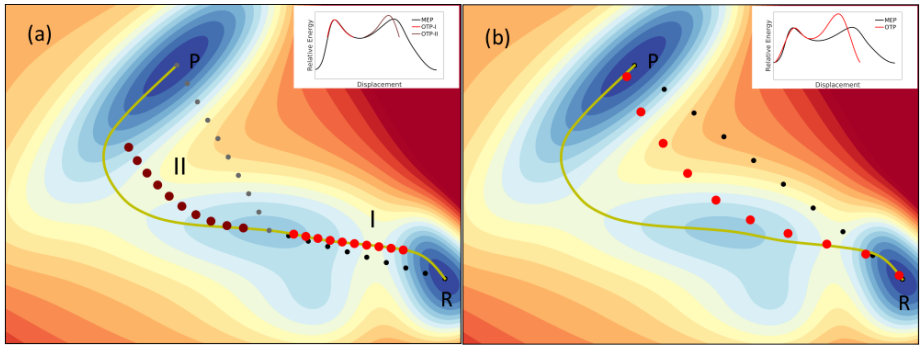


Figure 4.20. Illustration of the LI-NEB method on the two-dimensional Müller-Brown energy surface, using different values of E^{sys}. The minimum energy path connecting the reactant (denoted by **R***) and product (denoted by* **P***) is shown by a yellow curve. The MEP goes through an intermediate energy minimum (with energy E_m) and two first order saddle points (with energy E_I and E_{II}). Linear interpolation between the minima are used as initial paths for the LI-NEB calculations and shown by small filled circles. In (a), optimal tunneling paths for energy $E^{sys} > E_m$, the images along the two paths are shown by brown and red circles. In (b), the optimal tunneling path for energy $E^{sys} < E_m$, the images along the path is shown by red circles. The energy profiles for the minimum energy path and the optimal tunnelings paths are shown as insets.*

In Article I, the LI-NEB method and HQTST are used to compute the thermally assisted tunneling rate of H_2 dissociation from ammoniumborane as a function of temperature, using the B3LYP density functional. The results are also compared to Vineyard's HTST (including rotational partition functions) and quasi-quantum HTST rates, see Fig. 4.21. This reaction presents an interesting example with a unusually high cross-over temperature, estimated to be 333 K. The set of LI-NEB calculations carried out for the dissociation of ammoniumborane were started from E_{system} of 0.05 eV below that of the saddle point energy and an initial path composed of $M = 10$ images interpolated along the relevant segment of the MEP. The energy of the system was

then reduced in a stepwise fashion covering an interval of 1 eV with each LI-NEB calculation started from the previous OTP calculation of higher energy. The instanton was constructed using $P = 240$ images. For the HQTST rate calculation, the Hessian matrix was computed only for the 10 OTP images and a cubic spline interpolation used to form the Hessian matrix of the instanton. As is evident from Fig. 4.21, the HQTST rate is about 10 orders of magnitude larger than the HTST rate. This example shows just how important it is to incorporate nuclear quantum effects when computing the rate at a low temperature. Also, Li et al. (Li et al., 2005) performed SCT calculations, which are based on a harmonic expansion around the classical reaction coordinate, on the same system and found it to be 6 orders of magnitude larger than the classical HTST rate at 200 K, i.e. 4 orders of magnitude smaller than the HQTST rate. This shows just how important it is to accurately account for corner-cutting effects in calculations of tunneling.

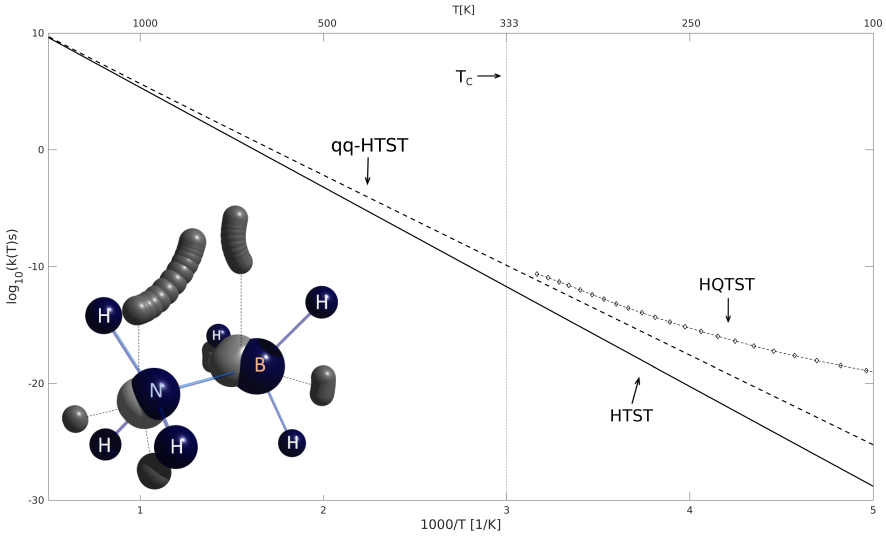


Figure 4.21. The HTST, qq-HTST and HQTST rates as a function of temperature for H_2 dissociation from ammoniumborane. The cross-over temperature, $T_c = 333$ K, is shown as a vertical solid line. The reactant energy minimum configuration (dark blue spheres and light blue bonds) and the optimal tunneling path at 200 K (grey) are super-imposed and shown as an inset.

One of the fundamental difference between LI-NEB calculations of OTPs and the direct search of instantons using MMF is that the distribution of images is controllable in LI-NEB calculations, while in MMF/instanton calculations the images tend to aggregate in the low energy regions. This allows significantly fewer images to be used to resolve the OTP, compared to the instanton. Moreover, because the images are distributed evenly along the OTP, the Hessian matrix of the instanton required to compute the HQTST rate can be formed by interpolation of the Hessian matrices computed along the OTP. Also, LI-NEB calculations appear to be able to use less strict convergence

thresholds compared to MMF/instanton calculations and still attain the same accuracy in the computation of the rate constant. All of these aspects contribute to a significant reduction in the number of energy/force evaluations needed for HQTST computations.

4.3 Solution to a challenging electronic structure problem

The methods and applications presented in the previous sections are focused on the motion of the atom nuclei on an energy surface, i.e. the second part of the Born-Oppenheimer approximation. However, the selection of the appropriate theory to accurately describe the energy surface of a given system is often very challenging.

One such example, presented in this dissertation, is the radical cation of N,N′-dimethylpiperazine (DMP^+), where the positive charge can either be localized on one of the nitrogen atoms or delocalized over the two nitrogen atoms, see Fig. 4.22 (b) & (c). The existence of the two electronic states has been observed experimentally with the energy difference between the two states found being 0.33 eV, where the delocalized state is more stable. (Deb et al., 2013; Cheng et al., 2016) In contrast, commonly used density functionals tend to favor charge delocalization and fail to predict the existence of the localized state of DMP^+. This finding is supported by the 'gold-standard' of computational chemistry, CCSD(T). This controversy has sparked a debate in the literature on whether the localized state, in fact, exists or if DFT and CCSD(T) are simply not accurate enough to properly describe the energy surface of DMP^+. (Ali et al., 2018; Cheng et al., 2018)

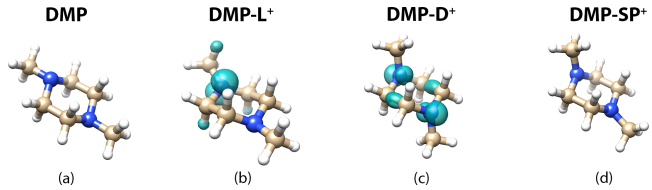


Figure 4.22. Configurations of neutral N,N′-Dimethylpiperazine (a), the localized cation (b), delocalized cation (c), and the saddle point configuration for the transition between localized and delocalized electronic state (d). The spin densities included in (b) and (c) correspond to isosurface level of 0.01 electrons/Å^3.

To establish whether the localized electronic state of DMP^+ exist, the energy surface of DMP^+ is constructed and visualized using various, both single and multi-reference, electronic structure methods. The surface is constructed, using 78 point grid where the dihedral angles are varied from 70 to 175°. Each point on the grid represents an optimized configuration with a fixed dihedral angle. The constrained optimization is carried out using the BHLYP level of theory. Then, a single point energy evaluation is carried out on the grid using the highly accurate multireference wavefunction theory FIC-MRCI, and different density functionals, BLYP, B3LYP and BHLYP, as well as, coupled cluster theory with singles and double excitations and perturbative triples, i.e. CCSD and CCSD(T). The different energy surfaces are visualized in Fig. 4.23.

The highest level of theory method used in this work is the multi-reference method

FIC-MRCI+Q(11,12), where MRCI is carried out on an active space of 11 electrons and 12 orbitals. These high-level calculations are in agreement with the experiments and confirm the existence of both the localized and delocalized states with an energy difference of 0.336 eV, in excellent agreement to the experimental value. The potential energy barrier height, separating the two energy minima is found to be 0.050 eV. The commonly used density functionals, BLYP and B3LYP, fail to find the localized electronic state, i.e. only the delocalized energy minimum is found on the surface. Interestingly, two energy minima are present on the BHLYP surface corresponding to the localized and delocalized states with an energy difference of 0.178 eV. The energy barrier for the transition from the localized to the delocalized state is found to be 0.033 eV. Moreover, the resemblance of the BHLYP energy surface to that of FIC-MRCI+Q(11,12) is remarkable. BHLYP is a hybrid density functional that uses 50% exact exchange from Hartree-Fock, effectively balancing the tendency of DFT for charge delocalization with the tendency of Hartree-Fock for localization. The CCSD energy surface, also, shows the existence of both the localized and delocalized state with an energy difference of 0.221 eV and a barrier height of 0.012 eV. Interestingly, when the perturbative triples are added to CCSD, i.e. the CCSD(T) method, the localized energy minimum disappears from the energy surface. This is in agreement to previous findings. (Cheng et al., 2016; Ali et al., 2018)

In conclusion, the DMP^+ molecular cation represents a challenging example where commonly used electronic structure methods, such as BLYP, B3LYP and CCSD(T), fail to properly describe the energy surface of the system. This has now been confirmed by high-level, arguably state-of-the-art, multireference wavefunction method and experimental measurements. The work presented in Article IV, shows just how challenging it can be to select the correct level of theory to accurately describe the energy surface of a given system. Often the standard, commonly used, tools may not present a suitable choice. For this particular system, the BHLYP could be used to obtain a qualitatively correct description of the DMP^+ cation and be used to explore the potential energy surface (e.g., locate energy minima, saddle points and MEPs), while energy surface exploration using higher-level of theory methods like FIC-MRCI+Q(11,12) (or even CCSD) are still computationally infeasible.

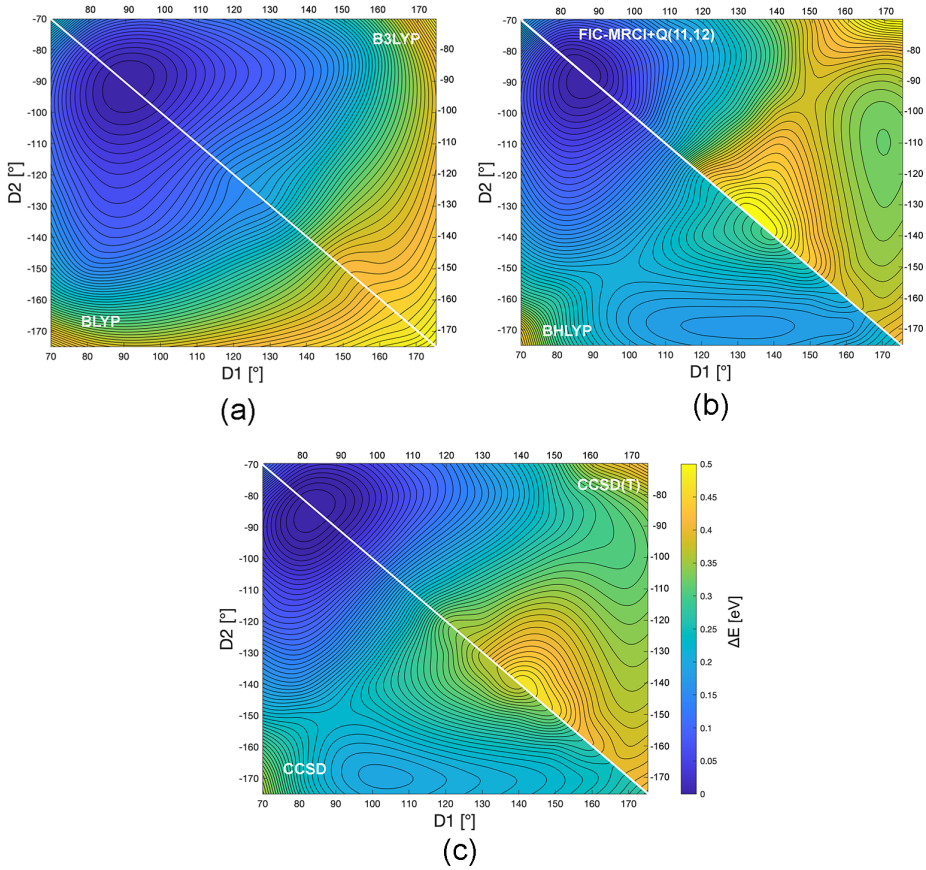


Figure 4.23. Potential energy surfaces of DMP^+ as a function of the two dihedral angles. The calculations are carried out using (a) B3LYP (upper panel) and BLYP (lower panel), (b) FIC-MRCI+Q(11,12) (upper panel) and BHLYP (lower panel) and (c) CCSD(T) (upper panel) and CCSD (lower panel) theory.

References

Ali, Z. A., Aquino, F. W., and Wong, B. M. (2018). The diamine cation is not a chemical example where density functional theory fails. *Nature Communications*, 9(1):4733.

Andersen, H. C. (1980). Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of chemical physics*, 72(4):2384–2393.

Andersson, S., Nyman, G., Arnaldsson, A., Manthe, U., and Jónsson, H. (2009). Comparison of Quantum Dynamics and Quantum Transition State Theory Estimates of the H + CH 4 Reaction Rate. *The Journal of Physical Chemistry A*, 113(16):4468–4478.

Arnaldsson, A. (2007). Calculation of quantum mechanical rate constants directly from ab initio atomic forces.

Ásgeirsson, V. and Jónsson, H. (2020). Exploring potential energy surfaces with saddle point searches. *Handbook of Materials Modeling: Methods: Theory and Modeling*, pages 689–714.

Ásgeirsson, V., Jónsson, H., and Wikfeldt, K. T. (2017). Long-Time Scale Simulations of Tunneling-Assisted Diffusion of Hydrogen on Ice Surfaces at Low Temperature. *The Journal of Physical Chemistry C*, 121(3):1648–1657.

Baker, J. (1986). An algorithm for the location of transition states. *Journal of Computational Chemistry*, 7(4):385–395.

Baker, J., Kessi, A., and Delley, B. (1996). The generation and use of delocalized internal coordinates in geometry optimization. *The Journal of Chemical Physics*, 105(1):192–212.

Banerjee, A., Adams, N., Simons, J., and Shepard, R. (1985). Search for stationary points on surfaces. *The Journal of Physical Chemistry*, 89(1):52–57.

Barker, J. A. (1979). A quantum-statistical Monte Carlo method; path integrals with boundary conditions. *The Journal of Chemical Physics*, 70(6):2914–2918.

Bartlett, R. J. (1989). Coupled-cluster approach to molecular structure and spectra: a step toward predictive quantum chemistry. *The Journal of Physical Chemistry*, 93(5):1697–1708.

Bartlett, R. J. and Musiał, M. (2007). Coupled-cluster theory in quantum chemistry. *Reviews of Modern Physics*, 79(1):291–352.

Becke, A. D. (1988). Correlation energy of an inhomogeneous electron gas: A coordinate-space model. *The Journal of Chemical Physics*, 88(2):1053–1062.

Becke, A. D. (2014). Perspective: Fifty years of density-functional theory in chemical physics. *The Journal of chemical physics*, 140(18):18A301.

Behn, A., Zimmerman, P. M., Bell, A. T., and Head-Gordon, M. (2011a). Efficient exploration of reaction paths via a freezing string method. *Journal of Chemical Physics*, 135(22).

Behn, A., Zimmerman, P. M., Bell, A. T., and Head-Gordon, M. (2011b). Incorporating linear synchronous transit interpolation into the growing string method: Algorithm

and applications. *Journal of Chemical Theory and Computation*, 7(12):4019–4025.

Bell, R. P. (1959). The tunnel effect correction for parabolic potential barriers. *Transactions of the Faraday Society*, 55:1.

Benderskii, V. A., Makarov, D. E., and Wight, C. A. (2009). *Chemical dynamics at low temperatures*, volume 188. John Wiley & Sons.

Bernhard Schlegel, H. (1984). Estimating the hessian for gradient-type geometry optimizations. *Theoretica Chimica Acta*, 66(5):333–340.

Bessarab, P. F., Uzdin, V. M., and Jónsson, H. (2015). Method for finding mechanism and activation energy of magnetic transitions, applied to skyrmion and antivortex annihilation. *Computer Physics Communications*, 196:335–347.

Beyer, A. N., Richardson, J. O., Knowles, P. J., Rommel, J., and Althorpe, S. C. (2016). Quantum Tunneling Rates of Gas-Phase Reactions from On-the-Fly Instanton Calculations. *The Journal of Physical Chemistry Letters*, 7(21):4374–4379.

Birkholz, A. B. and Schlegel, H. B. (2015). Using bonding to guide transition state optimization. *Journal of computational chemistry*, 36(15):1157–1166.

Bitzek, E., Koskinen, P., Gähler, F., Moseler, M., and Gumbsch, P. (2006). Structural relaxation made simple. *Physical review letters*, 97(17):170201.

Bligaard, T. and Jónsson, H. (2005). Optimization of hyperplanar transition states: Application to 2D test problems. *Computer Physics Communications*, 169(1-3):284–288.

Bofill, J. M. (1994). Updated Hessian matrix and the restricted step method for locating transition structures. *Journal of Computational Chemistry*, 15(1):1–11.

Bohner, M. U., Meisner, J., and Kästner, J. (2013). A quadratically-converging nudged elastic band optimizer. *Journal of chemical theory and computation*, 9(8):3498–3504.

Bohner, M. U., Zeman, J., Smiatek, J., and Arnold, A. (2014). Nudged-elastic band used to find reaction coordinates based on the free energy. *Journal of Chemical Physics*, 140(7).

Born, M. and Oppenheimer, R. (1927). Zur Quantentheorie der Molekeln. *Annalen der Physik*, 389(20):457–484.

Broyden, C. G. (1970). The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations. *IMA Journal of Applied Mathematics*, 6(1):76–90.

Burger, S. K. and Yang, W. (2006). Quadratic string method for determining the minimum-energy path based on multiobjective optimization. *Journal of Chemical Physics*, 124(5):1–14.

Burger, S. K. and Yang, W. (2007). Sequential quadratic programming method for determining the minimum energy path. *The Journal of Chemical Physics*, 127(16):164107.

Burke, K. (2012). Perspective on density functional theory. *Journal of Chemical Physics*, 136(15).

Burke, K. and Wagner, L. O. (2013). Dft in a nutshell. *International Journal of Quantum Chemistry*, 113(2):96–101.

Callan, C. G. and Coleman, S. (1977). Fate of the false vacuum. II. First quantum corrections. *Physical Review D*, 16(6):1762–1768.

Castro, C. and Karney, W. L. (2020). Heavy-Atom Tunneling in Organic Reactions. *Angewandte Chemie - International Edition*, 59(22):8355–8366.

Ceperley, D. M. and Alder, B. J. (1980). Ground State of the Electron Gas by a

Stochastic Method. *Physical Review Letters*, 45(7):566–569.

Cerjan, C. J. and Miller, W. H. (1981). On finding transition states. *The Journal of chemical physics*, 75(6):2800–2806.

Chandler, D. and Wolynes, P. G. (1981). Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids. *The Journal of Chemical Physics*, 74(7):4078–4095.

Cheng, X., Jónsson, E., Jónsson, H., and Weber, P. M. (2018). Reply to: "The diamine cation is not a chemical example where density functional theory fails". *Nature Communications*, 9(1):5348.

Cheng, X., Zhang, Y., Jónsson, E., Jónsson, H., and Weber, P. M. (2016). Charge localization in a diamine cation provides a test of energy functionals and self-interaction correction. *Nature Communications*, 7(1):11013.

Chill, S. T., Stevenson, J., Ruehle, V., Shang, C., Xiao, P., Farrell, J. D., Wales, D. J., and Henkelman, G. (2014a). Benchmarks for Characterization of Minima, Transition States, and Pathways in Atomic, Molecular, and Condensed Matter Systems. *Journal of Chemical Theory and Computation*, 10(12):5476–5482.

Chill, S. T., Welborn, M., Terrell, R., Zhang, L., Berthet, J.-C., Pedersen, A., Jónsson, H., and Henkelman, G. (2014b). EON: software for long time simulations of atomic scale systems. *Modelling and Simulation in Materials Science and Engineering*, 22(5):055002.

Choi, C. and Elber, R. (1991). Reaction path study of helix formation in tetrapeptides: Effect of side chains. *The Journal of Chemical Physics*, 94(1):751–760.

Ciccotti, G., Ferrario, M., Laria, D., and Kapral, R. (1995). Simulation of quantum and classical activated processes in the condensed phase. In *Progress in Computational Physics of Matter*, pages 150–190. World Scientific.

Coleman, S. (1977). Fate of the false vacuum: Semiclassical theory. *Physical Review D*, 15(10):2929–2936.

Crehuet, R. and Field, M. J. (2003). A temperature-dependent nudged-elastic-band algorithm. *Journal of Chemical Physics*, 118(21):9563–9571.

Czerminski, R. and Elber, R. (1990). Self-avoiding walk between two fixed points as a tool to calculate reaction paths in large molecular systems. *International Journal of Quantum Chemistry*, 38(S24):167–185.

Deb, S., Cheng, X., and Weber, P. M. (2013). Structural dynamics and charge transfer in electronically excited n,n′-dimethylpiperazine. *The Journal of Physical Chemistry Letters*, 4(16):2780–2784.

Dennis Jr, J. E. and Schnabel, R. B. (1996). *Numerical methods for unconstrained optimization and nonlinear equations*. SIAM.

Dirac, P. A. (1930). Note on exchange phenomena in the thomas atom. In *Mathematical proceedings of the Cambridge philosophical society*, volume 26, pages 376–385. Cambridge University Press.

D.M. Einarsdóttir, A. Arnaldsson, F. O. and Jónsson, H. (2012). Path optimization with application to tunneling. *Lecture Notes in Computer Science*, 7134(45).

Doubleday, C., Armas, R., Walker, D., Cosgriff, C. V., and Greer, E. M. (2017). Heavy-Atom Tunneling Calculations in Thirteen Organic Reactions: Tunneling Contributions are Substantial, and Bell's Formula Closely Approximates Multidimensional Tunneling at $\geq$ 250 K. *Angewandte Chemie*, 129(42):13279–13282.

Dunham, J. L. (1932). The Wentzel-Brillouin-Kramers Method of Solving the Wave Equation. *Physical Review*, 41(6):713–720.

Dunning, T. H. (1989). Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90(2):1007–1023.

E, W., Ren, W., and Vanden-Eijnden, E. (2002). String method for the study of rare events. *Physical Review B*, 66(5):052301.

Elber, R. and Karplus, M. (1987). A method for determining reaction paths in large molecules: Application to myoglobin. *Chemical Physics Letters*, 139(5):375–380.

Evans, M. and Polanyi, M. (1938). Inertia and driving force of chemical reactions. *Transactions of the Faraday Society*, 34:11–24.

Eyring, H. (1935). The activated complex in chemical reactions. *The Journal of Chemical Physics*, 3(2):107–115.

Fang, Y., Liu, F., Barber, V. P., Klippenstein, S. J., McCoy, A. B., and Lester, M. I. (2016). Deep tunneling in the unimolecular decay of CH 3 CHOO Criegee intermediates to OH radical products. *The Journal of Chemical Physics*, 145(23):234308.

Fermi, E. (1927). A statistical method for the determination of some atomic properties. *Rend. Accad. Naz. Lincei*, 6:602–607.

Feynman, R. P., Hibbs, A. R., and Styer, D. F. (2010). *Quantum mechanics and path integrals*. Courier Corporation.

Fischer, T. H. and Almlöf, J. (1992). General methods for geometry and wave function optimization. *Journal of Physical Chemistry*, 96(24):9768–9774.

Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, 13(3):317–322.

Fletcher, R. (2013). *Practical methods of optimization*. John Wiley & Sons.

Fukui, K. (1981). The path of chemical reactions - the IRC approach. *Accounts of Chemical Research*, 14(12):363–368.

Galván, I. F. and Field, M. J. (2008). Improving the efficiency of the NEB reaction path finding algorithm. *Journal of Computational Chemistry*, 29(1):139–143.

Garrett, B. C., Abusalbi, N., Kouri, D. J., and Truhlar, D. G. (1985). Test of variational transition state theory and the least-action approximation for multidimensional tunneling probabilities against accurate quantal rate constants for a collinear reaction involving tunneling into an excited state. *The Journal of Chemical Physics*, 83(5):2252–2258.

Garrett, B. C., Truhlar, D. G., Wagner, A. F., and Dunning, T. H. (1983). Variational transition state theory and tunneling for a heavy-light-heavy reaction using an ab initio potential energy surface. 37Cl+H(D) 35Cl→H(D) 37Cl+35Cl. *The Journal of Chemical Physics*, 78(7):4400–4413.

Gill, P. M., Johnson, B. G., Pople, J. A., and Frisch, M. J. (1992). The performance of the becke–lee–yang–parr (b–lyp) density functional theory with various basis sets. *Chemical Physics Letters*, 197(4-5):499–505.

Gillan, M. J. (1987). Quantum-classical crossover of the transition rate in the damped double well. *Journal of Physics C: Solid State Physics*, 20(24):3621–3641.

Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of Computation*, 24(109):23–23.

Goodrow, A., Bell, A. T., and Head-Gordon, M. (2008). Development and application of a hybrid method involving interpolation and ab initio calculations for the

determination of transition states. *Journal of Chemical Physics*, 129(17).

Goodrow, A., Bell, A. T., and Head-Gordon, M. (2009). Transition state-finding strategies for use with the growing string method. *Journal of Chemical Physics*, 130(24).

Goumans, T. P. M., Catlow, C. R. A., Brown, W. A., Kästner, J., and Sherwood, P. (2009). An embedded cluster study of the formation of water on interstellar dust grains. *Physical Chemistry Chemical Physics*, 11(26):5431.

Griffiths, D. J. and Schroeter, D. F. (2018). *Introduction to quantum mechanics*. Cambridge University Press.

Grimme, S. (2004). Accurate description of van der waals complexes by density functional theory including empirical corrections. *Journal of computational chemistry*, 25(12):1463–1473.

Grimme, S. (2006). Semiempirical gga-type density functional constructed with a long-range dispersion correction. *Journal of computational chemistry*, 27(15):1787–1799.

Halgren, T. A. and Lipscomb, W. N. (1977). The synchronous-transit method for determining reaction pathways and locating molecular transition states. *Chemical Physics Letters*, 49(2):225–232.

Henkelman, G., Arnaldsson, A., and Jónsson, H. (2006). Theoretical calculations of ch_4 and h_2 associative desorption from ni (111): could subsurface hydrogen play an important role? *The Journal of chemical physics*, 124(4):044706.

Henkelman, G., Jóhannesson, G., and Jónsson, H. (2002). Methods for finding saddle points and minimum energy paths. In *Theoretical methods in condensed phase chemistry*, pages 269–302. Springer.

Henkelman, G., Jóhannesson, G., and Jónsson, H. (2005). Methods for Finding Saddle Points and Minimum Energy Paths. *Theoretical Methods in Condensed Phase Chemistry*, pages 269–302.

Henkelman, G. and Jónsson, H. (1999). A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *The Journal of chemical physics*, 111(15):7010–7022.

Henkelman, G. and Jónsson, H. (2000). Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *The Journal of chemical physics*, 113(22):9978–9985.

Henkelman, G. and Jónsson, H. (2001). Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table. *The Journal of Chemical Physics*, 115(21):9657–9666.

Henkelman, G., Uberuaga, B. P., and Jónsson, H. (2000). A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of chemical physics*, 113(22):9901–9904.

Herbol, H. C., Stevenson, J., and Clancy, P. (2017). Computational Implementation of Nudged Elastic Band, Rigid Rotation, and Corresponding Force Optimization. *Journal of Chemical Theory and Computation*, 13(7):3250–3259.

Hestenes, M. R., Stiefel, E., et al. (1952). *Methods of conjugate gradients for solving linear systems*, volume 49. NBS Washington, DC.

Heyden, A., Bell, A. T., and Keil, F. J. (2005). Efficient methods for finding transition states in chemical reactions: Comparison of improved dimer method and partitioned rational function optimization method. *The Journal of chemical physics*,

123(22):224101.

Hohenberg, P. and Kohn, W. (1964). Inhomogeneous electron gas. *Physical review*, 136(3B):B864.

Hratchian, H. P. and Schlegel, H. B. (2005). Using Hessian Updating To Increase the Efficiency of a Hessian Based Predictor-Corrector Reaction Path Following Method. *Journal of Chemical Theory and Computation*, 1(1):61–69.

Ishida, K., Morokuma, K., and Komornicki, A. (1977). The intrinsic reaction coordinate. An ab-initio calculation for HNC $\rightarrow$ HCN and H^-+CH_4 $\rightarrow$ CH_4+H^-. *The Journal of Chemical Physics*, 66(5):2153–2156.

Jensen, F. (2017). *Introduction to computational chemistry*. John wiley & sons.

Jóhannesson, G. H. and Jónsson, H. (2001). Optimization of hyperplanar transition states. *The Journal of Chemical Physics*, 115(21):9644–9656.

Johnston, H. S. and Heicklen, J. (1962). Tunnelling corrections for unsymmetrical Eckart potential energy barriers. *The Journal of Physical Chemistry*, 66(3):532–533.

Jonsson, H. (2011). Simulation of surface processes. *Proceedings of the National Academy of Sciences*, 108(3):944–949.

Jónsson, H., Mills, G., and Jacobsen, K. W. (1998). Nudged elastic band method for finding minimum energy paths of transitions.

Kästner, J. (2014). Theory and simulation of atom tunneling in chemical reactions. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(2):158–168.

Kästner, J. and Sherwood, P. (2008). Superlinearly converging dimer method for transition state search. *Journal of Chemical Physics*, 128(1).

Keck, J. C. (1967). Variational theory of reaction rates. *Adv. Chem. Phys*, 13:85–121.

Koch, W. and Holthausen, M. (2001). A chemist's guide to density functional theory, wiley.

Kohn, W. and Sham, L. J. (1965). Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133.

Koistinen, O.-P., Ásgeirsson, V., Vehtari, A., and Jónsson, H. (2019). Nudged elastic band calculations accelerated with gaussian process regression based on inverse interatomic distances. *Journal of chemical theory and computation*, 15(12):6738–6751.

Koistinen, O.-P., Dagbjartsdóttir, F. B., Ásgeirsson, V., Vehtari, A., and Jónsson, H. (2017). Nudged elastic band calculations accelerated with gaussian process regression. *The Journal of chemical physics*, 147(15):152720.

Kolsbjerg, E. L., Groves, M. N., and Hammer, B. (2016). An automated nudged elastic band method. *The Journal of Chemical Physics*, 145(9):094107.

Koslover, E. F. and Wales, D. J. (2007). Comparison of double-ended transition state search methods. *Journal of Chemical Physics*, 127(13).

Langer, J. (1969). Statistical theory of the decay of metastable states. *Annals of Physics*, 54(2):258–275.

Lauhon, L. J. and Ho, W. (2000). Direct Observation of the Quantum Tunneling of Single Hydrogen Atoms with a Scanning Tunneling Microscope. *Physical Review Letters*, 85(21):4566–4569.

Lee, C., Yang, W., and Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37(2):785–789.

Li, Q. S., Zhang, J., and Zhang, S. (2005). A dft and ab initio direct dynamics study on the hydrogen abstract reaction of h3bnh3 → h2+ h2bnh2. *Chemical physics letters*, 404(1-3):100–106.

Malek, R. and Mousseau, N. (2000). Dynamics of Lennard-Jones clusters: A characterization of the activation-relaxation technique. *Physical Review E*, 62(6):7723–7728.

Manthe, U., Meyer, H. D., and Cederbaum, L. S. (1992). Wave-packet dynamics within the multiconfiguration Hartree framework: General aspects and application to NOCl. *The Journal of Chemical Physics*, 97(5):3199–3213.

Maragakis, P., Andreev, S. A., Brumer, Y., Reichman, D. R., and Kaxiras, E. (2002). Adaptive nudged elastic band approach for transition state calculation. *Journal of Chemical Physics*.

Maras, E., Pizzagalli, L., Ala-Nissila, T., and Jonsson, H. (2017). Atomic scale formation mechanism of edge dislocation relieving lattice strain in a gesi overlayer on si (001). *Scientific reports*, 7(1):1–9.

Maras, E., Trushin, O., Stukowski, A., Ala-Nissila, T., and Jonsson, H. (2016). Global transition path search for dislocation formation in ge on si (001). *Computer Physics Communications*, 205:13–21.

Marcus, R. A. (1952). Unimolecular Dissociations and Free Radical Recombination Reactions. *The Journal of Chemical Physics*, 20(3):359–364.

Marcus, R. A. (1966). On the Analytical Mechanics of Chemical Reactions. Quantum Mechanics of Linear Collisions. *The Journal of Chemical Physics*, 45(12):4493–4499.

Marcus, R. A. and Coltrin, M. E. (1977). A new tunneling path for reactions such as $H+H_2 \rightarrow H_2+H$. *The Journal of Chemical Physics*, 67(6):2609–2613.

Marcus, R. A. and Rice, O. K. (1951). The Kinetics of the Recombination of Methyl Radicals and Iodine Atoms. *The Journal of Physical Chemistry*, 55(6):894–908.

Maronsson, J. B., Jónsson, H., and Vegge, T. (2012). A method for finding the ridge between saddle points applied to rare event rate estimates. *Physical Chemistry Chemical Physics*, 14(8):2884.

Masgrau, L. (2006). Atomic Description of an Enzyme Reaction Dominated by Proton Tunneling. *Science*, 312(5771):237–241.

Mathiesen, N. R., Jónsson, H., Vegge, T., and García Lastra, J. M. (2019). R-NEB: Accelerated Nudged Elastic Band Calculations by Use of Reflection Symmetry. *Journal of Chemical Theory and Computation*.

McPherson, K. E., Bartolotti, L. J., Morehead, A. T., and Sargent, A. L. (2016). Utility of the Nudged Elastic Band Method in Identifying the Minimum Energy Path of an Elementary Organometallic Reaction Step. *Organometallics*, 35(11):1861–1865.

Meisner, J. and Kästner, J. (2016). Atom Tunneling in Chemistry. *Angewandte Chemie International Edition*, 55(18):5400–5413.

Melander, M., Laasonen, K., and Jónsson, H. (2015). Removing external degrees of freedom from transition-state search methods using quaternions. *Journal of chemical theory and computation*, 11(3):1055–1062.

Messina, M., Schenter, G. K., and Garrett, B. C. (1995). A variational centroid density procedure for the calculation of transmission coefficients for asymmetric barriers at low temperature. *The Journal of Chemical Physics*, 103(9):3430–3435.

Meyer, H. D., Manthe, U., and Cederbaum, L. S. (1990). The multi-configurational

time-dependent Hartree approach. *Chemical Physics Letters*, 165(1):73–78.
Miller, W. H. (1975). Semiclassical limit of quantum mechanical transition state theory for nonseparable systems. *The Journal of Chemical Physics*, 62(5):1899–1906.
Miller, W. H., Handy, N. C., and Adams, J. E. (1980). Reaction path Hamiltonian for polyatomic molecules. *The Journal of Chemical Physics*, 72(1):99–112.
Mills, G., Jónsson, H., and Schenter, G. K. (1995). Reversible work transition state theory: application to dissociative adsorption of hydrogen. *Surface Science*, 324(2-3):305–337.
Mills, G., Schenter, G., Makarov, D., and Jónsson, H. (1997). Generalized path integral based quantum transition state theory. *Chemical Physics Letters*, 278(1-3):91–96.
Mills, G., Schenter, G., Makarov, D., and Jónsson, H. (1998). Raw quantum transition state theory.
Møller, C. and Plesset, M. S. (1934). Note on an approximation treatment for many-electron systems. *Physical review*, 46(7):618.
Müller, K. and Brown, L. D. (1979). Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theoretica chimica acta*, 53(1):75–93.
Munro, L. J. and Wales, D. J. (1999). Defect migration in crystalline silicon. *Physical Review B*, 59(6):3969–3980.
Murtagh, B. A. (1970). Computational experience with quadratically convergent minimisation methods. *The Computer Journal*, 13(2):185–194.
Neese, F. (2018). Software update: The orca program system, version 4.0. *WIREs Comput. Mol. Sci.*, 8(1):e1327.
Nesbet, R. (1955). Configuration interaction in orbital theories. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 230(1182):312–321.
Nocedal, J. (1980). Updating quasi-Newton matrices with limited storage. *Mathematics of Computation*, 35(151):773–773.
Nocedal, J. and Wright, S. (2006). *Numerical optimization*. Springer Science & Business Media.
Nosikov, I. A., Klimenko, M. V., Zhbankov, G. A., Podlesnyi, A. V., Ivanova, V. A., and Bessarab, P. F. (2020). Generalized Force Approach to Point-to-Point Ionospheric Ray Tracing and Systematic Identification of High and Low Rays. *IEEE Transactions on Antennas and Propagation*, 68(1):455–467.
Olender, R. and Elber, R. (1997). Yet another look at the steepest descent path. *Journal of Molecular Structure: theochem*, 398-399:63–71.
Olsen, R. A., Kroes, G. J., Henkelman, G., Arnaldsson, A., and Jónsson, H. (2004). Comparison of methods for finding saddle points without knowledge of the final states. *Journal of Chemical Physics*, 121(20):9776–9792.
Peng, C. and Bernhard Schlegel, H. (1993). Combining synchronous transit and quasi-newton methods to find transition states. *Israel Journal of Chemistry*, 33(4):449–454.
Perdew, J. P., Burke, K., and Ernzerhof, M. (1996a). Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865.
Perdew, J. P., Ernzerhof, M., and Burke, K. (1996b). Rationale for mixing exact exchange with density functional approximations. *The Journal of chemical physics*, 105(22):9982–9985.

Perdew, J. P. and Ruzsinszky, A. (2010). Fourteen easy lessons in density functional theory. *International Journal of Quantum Chemistry*, 110(15):2801–2807.

Perdew, J. P., Ruzsinszky, A., Constantin, L. A., Sun, J., and Csonka, G. I. (2009). Some fundamental issues in ground-state density functional theory: A guide for the perplexed. *Journal of chemical theory and computation*, 5(4):902–908.

Perdew, J. P. and Schmidt, K. (2001). Jacob's ladder of density functional approximations for the exchange-correlation energy. In *AIP Conference Proceedings*, volume 577, pages 1–20. American Institute of Physics.

Peters, B. (2017). *Reaction rate theory and rare events*. Elsevier.

Peters, B., Heyden, A., Bell, A. T., and Chakraborty, A. (2004). A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *The Journal of chemical physics*, 120(17):7877–7886.

Peterson, A. A. (2016). Acceleration of saddle-point searches with machine learning. *The Journal of chemical physics*, 145(7):074106.

Plasencia Gutiérrez, M., Argáez, C., and Jónsson, H. (2017). Improved minimum mode following method for finding first order saddle points. *Journal of chemical theory and computation*, 13(1):125–134.

Pople, J. A., Head-Gordon, M., and Raghavachari, K. (1987). Quadratic configuration interaction. a general technique for determining electron correlation energies. *J. Chem. Phys.*, 87(10):5968–5975.

Ren, W. (2003). Higher Order String Method for Finding Minimum Energy Paths. *Communications in Mathematical Sciences*, 1(2):377–384.

Ren, W. and Vanden-Eijnden, E. (2013). A climbing string method for saddle point search. *Journal of Chemical Physics*, 138(13).

Richardson, J. O. (2016). Derivation of instanton rate theory from first principles. *The Journal of Chemical Physics*, 144(11):114106.

Richardson, J. O. (2018). Perspective: Ring-polymer instanton theory. *Journal of Chemical Physics*, 148(20).

Röder, K. and Wales, D. J. (2020). Improving double-ended transition state searches for soft-matter systems. *Journal of Chemical Physics*, 153(3).

Rommel, J. B., Goumans, T. P. M., and Kästner, J. (2011). Locating Instantons in Many Degrees of Freedom. *Journal of Chemical Theory and Computation*, 7(3):690–698.

Rommel, J. B. and Kästner, J. (2011). Adaptive integration grids in instanton theory improve the numerical accuracy at low temperature. *The Journal of chemical physics*, 134(18):184107.

Rommel, J. B., Liu, Y., Werner, H.-J., and Kästner, J. (2012). Role of Tunneling in the Enzyme Glutamate Mutase. *The Journal of Physical Chemistry B*, 116(46):13682–13689.

Roothaan, C. C. J. (1951). New Developments in Molecular Orbital Theory. *Reviews of Modern Physics*, 23(2):69–89.

Samanta, A. and E, W. (2013). Optimization-Based String Method for Finding Minimum Energy Path. *Communications in Computational Physics*, 14(2):265–275.

Schäfer, A., Horn, H., and Ahlrichs, R. (1992). Fully optimized contracted gaussian basis sets for atoms li to kr. *The Journal of Chemical Physics*, 97(4):2571–2577.

Schlegel, H. B. (1982). Optimization of equilibrium geometries and transition structures. *Journal of Computational Chemistry*, 3(2):214–218.

Schlegel, H. B. (2003). Exploring potential energy surfaces for chemical reactions: An overview of some practical methods. *Journal of Computational Chemistry*, 24(12):1514–1527.

Schlegel, H. B. (2011). Geometry optimization. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5):790–809.

Schmerwitz, Y., Ásgeirsson, V., Engels, B., and Jónsson, H. (2021). Growing nudged elastic band method for initial reaction path generation on the image-dependent pair potential surface. *To be submitted*.

Shanno, D. F. (1970). Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, 24(111):647–647.

Shannon, R. J., Blitz, M. A., Goddard, A., and Heard, D. E. (2013). Accelerated chemistry in the reaction between the hydroxyl radical and methanol at interstellar temperatures facilitated by tunnelling. *Nature Chemistry*, 5(9):745–749.

Sheppard, D., Terrell, R., and Henkelman, G. (2008). Optimization methods for finding minimum energy paths. *The Journal of chemical physics*, 128(13):134106.

Sheppard, D., Xiao, P., Chemelewski, W., Johnson, D. D., and Henkelman, G. (2012). A generalized solid-state nudged elastic band method. *The Journal of Chemical Physics*, 136(7):074103.

Skodje, R. T., Truhlar, D. G., and Garrett, B. C. (1981). A general small-curvature approximation for transition-state-theory transmission coefficients. *Journal of Physical Chemistry*, 85(21):3019–3023.

Slater, J. C. (1928). The Self Consistent Field and the Structure of Atoms. *Physical Review*, 32(3):339–348.

Smidstrup, S., Pedersen, A., Stokbro, K., and Jónsson, H. (2014). Improved initial guess for minimum energy path calculations. *The Journal of chemical physics*, 140(21):214106.

Steckel, J. A. (2009). *Density functional theory: a practical introduction*. Wiley.

Stephens, P. J., Devlin, F. J., Chabalowski, C. F., and Frisch, M. J. (1994). Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry*, 98(45):11623–11627.

Sun, J., Remsing, R. C., Zhang, Y., Sun, Z., Ruzsinszky, A., Peng, H., Yang, Z., Paul, A., Waghmare, U., Wu, X., Klein, M. L., and Perdew, J. P. (2016). Accurate first-principles structures and energies of diversely bonded systems from an efficient density functional. *Nature Chemistry*, 8(9):831–836.

Sutcliffe, M. J. and Scrutton, N. S. (2002). A new conceptual framework for enzyme catalysis. *European Journal of Biochemistry*, 269(13):3096–3102.

Tao, J., Perdew, J. P., Staroverov, V. N., and Scuseria, G. E. (2003). Climbing the Density Functional Ladder: Nonempirical Meta–Generalized Gradient Approximation Designed for Molecules and Solids. *Physical Review Letters*, 91(14):146401.

Thomas, L. H. (1927). The calculation of atomic fields. In *Mathematical proceedings of the Cambridge philosophical society*, volume 23, pages 542–548. Cambridge University Press.

Truhlar, D. G., Garrett, B. C., and Klippenstein, S. J. (1996). Current status of transition-state theory. *The Journal of physical chemistry*, 100(31):12771–12800.

Truhlar, D. G. and Kuppermann, A. (1971). Exact Tunneling Calculations. *Journal of the American Chemical Society*, 93(8):1840–1851.

Trygubenko, S. A. and Wales, D. J. (2004). A doubly nudged elastic band method for finding transition states. *The Journal of Chemical Physics*, 120(5):2082–2094.

Vineyard, G. H. (1957). Frequency factors and isotope effects in solid state rate processes. *Journal of Physics and Chemistry of Solids*, 3(1-2):121–127.

Vosko, S. H., Wilk, L., and Nusair, M. (1980). Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics*, 58(8):1200–1211.

Voter, A. F. (1997). Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events. *Physical Review Letters*, 78(20):3908–3911.

Voter, A. F. and Doll, J. D. (1985). Dynamical corrections to transition state theory for multistate systems: Surface self-diffusion in the rare-event regime. *The Journal of Chemical Physics*, 82(1):80–92.

Wigner, E. (1932). Über das überschreiten von potentialschwellen bei chemischen reaktionen. *Z. Phys. Chem.*, 15:203.

Wigner, E. (1938). The transition state method. *Transactions of the Faraday Society*, 34:29–41.

Witt, W. C., del Rio, B. G., Dieterich, J. M., and Carter, E. A. (2018). Orbital-free density functional theory for materials research. *Journal of Materials Research*, 33(7):777–795.

Zarkevich, N. A. and Johnson, D. D. (2015). Nudged-elastic band method with two climbing images: Finding transition states in complex energy landscapes. *Journal of Chemical Physics*.

Zhang, J., Zhang, H., Ye, H., and Zheng, Y. (2016). Free-end adaptive nudged elastic band method for locating transition states in minimum energy path calculation. *The Journal of chemical physics*, 145(9):094104.

Zhu, T., Li, J., Samanta, A., Kim, H. G., and Suresh, S. (2007). Interfacial plasticity governs strain rate sensitivity and ductility in nanostructured metals. *Proceedings of the National Academy of Sciences*, 104(9):3031–3036.

Zhu, X., Thompson, K. C., and Martínez, T. J. (2019). Geodesic interpolation for reaction pathways. *Journal of Chemical Physics*, 150(16).

Zimmerman, P. (2013a). Reliable transition state searches integrated with the growing string method. *Journal of chemical theory and computation*, 9(7):3043–3050.

Zimmerman, P. M. (2013b). Growing string method with interpolation and optimization in internal coordinates: Method and examples. *Journal of Chemical Physics*, 138(18).

Zimmerman, P. M. (2015). Single-ended transition state finding with the growing string method. *Journal of Computational Chemistry*, 36(9):601–611.

Article I

Efficient evaluation of atom tunneling combined with electronic structure calculations

Ásgeirsson, V., Arnaldsson, A. and Jónsson, H.
J. Chem. Phys., **148**, 102334, 2018

THE JOURNAL OF CHEMICAL PHYSICS **148**, 102334 (2018)

Efficient evaluation of atom tunneling combined with electronic structure calculations

Vilhjálmur Ásgeirsson,[1] Andri Arnaldsson,[1,2] and Hannes Jónsson[1,3,a)]
[1]*Science Institute and Faculty of Physical Sciences, University of Iceland VR-III, 107 Reykjavík, Iceland*
[2]*Vatnaskil, Síduúli 28, 108 Reykjavík, Iceland*
[3]*Center for Nonlinear Studies, Los Alamos, New Mexico 87545, USA*

(Received 1 October 2017; accepted 18 December 2017; published online 9 January 2018)

Methodology for finding optimal tunneling paths and evaluating tunneling rates for atomic rearrangements is described. First, an optimal JWKB tunneling path for a system with fixed energy is obtained using a line integral extension of the nudged elastic band method. Then, a calculation of the dynamics along the path is used to determine the temperature at which it corresponds to an optimal Feynman path for thermally activated tunneling (instanton) and a harmonic approximation is used to estimate the transition rate. The method is illustrated with calculations for a modified two-dimensional Müller-Brown surface but is efficient enough to be used in combination with electronic structure calculations of the energy and atomic forces in systems containing many atoms. An example is presented where tunneling is the dominant mechanism well above room temperature as an H_3BNH_3 molecule dissociates to form H_2. Also, a solid-state example is presented where density functional theory calculations of H atom tunneling in a Ta crystal give close agreement with experimental measurements on hydrogen diffusion over a wide range in temperature. *Published by AIP Publishing.*
https://doi.org/10.1063/1.5007180

I. INTRODUCTION

At low enough temperature, quantum mechanical tunneling rather than the classical over-the-barrier mechanism becomes the dominant transition mechanism for atomic rearrangements such as chemical reactions and diffusion events. The lighter the atoms, the higher the crossover temperature is between the two mechanisms. While tunneling is typically only important well below room temperature, there are cases where it is the preferred mechanism at room temperature. One such example, for which calculations are presented below, is the dissociation of an ammoniaborane molecule, H_3BNH_3, to give H_2. For this reaction, tunneling is estimated to be the dominant transition mechanism at room temperature. When the crossover from over-the-barrier mechanism to tunneling is smooth, the crossover temperature can be estimated from the first-order saddle point characterizing the classical transition mechanism.[1] The saddle point is the (highest) maximum along the minimum energy path (MEP) connecting the initial and final state minima on the potential energy surface. There, one of the eigenvalues of the Hessian is negative and corresponds to an imaginary frequency. Based on its magnitude, $|\omega^{\ddagger}|$, the crossover temperature can be estimated as

$$T_c = \frac{\hbar|\omega^{\ddagger}|}{2\pi k_B}, \tag{1}$$

where $\hbar$ and k_B are the Planck and Boltzmann constants, respectively. This estimate is based on a harmonic approximation and gives only a rough estimate of the temperature at which tunneling starts to play an important role, as can be seen from the examples below. The narrower the energy barrier, the larger the $\omega^{\ddagger}$ is and the higher the crossover temperature is. Above T_c, there can still be large nuclear quantum effects influencing the transition rate, but these are usually referred to as zero point energy effects.

Below T_c, the calculation of the transition rate needs to take tunneling into account. This requires a full quantum mechanical description of the position of the atomic nuclei. The statistical Feynman path integral formalism[2] is particularly convenient and lends itself well to numerical calculations. There, the quantum statistical mechanics of the system become mathematically equivalent to the classical statistical mechanics of a ring polymer of P replicas of the system connected by temperature dependent springs.[3,4] Rigorous thermal averages are obtained as $P \to \infty$. A configuration of the ring polymer for a system with N atomic coordinates, $\mathbf{r} = \{x_1, x_2, \ldots, x_N\}$, is specified by the coordinates of the replicas, $\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_P\}$. The energy surface characterizing the classical system, $V(\mathbf{r})$, then effectively gets replaced by an effective quantum mechanical energy surface for the ring polymer,

$$V^{rp}(\mathbf{R}, T) = \sum_{j=1}^{P} \frac{k_{sp}}{2}|\mathbf{r}_{mod(j,P)+1} - \mathbf{r}_j|^2 + \frac{V(\mathbf{r}_j)}{P}, \tag{2}$$

where the spring constant is dependent on the temperature, mass, and P,

$$k_{sp} = \mu P\left(\frac{k_B T}{\hbar}\right)^2. \tag{3}$$

The distribution of the replicas in configuration space represents the quantum delocalization of the nuclei. Here, μ is

a)Electronic mail: hj@hi.is

the effective mass which is conveniently incorporated in the calculations by using mass weighted coordinates. Early formulations of a quantum mechanical extension of transition state theory[5] are based on Feynman path integrals centered on the average coordinate of the replicas in the ring polymer, the centroid, essentially treating it as a classical coordinate.[1,6–10] This, however, turned out to work only for symmetric energy barriers. In general, a good definition of a quantum transition state requires going beyond a description in terms of the centroid.[11,12] It requires specifying a constraint that incorporates the shape of the ring polymer as well as the location of the centroid.

The reason for this can be seen by visualising the effective energy surface V^{rp} as a function of temperature.[11–13] Figure 1 shows a contour graph of V^{rp} for a one-dimensional asymmetric Eckart barrier in terms of the first two Fourier components of the ring polymer. The energy ridge separating the initial state from the final state does not line up with a constraint placed on the centroid (a vertical line in this figure). It only does so by chance for a symmetric barrier. Minima on the energy ridge, i.e., first-order saddle points on the V^{rp} surface, corresponding to non-zero quantum delocalization, appear as temperature is lowered below T_c. They represent centers of regions where the ring polymer most easily can pass through the energy ridge on its way from the initial state to the final state. An optimal path on this effective energy surface connecting the initial and final state minima goes through a first-order saddle point on the V^{rp} surface and is the quantum mechanical analog of the classical MEP. It is referred to as the minimum action path (MAP), see Fig. 1.[11,12] The saddle points on the V^{rp} surface are often referred to as instantons.

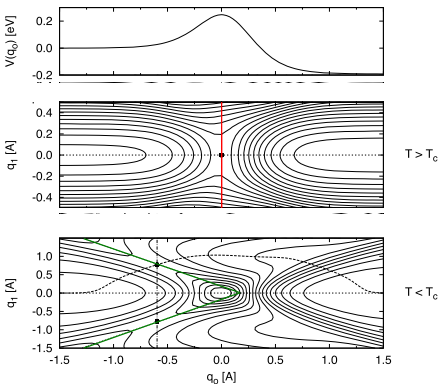


FIG. 1. Illustration of the temperature dependent, quantum mechanical effective energy surface, V^{rp}, given by Eq. (2) for a one-dimensional asymmetric Eckart barrier. Top panel: The Eckart potential energy curve, $V(q_0)$. Middle panel: Contour graph of $V^{rp}(q_0, q_1)$ when the ring polymer is of the form $q(\tau) = q_0 + q_1 \sin(2\pi\tau/\beta\hbar)$ with q_0 being the centroid coordinate and q_1 being the first Fourier component representing quantum delocalization. When $T > T_c$, the effective energy, V^{rp}, is minimal for a given q_0 with all replicas in the ring polymer at the same location. The reaction coordinate is then the classical minimum energy path, shown with a dashed horizontal line. The HTST transition state is shown with a red line going through the maximum of $V(q_0)$ (black circle), a first-order saddle point on the V^{rp} surface. Bottom panel: V^{rp} when $T < T_c$. The minimal value of V^{rp} for a given q_0 now corresponds to an extended ring polymer, representative of quantum delocalization and tunneling through the barrier. Two saddle points (instantons), equivalent by symmetry, emerge for non-zero q_1 (black circles). They lie on the minimum action path, a path of minimal V^{rp}, connecting the initial and final states. The conical HQTST transition state is marked with green lines. Similar figures are found in Refs. 11 and 12.

A quantum mechanical extension of transition state theory (QTST) has been formulated based on these considerations.[11,12] It is based on the ansatz that a quantum mechanical transition state, the bottleneck for a tunneling event, should represent a region in the ring polymer configuration space that corresponds to maximal free energy as it transits from the initial state to the final state. The tunneling rate is then determined by the rise in free energy from the initial state to this quantum mechanical transition state. It is necessary to go beyond a hyperplanar representation of the dividing surface and form a cone because the labeling of the P replicas of the systems is arbitrary and there is a mode of constant energy corresponding to translation of the replicas along the path, the so-called zero mode. Such a conical transition state dividing surface that includes the first-order saddle points is shown in Fig. 1. This ansatz for QTST has more recently been supported by direct dynamical simulations of ring polymers where remarkably good agreement with accurate tunneling rates has been obtained.[14,15] QTST can be related to the transition-state theory limit of ring-polymer molecular dynamics[16] and thus to the generalized flux correlation function.[17]

QTST can be implemented by calculating the free energy increase going from the initial state to a variationally optimized quantum transition state evaluated by reversible work calculations, as has, for example, been done for H_2 associative desorption and dissociative adsorption on surfaces.[11,12,18] However, such free energy calculations require extensive sampling of configuration space and are computationally too demanding in most cases to be combined with electronic structure calculations of the energy of the system and the atomic forces.

Even for classical systems, transition rates are most often carried out without explicit evaluation of the free energy. Instead, a quadratic approximation of the energy surface in the critical regions, i.e., around the initial state minimum and the first-order saddle point, is made for degrees of freedom corresponding to changes in relative distances between atoms. For finite systems, such as molecules and clusters, translation and rotation need to be treated separately if there is a change in these degrees of freedom during the transition. Neglecting such cases, the quadratic approximation leads to a closed expression for the rate constant, a harmonic approximation to transition state theory (HTST), which is by far the most widely used approach in calculations of rates of transitions involving atomic rearrangements. The expression for the rate constant, leaving out possible changes in translational and rotational degrees of freedom in the system during the transition, is[19]

$$k^{\mathrm{HTST}}(T) = \frac{1}{2\pi} \frac{\prod_i^N \omega_i^m}{\prod_i^{N-1} \omega_i^{\ddagger}} e^{-(V(\mathbf{r}_\ddagger) - V(\mathbf{r}_m))/k_B T}, \qquad (4)$$

where $V(\mathbf{r}_\ddagger)$ is the energy at the first-order saddle point, $V(\mathbf{r}_m)$ is the energy at the initial state minimum, and ω_i are the corresponding vibrational frequencies obtained from the eigenvalues of the Hessian matrix, the matrix of second derivatives of V with respect to the coordinates of the atoms. There is one fewer eigenvalue for the transition state dividing surface than the initial state since the dividing surface is a hyperplane with one lower dimensionality than the initial state. The normal to the hyperplane is taken to be parallel to the eigenvector corresponding to the negative eigenvalue at the first-order saddle point. The HTST expression for the rate constant corresponds to the Arrhenius law with activation energy $\Delta E = V(\mathbf{r}_\ddagger) - V(\mathbf{r}_m)$. The pre-exponential factor takes into account the vibrational entropy of the initial and transition states.

Nuclear quantum effects can to some extent be taken into account in an extension of HTST where the quantum mechanical harmonic oscillator partition function is used for each vibrational mode in the transition state and in the initial state, instead of the classical, high-temperature limit of the vibrational partition function. The rate expression then becomes

$$k^{\text{qq-HTST}}(T) = \frac{k_B T}{2\pi\hbar} \frac{\prod_i^{3N} 2\sinh\left(\hbar\omega_i^R/2k_BT\right)}{\prod_i^{3N-1} 2\sinh\left(\hbar\omega_i^\ddagger/2k_BT\right)} e^{-\Delta E/k_BT}. \tag{5}$$

We will refer to this as quasi-quantum TST (qq-HTST). This level of approximation has, for example, been used in calculations of H_2 and CH_4 dissociative adsorption on surfaces.[20] The qq-HTST approximation includes the effect of zero point motion but not tunneling. In the high-temperature limit, the zero point energy effects disappear and the qq-HTST and HTST rate approximations coincide.

The quantum mechanical tunneling rate can be estimated in a way that is analogous to HTST, i.e., within a harmonic approximation around a first-order saddle point on the effective energy surface, $V^{\text{rp}}(\mathbf{R})$, and around the initial state minimum.[13,21–26] We will refer to this approach as harmonic quantum transition state theory (HQTST), but it is more generally referred to as instanton theory or $\text{Im}\mathcal{F}$ theory. The first-order saddle points on the V^{rp} surface correspond to extended tunneling paths where the system advances from the initial state to the final state by shifting one replica after another over the energy barrier, rather than the whole ring polymer as in the over-the-barrier mechanism. The HQTST rate expression (excluding rotational and translational contributions) can be written as

$$k^{\text{HQTST}}(T) = \frac{1}{Q_R}\sqrt{\frac{S_0}{2\pi\hbar}} \frac{k_BTP}{\hbar\prod_{i=1}^{NP-1}|\Omega_{i,\ddagger}^{rp}|} e^{-(V^{rp}(\mathbf{R}_\ddagger)-V(\mathbf{r}_m))/k_BT}, \tag{6}$$

where $\Omega_{i,\ddagger}^{rp}$ denote the vibrational frequencies of the ring polymer at the first-order saddle point $\mathbf{R}_\ddagger$ (obtained from the NP-dimensional Hessian matrix of the ring polymer) and Q_R is the initial state partition function, which is approximated here as

$$Q_R = \prod_{i=1}^{N}\prod_{k=1}^{P}\left[4\sin^2\left(\frac{\pi k}{P}\right) + \left(\frac{\hbar\omega_i}{k_BTP}\right)^2\right]^{-\frac{1}{2}}. \tag{7}$$

At the initial state, the ring polymer collapses to the minimum except for thermal fluctuations. There are many similarities but also differences between this expression for the tunneling rate and the classical over-the-barrier rate given by Eq. (4). Again, the evaluation of the rate expression involves calculating the eigenvalues of the Hessian matrix, now the second derivatives of V^{rp} with respect to the coordinates of the atoms in all the replicas at the first-order saddle point, $\mathbf{R}_\ddagger$. At the initial state minimum, it is sufficient to evaluate the vibrational frequencies of the N-dimensional Hessian since all the replicas collapse to the minimum energy configuration. The effect of the springs is added explicitly. One of the degrees of freedom of the ring polymer corresponds to a rotation of the replicas along the path and needs to be treated separately, whereas it involves no change in the energy. It gives only an entropic contribution which is proportional to the length of the ring polymer,

$$S_0 = \frac{\mu P k_B T}{\hbar}\sum_{j=1}^{P}|\mathbf{r}_{mod(j,P)+1} - \mathbf{r}_{j-1}|^2. \tag{8}$$

When calculations are carried out for finite systems, such as molecules and clusters of atoms, overall rotation and translation need to be projected out of the NP-dimensional Hessian matrix and the entropy associated with these degrees of freedom evaluated separately. The pre-exponential factor in Eq. (6) and the initial state partition function, Q_R, are then written as a product of the translational, rotational, and vibrational contributions to the ring polymer partition function at the first-order saddle point and the initial state. The ring-polymer is taken to be a "super-molecule" and the rotation treated classically, where each replica only accounts for $1/P$-th of the total mass. The translational contributions are omitted since they are the same for the saddle point and initial state minimum and cancel out in the expression for the rate constant. For a more in-depth discussion of finite systems, the reader is referred to Ref. 27.

The HQTST approximation has been found to work remarkably well in many cases, for example, it gives results that are in close agreement with accurate multiconfigurational, time-dependent Hartree calculations of H atom reaction with CH_4 to form H_2 and CH_3.[28] Also, calculations of H_2 associative desorption from a Cu(110) surface agree well with full QTST calculations.[18,29] HQTST has also been applied to various biochemical reactions.[30,31] A recent review lists several other applications.[32]

In some previous HQTST calculations, the first-order saddle point on the ring polymer energy surface, V^{rp}, was found by calculating a nudged elastic band of Feynman paths so as to determine the full MAP[8,10–12] or by using the minimum mode following (MMF) method to converge on the first-order saddle point (the maximum along the path).[33,34] This, however, turns out to be a computationally demanding task. Tight convergence is needed in order to determine the saddle point accurately enough, so as to obtain converged eigenvalues of the Hessian matrix. This has turned out to be particularly challenging when the energy and atomic forces are obtained from *ab initio* or density functional theory (DFT) calculations.

In this article, an alternative method for evaluating HQTST tunneling rates is discussed. It is found to perform well in calculations combined with electronic structure evaluations of the energy and atomic forces. First, an optimal tunneling path (OTP) for a system of fixed energy is found within the JWKB approximation.[35] The method for finding the OTP involves minimizing a line integral over a path and we refer to the optimization method as the line integral nudged elastic band method (LI-NEB). It can be considered to be an extension of the NEB method for finding minimum energy paths.[10,36] There, discretization points representing the path are placed in some convenient way, preferably evenly distributed. The shape of the OTP corresponds also to a classical periodic orbit on the inverted potential surface and coincides with the shape of the ring polymer at a first-order saddle point on the V^{rp} surface, as discussed in Sec. II B. By carrying out the calculation of the classical dynamics along the path, the corresponding temperature and the location of the replicas of the ring polymer at the first-order saddle point can be determined. Finally, the NP-dimensional Hessian matrix for the ring polymer is constructed, the eigenvalues are calculated, and the thermal tunneling rate is evaluated. A preliminary report of this approach in a conference proceedings has previously been published.[29] Here, we revisit the methodology and apply it to molecular dissociation and diffusion in a solid as well as an illustration based on a modified Müller-Brown potential. Related methods, similarly based on an abbreviated Hamilton-Jacobi action, have been used for optimizing paths in various contexts.[37–39]

It has recently been shown that calculations of MEPs using the NEB method can be accelerated by using Gaussian process regression (GPR), where the number of energy and force evaluations needed to reach convergence could be reduced by an order of magnitude.[40,41] We expect that the computational effort of LI-NEB calculations (based on *ab initio* or DFT calculations) could be reduced significantly using GPR. In fact, since the OTP calculation and subsequent evaluation of the thermal rate constant involve a large number of energy and force evaluations over a rather limited region of the energy surface, the use of GPR can be expected to lead to an even larger reduction in computational effort in HQTST calculations than in classical HTST calculations.

The article is organized as follows: The methodology is described in Sec. II, first the LI-NEB method for finding OTPs, illustrated with a calculation for a modified Müller-Brown two-dimensional energy surface.[42] Then, the equivalence of an OTP for a system at a given energy and a ring polymer saddle point configuration for tunneling in a system coupled to a heat bath is discussed, and the method for generating the distribution of the replicas and evaluating the rate of tunneling is described. We address the accuracy and efficiency of the approach by computing the reactive flux for associative desorption of H_2 from Cu(110). In Sec. III, two applications of HQTST are described, the dissociation of an H_3BNH_3 molecule to form H_2 and H atom diffusion in a Ta crystal. In the latter, the results are compared with experimental measurements of the diffusion rate. The article concludes with a discussion section.

II. METHODS

In this section, we describe a robust and efficient method for finding an OTP. Then, the equivalence between the shape of the OTP and a classical periodic orbit on the inverted potential surface is discussed. The construction of the ring polymer representing tunneling in a system coupled to a heat bath at a certain temperature is then described. Finally, the calculation of the thermal tunneling rate is discussed.

A. Microcanonical tunneling path

In the JWKB approximation for a system with energy E, the tunneling rate is proportional to $1/(1+\exp 2\tilde{\theta})$ where $\tilde{\theta}$ is a line integral along a path $\mathbf{r}_p = \mathbf{r}(s)$ traced out by a parameter $s \in [0, 1]$

$$\tilde{\theta} = \tilde{\theta}[\mathbf{r}_p] = \frac{1}{\hbar}\int_{\mathbf{r}_1}^{\mathbf{r}_n} \sqrt{2\mu(V(\mathbf{r})-E)}\,dr \tag{9}$$

between the classical turning points $\mathbf{r}_1$ and $\mathbf{r}_n$. The turning points are located on opposite sides of the energy barrier such that $V(\mathbf{r}_1) = V(\mathbf{r}_n) = E$. Here, $\mathrm{dr} = |\mathbf{dr}|$ is the length of an infinitesimal displacement along the path. The maximum tunneling probability is obtained for the path that minimizes $\tilde{\theta}$, i.e., the OTP.

An efficient numerical method for finding the OTP will now be described. We assume that the gradient of the potential energy, ∇V, can be evaluated readily, but the second derivatives are not needed. The path will be represented by a discrete set of points ordered along the path $\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n\}$, and the integral will be approximated using an interpolation between the points. The simplest interpolation scheme will be used, namely, line segments between the points and a trapezoidal rule for evaluating the integral. Alternatively, one can use a cubic polynomial interpolation and Simpson's rule.

The task is then to find the vectors $\mathbf{r}_j$ that minimize the integral estimated from the discrete representation of the path

$$\tilde{\theta}[\mathbf{r}_p] \approx \theta(\mathbf{r}_1, \ldots \mathbf{r}_n) = \frac{1}{2\hbar}\sum_{j=1}^{n}\left(\sqrt{2\mu(V(\mathbf{r}_j)-E)} + \sqrt{2\mu(V(\mathbf{r}_{j-1})-E)}\right)\left|\mathbf{r}_j - \mathbf{r}_{j-1}\right|. \tag{10}$$

There are $n-1$ discretization points representing the path between the two endpoints, $\mathbf{r}_1$ and $\mathbf{r}_n$. The location of these discretization points needs to be optimized so as to minimize θ and the positions of the endpoints adjusted by moving them along the iso-contours corresponding to the energy E.

The optimization starts from some trial set of discretization points $\{\mathbf{r}_1^0, \mathbf{r}_2^0, \ldots, \mathbf{r}_n^0\}$ and involves iterative displacements of these points until $\theta(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n)$ cannot be reduced further. Let the negative gradient of θ with respect to the discretization point, $\mathbf{r}_j$, be denoted by

$$\mathbf{g}_j = -\nabla_j\theta = \frac{1}{2}\left(\frac{\mu}{\xi_j\hbar^2}(d_j + d_{j+1})\mathbf{f}_j - (\xi_j + \xi_{j-1})\widehat{\mathbf{d}}_j + (\xi_{j+1} + \xi_j)\widehat{\mathbf{d}}_{j+1}\right), \tag{11}$$

where $\mathbf{f}_j$ is the negative of the gradient of the energy, and for compactness of notation, the definitions

$$\xi_i \equiv \frac{1}{\hbar}\sqrt{2\mu(V(\mathbf{r}_i)-E)}, \tag{12}$$

$$\widehat{\mathbf{f}} = -\nabla V/|\nabla V|, \tag{13}$$

and

$$\widehat{\mathbf{d}}_j = (\mathbf{r}_j - \mathbf{r}_{j-1})/d_j \text{ with } d_j = |\mathbf{r}_j - \mathbf{r}_{j-1}| \tag{14}$$

have been used.

This represents the direction of steepest descent for each one of the discretization points located in between the endpoints (i.e., the interior points). It is used in an iterative optimization algorithm to find the set of configurations $\{\mathbf{r}_2, \mathbf{r}_3, \ldots, \mathbf{r}_{n-1}\}$ that minimize θ. But, only the component of $\mathbf{g}_j$ that is perpendicular to the path, $\mathbf{g}_j^{\perp}$, should be included in the optimization. The distribution of the discretization points along the path is controlled separately and should not be affected by the value of θ. This projection is analogous to what is used in the NEB method for finding MEPs and is referred to as "nudging."[10,36] The negative gradient, $\mathbf{g}_j$, is projected along the path

$$\mathbf{g}_j^{\parallel} = (\mathbf{g}_j \cdot \widehat{\tau}_j)\widehat{\tau}_j \tag{15}$$

and the rest of the vector constitutes the perpendicular component

$$\mathbf{g}_j^{\perp} = \mathbf{g}_j - \mathbf{g}_j^{\parallel}, \tag{16}$$

where τ_j is the tangent of the path at point $\mathbf{r}_j$. The tangent can be estimated from the line segment between j and the adjacent point with a larger value of V, either $\mathbf{r}_{j+1}$ or $\mathbf{r}_{j-1}$. This turns out to give a more stable numerical algorithm than the line segment between the two adjacent points $j+1$ and $j-1$ as has been found in NEB calculations.[46]

The discretization points can be distributed along the path in various ways, for example, by using a restraint method where a "spring" acting between adjacent discretization points is added. If the spring constant, k, is the same for all pairs of adjacent discretization points, then the points will be equally spaced along the path once convergence has been reached. If a different distribution is desired, the values of k for each adjacent pair of discretization points can be chosen accordingly and the equations can be modified in a straightforward way. Again, a projection is used to make sure that the spring forces do not affect the location and shape of the converged path. For the interior discretization points, $\{\mathbf{r}_2, \mathbf{r}_3, \ldots, \mathbf{r}_{n-1}\}$, the component of $\mathbf{g}_j$ parallel to the path is replaced by

$$\mathbf{g}_j^{sp} = k\left(\left|\mathbf{r}_{j+1} - \mathbf{r}_j\right| - \left|\mathbf{r}_j - \mathbf{r}_{j-1}\right|\right)\widehat{\tau}_j, \tag{17}$$

where k is analogous to a spring constant. A wide range of values can be chosen for k without affecting the results, but the convergence rate is in general faster if the $\mathbf{g}_j^{sp}$ are roughly of the same order of magnitude as $\mathbf{g}_j$. The total $\mathbf{g}$ that is used in the optimization is then given by the vector sum

$$\mathbf{g}_j^{\text{opt}} = \mathbf{g}_j^{\perp} + \mathbf{g}_j^{sp} \tag{18}$$

for $j = 2, \ldots, n-1$.

The steepest descent direction for the endpoints is defined differently since they should only move along the iso-contours corresponding to $V(\mathbf{r}) = E$. The component of $\mathbf{g}^{sp}$ parallel to the gradient of V needs to be zeroed so the endpoints only get displaced along the iso-contour. Furthermore, a harmonic restraint term, $\kappa(V_0 - E)^2$, is added to pull the endpoints towards the iso-contour if curvature results in a drift away from the iso-contour. The steepest descent direction for endpoint $\mathbf{r}_1$ is

$$\mathbf{g}_1^{\text{opt}} = \mathbf{g}_1^{sp} - \left(\mathbf{g}_1^{sp} \cdot \widehat{\mathbf{f}}(\mathbf{r}_1) - \kappa(V(\mathbf{r}_1) - E)\right)\widehat{\mathbf{f}}(\mathbf{r}_1), \tag{19}$$

where

$$\mathbf{g}_1^{\text{sp}} = k\,(\mathbf{r}_2 - \mathbf{r}_1). \tag{20}$$

An analogous expression gives the steepest descent direction for the other endpoint, $\mathbf{r}_n$,

$$\mathbf{g}_n^{\text{opt}} = \mathbf{g}_n^{sp} - \left(\mathbf{g}_n^{sp} \cdot \widehat{\mathbf{f}}(\mathbf{r}_n) - \kappa(V(\mathbf{r}_n) - E)\right)\widehat{\mathbf{f}}(\mathbf{r}_n), \tag{21}$$

where

$$\mathbf{g}_n^{\text{sp}} = k\,(\mathbf{r}_{n-1} - \mathbf{r}_n). \tag{22}$$

The parameter κ can be adjusted with respect to the spring constant, k, in order to maximize the convergence rate of the OTP. In general, these two constants should be chosen so that the corresponding terms are of similar order of magnitude to the remaining terms in the gradients. This can be done with a few test runs prior to the production calculations. The vectors $\mathbf{g}_1^{\text{opt}}$ and $\mathbf{g}_n^{\text{opt}}$ give the steepest descent directions for the two endpoints used in the iterative optimization while Eq. (18) applies to the interior discretization points.

In a steepest descent algorithm, all the discretization points $\mathbf{r}_j$ would be displaced in the direction of $\mathbf{g}_j^{\text{opt}}$ at each iteration. A more efficient and robust approach is based on a velocity projection modification of the velocity Verlet algorithm[43] which has been presented and used in the context of NEB calculations.[36] In LI-NEB calculations, the coupling between discretization points is stronger than that in NEB calculations and the calculation may require somewhat larger number of iterations to converge. Especially for the first LI-NEB calculation, where the trial path is based on two points along the MEP instead of on a previous OTP path. Moreover, there remains considerable room for improving the convergence rate, e.g., by the choice of the optimization technique[44] and the method used to discretize the JWKB integral.

When the LI-NEB method as well as the regular NEB method is applied to a finite system it becomes important to remove the translational and rotational degrees of freedom during the optimization of the path. This is accomplished by using the quaternion approach[47] to minimize sequentially the root-mean squared deviation of adjacent replica configurations, in each optimization step.

A more stable and faster way of reaching the preferred distribution of points along the path uses springs not between the adjacent points on the path but rather between the current position and an ideal position for each point based on the overall length of the path. This approach has been described and found to be useful in calculations of long and complex MEPs for dislocation nucleation.[45]

Prototypical LI-NEB calculations of OTPs are shown in Fig. 2 for a modified two-dimensional Müller-Brown surface. There, three Gaussians have been added to the regular Müller-Brown surface to shift and deepen one of the minima so as to make the arrangement of the three minima farther away from being linear. The initial path for the iterative LI-NEB calculation can be chosen quite arbitrarily; the main condition is that the two endpoints be placed on opposite sides of the

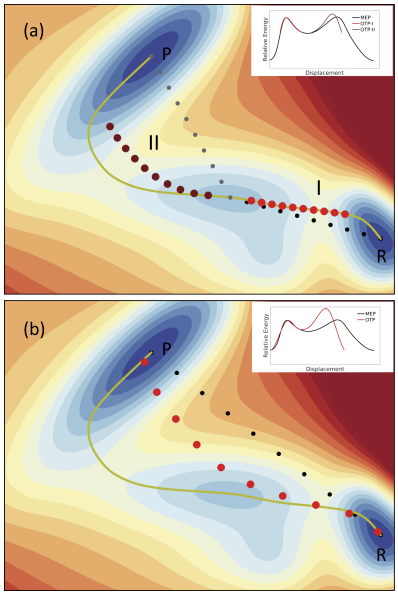


FIG. 2. Illustration of the path calculations for a two-dimensional modified Müller-Brown energy surface. The minimum energy path (green curve) between the initial and final states (R and P) goes through an intermediate minimum of energy E_m and two first-order saddle points of energy E_I and E_{II}. The initial paths were taken to be straight line interpolations between the minima (black disks). (a) Optimal tunneling paths for energy $E > E_m$. Tunneling can take place in two steps: first from R to the intermediate state (red disks) and then from the intermediate state to the final state P (brown disks). (b) Optimal tunneling path for energy $E < E_m$. Here, corner cutting can clearly be seen, as the optimal tunneling path does not go through the intermediate minimum and lies far from saddle point II. Insets: Energy along the minimum energy path and the optimal tunneling paths as a function of displacement.

energy barrier. Here, a straight line interpolation between the initial and final state minima is used. Alternatively, a much better choice for the initial path is to use the MEP or a part of it for the LI-NEB calculation. There are two different types of tunneling paths in this case depending on whether the energy is higher or lower than the energy of the intermediate minimum on the surface. For the higher energy range, the system tunnels from the initial state to the intermediate state. For the lower energy range, the tunneling occurs from the initial state to the final state directly. The endpoints are initially not at the right energy, but the iterative optimization of the path brings the endpoints to the contour corresponding to the total energy of the system, E, and then moves them along the energy contour V = E so as to minimize θ.

B. Classical periodic orbit

A classical trajectory from an initial position $\mathbf{r}_0$ at time $t = 0$ to a final position $\mathbf{r}_n$ at time $t = t_n$ is a path for which the action functional, S, defined as the integral of the Lagrangian $L(\mathbf{r}, \dot{\mathbf{r}}) = \frac{1}{2}\mu\dot{\mathbf{r}}^2 - V(\mathbf{r})$ over the time interval

$$S(\mathbf{r}, t) = \int_{t_0}^{t_n} L(\mathbf{r}, \dot{\mathbf{r}})dt \tag{23}$$

is stationary, i.e., $\delta S = 0$. The Lagrangian can be written in terms of the Hamiltonian, $H(\mathbf{r}, \mathbf{p})$, where $\mathbf{p}$ is the momentum $\mathbf{p} = \mu\dot{\mathbf{r}}$. The Lagrangian can be rewritten as $L(\mathbf{r}, \dot{\mathbf{r}}) = p|\dot{\mathbf{r}}| - H(\mathbf{r}, \mathbf{p})$ and the action can be written as

$$\underset{\substack{\text{Classical path}\\(\mathbf{r}_0,t_0)\to(\mathbf{r}_n,t_n)}}{S} = \int \left(p\frac{dr}{dt} - H\right) dt = \int_{\mathbf{r}_0}^{\mathbf{r}_n} p(\mathbf{r})dr - E(t_n - t_0), \tag{24}$$

where the conservation of total energy, $E = H(\mathbf{r}, \mathbf{p})$, along the classical path has been used in the last term, and the magnitude of the momentum, $p = |\mathbf{p}|$, can be obtained as a function of position along the path as

$$p(\mathbf{r}) = \sqrt{2\mu(E - V(\mathbf{r}))}. \tag{25}$$

A path that makes $\tilde{\theta}$ in Eq. (9) stationary is therefore a classical trajectory.

But, in the classically forbidden region where E < V(**r**), p(**r**) is imaginary. This can be interpreted as motion during an imaginary time interval. The replacement of t by it has the same effect as changing the sign of the potential energy V(**r**) in the classical equations of motion. The OTP, therefore, corresponds to a classical path on the inverted potential. Since the path lies between classical turning points, the velocity at the endpoints is zero and the path represents a closed classical trajectory, i.e., a periodic orbit.

C. Canonical tunneling path

A first-order saddle point (or an instanton), $\mathbf{R}_\ddagger$, on the effective quantum mechanical energy surface, V^{TP}, corresponds to stationary points of the Euclidean action, S_E, for the quantum system

$$S_E(\mathbf{R}) = \hbar V^{TP}(\mathbf{R})/k_B T, \tag{26}$$

i.e., it corresponds to $\delta S_E = 0$. This configuration of the ring polymer corresponds to an optimal tunneling path for a system coupled to a heat bath at a certain temperature. To evaluate the thermal rate constant, the corresponding temperature needs to be found and the replicas of the system placed in the right positions along the path to form the ring polymer. This can be accomplished by calculating the dynamical trajectory of the system along the one-dimensional tunneling path on the inverted potential, $-V(\mathbf{R})$. The trajectory is a periodic orbit with period $\tau = \hbar/k_B T$, where T is the temperature of the system.[23,24]

First, the period of the periodic orbit is found from the OTP. The force is evaluated from the inverted potential as $\mathbf{f}^p = +\nabla V$, and the magnitude of the component of $\mathbf{f}^p$ parallel to the path at a point $\mathbf{r}_j$ is evaluated as

$$f_j^{p\parallel} = \mathbf{f}_j^p \cdot \hat{\boldsymbol{\tau}}_j, \tag{27}$$

where $\hat{\boldsymbol{\tau}}_j$ is the unit tangent vector to the path. This is the force acting on the system along the one-dimensional tunneling path.

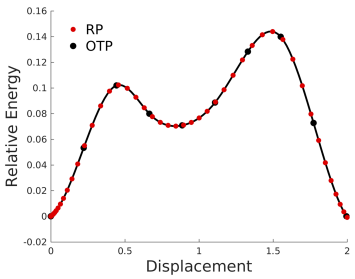


FIG. 3. Energy along the optimal tunneling path shown in Fig. 2(b) and the location of the n = 10 discretization points used in the LI-NEB calculation as well as 50 replicas in the corresponding saddle point configuration of the ring polymer (instanton). The figure illustrates the savings in computational effort by using the LI-NEB method rather than by optimizing the configuration of the ring polymer directly in a saddle point search. The latter involves more system replicas. Moreover, many of those replicas tend to cluster in regions of low energy, while the discretization points in the LI-NEB calculation are distributed evenly along the path.

A cubic interpolation of the coordinates $\mathbf{r}$ in the OTP as well as the parallel component of the force, $f_j^{\parallel}$, is used together with the velocity Verlet algorithm to carry out the dynamical simulation. Since the replicas double up along the path, it is enough to run the dynamics for only half an oscillation period. From the period of an oscillation, τ, the temperature corresponding to this tunneling path is found as $T = \hbar/k_B\tau$. The location of replica j is given by the coordinates of the system after time $t_j = j\tau/P$.

We find that to obtain a sufficient resolution of the OTP and hence a good estimate of the reactive flux, the LI-NEB calculation only requires 5-10 evenly distributed system replicas, depending on the system energy or temperature. Figure 3 shows a comparison of the discretization points in the OTP and the replicas in the saddle point configuration of the ring polymer calculated for the modified Muller-Brown potential surface. As can be seen there, more replicas are needed in the ring polymer than the discretization points in the OTP, and while the distribution of the latter can be controlled (chosen here to be equally spaced), the replicas of the ring polymer tend to cluster in regions of low energy.

D. The tunneling rate

A first-order saddle point on the V^{rp} energy surface can be used to estimate the thermal rate via the corresponding tunneling mechanism. The estimate of the rate constant (k^{HQTST}) (in the absence of rotational and translational degrees of freedom) is given by Eq. (6). Ω_j in the equation are the non-zero vibrational frequencies obtained from the eigenvalues of the Hessian matrix for the ring polymer consisting of P replicas of the system. In other words, the product sign in Eq. (6) only includes NP-1 modes because of the zero-mode, S_0, which cannot be treated with a quadratic approximation. The Hessian at each of the evenly distributed discretization points in the OTP, $H(\mathbf{r}_j), j = \{1, \ldots, n\}$, can be estimated by finite differences and each element in the matrix can then be interpolated with a cubic polynomial to estimate the Hessian matrix at the replicas in the ring polymer $H(\mathbf{r}_j^{rp}), j = \{1, \ldots, P\}$. The interpolated Hessian matrix can then be used to find the eigenvalues and the tunneling rate evaluated from the expression in Eq. (6). Generally, the number of replicas in the ring polymer is much larger than the number of replicas used to find the OTP ($P > n$). Therefore, by computing only the Hessian matrix of the OTP images in order to construct the Hessian matrix of the ring polymer, significant computational savings can be made, especially, when the methodology is used in conjunction with *ab initio* or DFT methods.

E. Performance

In the preliminary presentation of the LI-NEB approach to HQTST calculations, an application to the desorption of H_2 from a Cu(110) surface was given and the results were shown to agree closely with a full QTST calculation of the rate.[29] There, the energy of the system was evaluated from an empirical potential function. The same test problem has been used here to document the performance of the LI-NEB method and to make a comparison with the MMF method applied to HQTST.[28]

The simulated system consists of 216 Cu atoms in a 6-layer Cu(110) slab and two H atoms. The initial state is obtained by minimizing the energy with the two H atoms located in adjacent hollow sites. The reactive flux for associative desorption of H_2 from the surface is then evaluated for a range of energy values, corresponding to a temperature range of 260 K–50 K. Each LI-NEB calculation is started from an OTP obtained previously for a slightly higher energy, whereas the first calculation is started from a linear interpolation between the two points along an MEP (on opposite sides of the energy barrier) that are closest to the initial value of the energy. The LI-NEB calculations include only 10 system replicas in the OTPs and have a relatively loose convergence criterion of $|g_j^{\perp}| < 1 \cdot 10^{-2}$ Å^{-1} for the interior points ($j = 2, \ldots, n-1$). For the endpoints, $j = 1$ and $j = n$, the convergence criterion is $|g^{opt}| < 1 \cdot 10^{-2}$ Å^{-1}. The values of k and κ are chosen to be 20 Å^{-2} and 1000 eV^{-1}Å^{-1}, respectively.

For comparison, the MMF algorithm[33,34] was implemented using the Lanczos method for estimating the minimum mode.[48] This algorithm is used to generate the reference values for the reactive flux within a temperature range of 260 K–50 K. The reference calculations included 150 system replicas in the Feynman paths and a tight convergence criterion of $1 \cdot 10^{-8}$ eV/Å in the atomic forces.

The results obtained from the LI-NEB calculations of the OTP reproduced the reference values well for the entire temperature range. The natural logarithm of the flux was within 2% of the reference values, except at the lowest temperature of 50 K. In comparison, the MMF approach required 50 replicas in the Feynman paths and a convergence threshold of $F_{max} < 1 \cdot 10^{-3}$ eV/Å to achieve the same accuracy in the calculated flux. The optimization of the LI-NEB was carried out using velocity projection in the velocity Verlet algorithm.[36] Each step in the MMF optimization is taken in the direction of the conjugate search direction and the step size estimated according to a Newton line search. The number of energy and force evaluations required by the LI-NEB approach to compute the reactive flux was on average ca. 900 and it only

depended weakly on the temperature. For the MMF approach, between 2000 and 16 000 evaluations were needed to reach the same level of accuracy, with more evaluations required as the temperature was lowered.

These are, however, only preliminary tests and more calculations are needed to fully evaluate the efficiency of these methods. Neither of the two implementations used here have been optimized to their fullest extent. For example, improvements of the MMF method have been presented recently[49] but were not used here. In Sec. III, calculations of tunneling rates for two systems are presented where the energy and atomic forces were obtained from electronic structure calculations. There, an additional issue arises due to the error in the estimates of the atomic forces, as compared to the H_2/Cu(110) test problem discussed above where a potential function was used. We expect the LI-NEB method to be less sensitive to errors in the atomic forces, but this remains to be tested.

III. APPLICATIONS

A. Dissociation of H_3BNH_3

There has been significant interest in ammoniaborane, H_3BNH_3, for solid-state hydrogen storage, as the compound contains 20% hydrogen by weight. Ammoniaborane is a structural analog of ethane with C atoms replaced by B and N while being a solid at room temperature. When heated, it decomposes to yield H_2 and a solid residue of polymeric aminoborane, $[H_2BNH_2]_x$. Figure 4 shows the structure of the gas phase ammoniaborane molecule and the saddle point configuration for the decomposition into H_2 and H_2BNH_2. Upon further heating, $[H_2BNH_2]_x$ can yield additional H_2, but the thermal decomposition process of the solid is complex and one of the problems with the practical use of the compound for hydrogen storage is the release of $B_3N_3H_6$ (a BN-analog of benzene) and B_2H_6.[50]

Calculations of the decomposition of a gas phase ammoniaborane molecule have been carried out by Li *et al.*[51] using DFT with the B3LYP functional[52–54] and G3MP2B3.[55] In addition to the classical HTST rate constant, they reported an estimate of the effect of tunneling using the small curvature tunneling (SCT) correction.[56] Interestingly, the effect of tunneling is estimated to be exceptionally large for this transition. For a temperature of 200 K, for example, Li *et al.* reported an SCT estimate of the tunneling rate that is 6 orders of magnitude larger than the classical HTST rate.

We have carried out calculations of the gas phase decomposition of H_3BNH_3 to compare various levels of rate theory. The electronic structure calculations were performed at the B3LYP/def2-SV(P)-gcp[57–59] level of theory, as implemented in the ORCA software.[60] The minimum energy of the H_3BNH_3 molecule is obtained with a staggered configuration of C_{3v} symmetry (see the inset of Fig. 4). At the classical first-order saddle point configuration for the decomposition into H_2 and H_2NBH_2, the NH_3 group has rotated into an eclipsed configuration and one of the N–H bonds has stretched to bring an H atom towards one of the H atoms bonded to the B atom. The remaining H atoms bonded to the N atom come close to the planar configuration of the product state, H_2BNH_2. The activation energy estimated from the first-order saddle point with zero-point energy correction is 1.50 eV which is in good agreement with the calculations of Li *et al.* who reported 1.41 eV and 1.49 eV using G3MP2B3 and B3LYP/aug-cc-pVTZ, respectively. Also, the calculated HTST rate constant agrees well with what Li *et al.* reported.[51]

The LI-NEB calculations started with an energy of 0.05 eV below that of the first-order saddle point on the energy surface and an initial path lying along the MEP. The energy was subsequently reduced stepwise covering an interval of roughly 1 eV. The calculations employed 10 replicas along the OTP path. The number of replicas in the instanton was P = 240, and the NP-dimensional Hessian matrix of the instanton was obtained by cubic polynomial interpolation of the Hessian evaluated at the 10 replicas of the OTP path. The parameters k and κ were chosen to have a value of 20 $Å^{-2}$ and 100 $eV^{-1}Å^{-1}$, respectively. These values were found to lead to good convergence behavior. The same convergence criterion was used as for the H_2/Cu(110) problem discussed earlier.

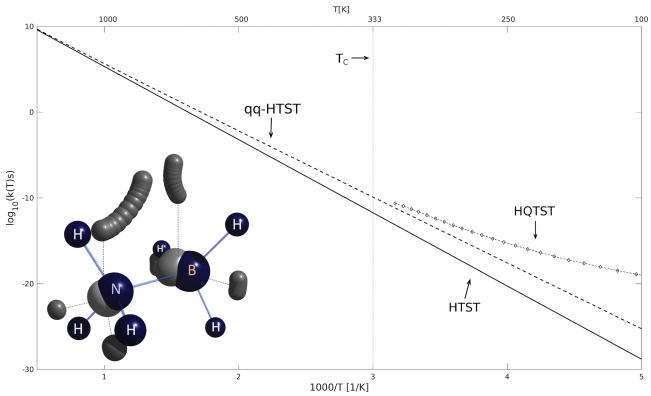


FIG. 4. Reaction rate as a function of inverse temperature for H_2 abstraction from a H_3BNH_3 molecule. Nuclear quantum effects are found to dominate the reaction rate even above room temperature. The crossover temperature is calculated to be T_c = 333 K, but the classical HTST approximation underestimates the reaction rate even at a higher temperature. The extension of HTST to include quantum mechanical vibrational partition functions, the qq-HTST approximation, takes into account zero point delocalization but not tunneling. Inset: Initial state (dark blue spheres and light blue bonds) and optimal tunneling path (gray) for 200 K.

The results of calculations carried out at various levels of rate theory are shown in Fig. 4. The negative eigenvalue of the Hessian at the classical first-order saddle point gives T_c = 330 K using Eq. (1). This high rate of tunneling compared with the classical over-the-barrier rate at such a high temperature is unusual. The reason for this is the narrowness of the energy barrier. The two H atoms that form the H_2 molecule can come close to each other because one of them is slightly positively charged and the other is slightly negatively charged, as one is bound to a B atom and the other is bound to an N atom.

The HQTST results give a tunneling rate that is nearly 10 orders of magnitude larger than the HTST rate at 200 K. This is a significantly larger difference than the one obtained from the SCT approximation.[51] The SCT approximation[56] is based on a harmonic expansion around the classical MEP, while the HQTST approximation is based on a harmonic expansion around the OTP. The SCT is, therefore, not as accurate as HQTST when the temperature is significantly below T_c. The OTP can be in a quite different region of the energy surface than the MEP, as illustrated in Fig. 2.

B. H atom diffusion in Ta crystal

Several studies of hydrogen diffusion in metals have indicated a crossover in the diffusion mechanism at low temperature, typically below room temperature.[61] This is usually interpreted as the onset of tunneling of the H atoms. Diffusion in BCC crystals shows stronger signs of tunneling than diffusion in FCC crystals as BCC metals have interstitial sites that lie closer together than in FCC metals. One of the clearest crossover has been observed for H diffusion in the dilute α-phase of hydrogen in a Ta crystal.

Results of two experimental measurements of hydrogen diffusion in Ta are shown in Fig. 5. In one of the experiments, the macroscopic diffusivity was measured using the Grosky effect.[62] In the other experiment, NMR spin-lattice relaxation

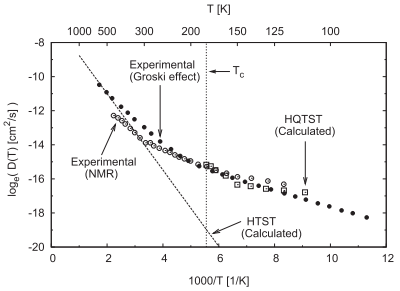


FIG. 5. Diffusion constant for hydrogen in a Ta crystal. The values calculated using the HQTST approximation (open squares) where the energy and atomic forces are obtained from DFT calculations are in close agreement with the values obtained from the Groski effect[62] (filled disks) and from NMR[63] (open circles) experiments. The crossover temperature is calculated to be T_c = 180 K, but the classical HTST approximation underestimates the reaction rate even at a significantly higher temperature. The agreement between the HQTST calculation and the experiments is remarkably good considering that no parameter in the calculation has been adjusted to reproduce the experimental data.

times were used to estimate the rate of transitions between sites in the lattice.[63] In the Grosky effect measurements, a concentration gradient is introduced by bending the sample. The hydrogen atoms migrate along the gradient causing a time-dependent anelastic strain. From the relaxation time of the anelastic strain, the diffusion constant can be obtained. Since the anelastic strain is macroscopic in nature, the diffusion constant obtained from the Grosky effect can be influenced by impurities and defects as hydrogen atoms can become trapped there, especially at low temperature. This will result in an underestimate of the diffusion rate in a perfect crystal. On the other hand, the NMR spin-lattice relaxation time is comparable to that of the elementary transitions and the diffusivity is estimated by assuming only jumps between two neighboring tetrahedral sites, which could lead to an underestimate at high temperature. As can be seen from Fig. 5, the Grosky effect measurements are found to give a lower diffusion rate than the NMR measurements at low temperature, while the opposite trend is observed at high temperature. Messer *et al.*[63] argue that the overall best estimate for H diffusion in a perfect Ta crystal is obtained from the Grosky effect measurements in the high-temperature range and NMR measurements in the low-temperature range.

Sundell and Wahnström carried out calculations of H atom hopping between the stable sites, T_d sites, in Ta.[64] The energy and atomic forces were evaluated using DFT. The H atom was found to self-trap in a T_d site with trapping energy of 0.3 eV. By constructing a three-dimensional grid to describe the position of the hydrogen atom in the T_d site as well as a hydrogen atom delocalized over two neighboring T_d sites, a three-dimensional Schrödinger equation was solved and the vibrational levels along with tunneling matrix elements[65] were estimated. The delocalization of the hydrogen atom was simulated by placing hydrogen atoms at the two sites and using a linear combination of the symmetrized Hellmann-Feynman forces.

By using the high-temperature limit of small polaron theory,[66] Sundell and Whanström calculated the rate of H atom jumps between sites over a temperature range of 100 K–200 K and obtained results in good agreement with the NMR measurements. The small polaron approximation is valid if the temperature is low enough for excited vibrational states to be ignored, but high enough for the lattice vibrations to be treated classically.

We carried out HQTST calculations of H atom diffusion in Ta where the energy of the system and the atomic forces were obtained from DFT calculations. The calculations made use of the PW91 functional,[67] a plane wave basis set, and periodic boundary conditions applied to a cubic cell of 16 Ta atoms and one H atom. The kinetic energy cutoff in the plane wave basis was 250 eV (18.4 Ry). The inner electrons were described by the projector augmented wave formalism.[68] The Brillouin zone was sampled by a 3 × 3 × 3 k-point mesh generated with the Monkhorst-Pack scheme.[69] After volume relaxation, the lattice constant was determined to be 3.32 Å, in excellent agreement with the experimental value of 3.31 Å.[70] The VASP software was used for these calculations.[71]

The HQTST method was used to calculate the rate of transitions between adjacent T_d sites. First, the NEB method

was used to find the MEP. From the maximum along the MEP, an activation energy of 0.20 eV was obtained. When all but the five Ta atoms surrounding the two T_d sites visited by the H atom were kept fixed, the activation energy only increased by 0.004 eV. From the unstable mode at the first-order saddle point, an estimate of T_c = 180 K was obtained.

The diffusion constant was estimated from the HQTST rate constant, $k(T)$, for transitions between adjacent T_d sites by assuming a random walk through the lattice. The expression for the diffusion constant, $D(T)$, is

$$D(T) = n\frac{l^2}{2d}k(T) \tag{28}$$

$$= 9.13 \cdot 10^{-17}k(T)\ (\mathrm{cm}^2/\mathrm{s}), \tag{29}$$

where l is the distance between two T_d sites (1.17 Å), $d = 3$ is the dimensionality of the lattice, and n is the number of symmetrically equivalent diffusion paths (4 in the present case).

The calculated diffusion constant is compared to the two sets of experimental data in Fig. 5. The agreement is remarkably good considering that no parameter has been fitted to reproduce the experimental data. The nuclear quantum effects are large even well above the harmonic estimate of T_c given by Eq. (1) as is apparent from a comparison of the HQTST and HTST results. The experimentally observed crossover in the diffusivity as a function of temperature can clearly be assigned to a shift from classical over-the-barrier hops of the H atoms between T_d sites at high temperature to thermally activated quantum mechanical tunneling between the sites at low temperature.

IV. DISCUSSION

The number of energy and atomic force evaluations needed in the application of the methodology described here for calculating OTPs and HQTST tunneling rates is of similar order of magnitude as NEB and HTST calculations of classical transition rates, except that the OTP needs to be updated as the energy and, thereby, the temperature is changed. The calculations can, therefore, be combined with electronic structure evaluations of the energy and atomic forces, such as *ab initio* or DFT, as is frequently done today in calculations of classical transition rates. The key aspect of the calculation is the controlled distribution of discretization points in the iterative optimization of the tunneling path. Preliminary results from calculations of the associative desorption of H_2 from the Cu(110) surface suggest that an optimization of the ring polymer configuration using a saddle point search method, such as the MMF method, is less efficient than the LI-NEB method for finding OTPs since it needs to include many more system replicas and most of those are redundant, clustering in the lower energy regions of the energy surface.

As has recently been demonstrated in the NEB calculations, the number of electronic structure calculations of the energy and the atomic forces can be reduced significantly by using GPR, a type of machine learning.[40,41] There, the results of the electronic structure calculations are used to create an approximate energy surface that can be evaluated without large computational effort. This can reduce the number of electronic structure calculations needed to converge on the OTP. Also, after finding the OTP, the approximate energy surface can be used to calculate more accurately the dynamics along the path to determine the period of the classical trajectory, the location of the replicas in the ring polymer, and the corresponding Hessian matrix. GPR also provides an estimate of the uncertainty in the approximate energy surface. If a calculation of an OTP or a classical trajectory on the approximate surface ventures into an area where the uncertainty is large, then an additional electronic structure calculation can be carried out at the most relevant point to reduce the uncertainty. In the present context, an OTP calculation would start with an MEP and an approximate energy surface giving an accurate representation of the true energy surface in the vicinity of the MEP. An LI-NEB calculation for an OTP for energy slightly below the saddle point energy can then be carried out on the approximate surface as a first approximation, and electronic structure calculations can then be carried out for the most relevant points on the converged OTP. The LI-NEB calculation is then repeated for the updated approximation of the energy surface, etc. Once the OTP has converged in this way and the approximate energy surface verified by electronic structure calculations at the discretization points, the classical dynamics simulations for determining the corresponding temperature can be carried out on the approximate energy surface. Again, if the dynamics bring the system to a point where the approximate energy surface has too large uncertainty, additional electronic structure calculations can be carried out at selected points.

Since the optimization of the paths is carried out by adjusting the location of each one of the discretization points simultaneously, the LI-NEB algorithm for path optimization lends itself well to parallel computing especially when combined with electronic structure calculations. Then, the evaluation of the energy and atomic forces becomes by far the largest computational effort and the overhead for parallel or even distributed computing becomes insignificant.

The HQTST approximation presents a significant improvement over frequently used approximations of tunneling rates, such as various one-dimensional approximations and the SCT approximation. The reason for this is that an optimal tunneling path is determined instead of using an expansion around the classical MEP. Since the computational effort in the LI-NEB calculation is of the same order of magnitude as a regular NEB calculation, especially when GPR is used, we foresee that the methodology presented here will become a method of choice for estimating tunneling rates from electronic structure calculations of, for example, chemical reactions and diffusion events.

In most systems, tunneling of atoms is only important below room temperature. But, there are exceptions as has been demonstrated here in the case of ammoniaborane decomposition. The reason tunneling is so fast in that case compared with the classical over-the-barrier mechanism is the narrowness of the energy barrier which results from partial positive and negative charges on the two H atoms that form the hydrogen molecule. There is no doubt whether other such examples and systematic calculations of tunneling rates could help

establish guidelines for identifying systems where tunneling plays an important role at a temperature of interest. The largest application area for atom tunneling calculations is no doubt astrochemistry where the relevant temperature is typically on the order of 10-50 K.[32,72,73]

The harmonic approximation for calculating classical transition rates, HTST, has been found to work well for a wide range of systems, especially transitions in solids and on the surfaces of solids. While methods that allow for full free energy calculations as the transition state dividing surface is optimized both with respect to location and orientation have been developed,[74,75] most calculations by far are carried out within the HTST approximation. Similarly, it is likely that the harmonic approximation for calculating the rate of tunneling is accurate enough in many cases, especially in light of the inherent errors in DFT calculations of energy barriers. The accuracy of HQTST has been demonstrated already for some systems by comparison with more elaborate quantum rate theories.[28,29] The close agreement obtained here between with experimental measurements of hydrogen diffusion in a Ta crystal also demonstrate the accuracy of the HQTST approach. We note that the calculations assume adiabaticity between the nuclear and electronic degrees of freedom, the Born-Oppenheimer approximation, and ignore possible electron-hole pair excitations during the diffusion events. While electron-hole pairs can play a role in atomic rearrangements in metals and on metal surfaces, those cases seem to be exceptions from the general applicability of the Born-Oppenheimer approximation in thermally activated transitions.

The extension of HTST to transitions between magnetic states has been developed[76,77] as well as a version of the NEB method tailored to magnetic systems.[78] But, as for atomic transitions, tunneling of magnetic moments rather than over-the-barrier hops becomes the dominant transition mechanism at low enough temperature. A method for estimating the crossover temperature for tunneling in magnetic systems has been presented,[79,80] and a method for finding magnetic instantons has been developed,[81] but an expression for the HQTST pre-exponential factor or thermally activated tunneling rates in magnetic systems remains to be formulated.

Finally, we point out that the crossover from over-the-barrier mechanism to tunneling mechanism as temperature is lowered for ring polymers representing a quantum mechanical system is analogous to the crossover from coiled to stretched transition of a molecular polymer as a function of length.[82,83] Also there, a harmonic transition state theory has been found to give accurate results for the transition rate, and the results of such calculations have been further improved by applying dynamical corrections.[84]

ACKNOWLEDGMENTS

We thank Judith Rommel and Johannes Kästner for helpful discussions. This work was supported by the Icelandic Research Fund, by the Center for Nonlinear Studies at the Los Alamos National Laboratory, and by a Doctoral Grant from the University of Iceland Research Fund. The calculations were carried out at the University of Iceland computing center (Reiknistofnun).

[1] M. J. Gillan, J. Phys. C: Solid State Phys. **20**, 3621 (1987).
[2] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[3] J. A. Barker, J. Chem. Phys. **70**, 2914 (1979).
[4] D. Chandler and P. G. Wolynes, J. Chem. Phys. **74**, 4078 (1981).
[5] E. Wigner, Trans. Faraday Soc. **34**, 29 (1938).
[6] M. J. Gillan, Phys. Rev. Lett. **58**, 563 (1987).
[7] G. A. Voth, D. Chandler, and W. H. Miller, J. Chem. Phys. **91**, 7749 (1989).
[8] G. Mills and H. Jónsson, Phys. Rev. Lett. **72**, 1124 (1994).
[9] G. K. Schenter, G. Mills, and H. Jónsson, J. Chem. Phys. **101**, 8964 (1994).
[10] G. Mills, H. Jónsson, and G. K. Schenter, Surf. Sci. **324**, 305 (1995).
[11] G. Mills, G. K. Schenter, D. E. Makarov, and H. Jónsson, Chem. Phys. Lett. **278**, 91 (1997).
[12] G. Mills, G. K. Schenter, D. E. Makarov, and H. Jónsson, "RAW quantum transition state theory," in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne, G. Ciccotti, and D. F. Coker (World Scientific, 1998), pp. 405–421.
[13] V. A. Benderskii, D. E. Makarov, and C. A. Wight, Adv. Chem. Phys. **88**, 1 (1994).
[14] I. R. Craig and D. E. Manolopoulos, J. Chem. Phys. **123**, 034102 (2005).
[15] S. Habershon, D. E. Manolopoulos, T. E. Markland, and T. F. Miller III, Annu. Rev. Phys. Chem. **64**, 387 (2013).
[16] J. O. Richardson and S. C. Althorpe, J. Chem. Phys. **131**, 214106 (2009).
[17] T. J. H. Hele and S. C. Althorpe, J. Chem. Phys. **138**, 084108 (2013).
[18] H. Jónsson, Proc. Natl. Acad. Sci. U. S. A. **108**, 944 (2011).
[19] G. H. Vineyard, J. Phys. Chem. Solids **3**, 121 (1957).
[20] G. Henkelman, A. Arnaldsson, and H. Jónsson, J. Chem. Phys. **124**, 044706 (2006).
[21] J. S. Langer, Ann. Phys. **54**, 258 (1969).
[22] W. H. Miller, J. Chem. Phys. **62**, 1899 (1975).
[23] S. Coleman, Phys. Rev. D **15**, 2929 (1977).
[24] C. G. Callan, Jr. and S. Coleman, Phys. Rev. D **16**, 1762 (1977).
[25] M. Messina, G. K. Schenter, and B. C. Garrett, J. Chem. Phys. **103**, 3430 (1995).
[26] J. O. Richardson, J. Chem. Phys. **144**, 114106 (2016).
[27] A. N. Beyer, J. O. Richardson, P. J. Knowles, J. Rommel, and S. C. Althorpe, J. Phys. Chem. Lett. **7**, 4374–4379 (2016).
[28] S. Andersson, G. Nyman, A. Arnaldsson, U. Manthe, and H. Jónsson, J. Phys. Chem. A **113**, 4468 (2009).
[29] D. M. Einarsdóttir, A. Arnaldsson, F. Óskarsson, and H. Jónsson, Lect.Notes Comput.Sci. **7134**, 45 (2012).
[30] J. B. Rommel and J. Kästner, J. Chem. Phys. **134**, 184107 (2011).
[31] J. B. Rommel, Y. Liu, H.-J. Werner, and J. Kästner, J. Phys. Chem. B **116**, 13682 (2012).
[32] J. Meisner and J. Kästner, Angew. Chem., Int. Ed. **55**, 5400 (2016).
[33] G. Henkelman and H. Jónsson, J. Chem. Phys. **111**, 7010 (1999).
[34] R. A. Olsen, G. J. Kroes, G. Henkelman, A. Arnaldsson, and H. Jónsson, J. Chem. Phys. **121**, 9776 (2004).
[35] E. Merzbacher, *Quantum Mechanics*, 3rd ed. (John Wiley and Sons, 1961).
[36] H. Jónsson, G. Mills, and K. W. Jacobsen, "Nudged elastic band method for finding minimum energy paths of transitions," in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne, G. Ciccotti, and D. F. Coker (World Scientific, 1998), pp. 385–404.
[37] P. Faccioli, M. Sega, F. Pederiva, and H. Orland, Phys. Rev. Lett. **97**, 108101 (2006).
[38] J. O. Richardson, J. Chem. Phys. **143**, 134116 (2015).
[39] M. T. Cvitas and S. C. Althorpe, J. Chem. Theory Comput. **12**, 787 (2016).
[40] O.-P. Koistinen, E. Maras, A. Vehtari, and H. Jónsson, Nanosyst.: Phys., Chem., Math. **7**, 925 (2016).
[41] O.-P. Koistinen, F. B. Dagbjartsdóttir, V. Ásgeirsson, A. Vehtari, and H. Jónsson, J. Chem. Phys. **147**, 152720 (2017).
[42] K. Müller and L. D. Brown, Theor. Chim. Acta **53**, 75 (1979).
[43] H. C. Andersen, J. Chem. Phys. **72**, 2384 (1980).
[44] M. U. Bohner, J. Meisner, and J. Kästner, J. Chem. Theory Comput. **9**, 3498–3504 (2013).
[45] E. Maras, L. Pizzagalli, T. Ala-Nissila, and H. Jónsson, Sci. Rep. **7**, 11966 (2017).
[46] G. Henkelman and H. Jónsson, J. Chem. Phys. **113**, 9978 (2000).
[47] M. Melander, K. Laasonen, and H. Jónsson, J. Chem. Theory Comput. **11**, 1055 (2015).
[48] C. Lanczos, *Applied Analysis* (Dover, New York, 1988).
[49] M. P. Gutiérrez, C. Argáez, and H. Jónsson, J. Chem. Theory Comput. **13**, 125 (2017).

[50]F. Baitalow, J. Baumann, G. Wolf, K. Jaenicke-Roessler, and G. Leitner, Thermochim. Acta **391**, 159 (2002).
[51]Q. S. Li, J. Zhang, and S. Zhang, Chem. Phys. Lett. **404**, 100 (2005).
[52]P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, J. Phys. Chem. **98**, 11623 (1994).
[53]A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[54]C. T. Lee, W. T. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[55]A. G. Baboul, L. A. Curtiss, P. C. Refern, and K. Raghavachari, J. Chem. Phys. **110**, 7650 (1999).
[56]R. T. Skodje and D. G. Truhlar, J. Chem. Phys. **77**, 5955 (1982).
[57]F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).
[58]A. Schäfer, H. Horn, and R. Ahlrichs, J. Chem. Phys. **97**, 2571 (1992).
[59]H. Kruse and S. Grimme, J. Chem. Phys. **136**(15), 154101 (2012).
[60]F. Neese, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**(1), 73–78 (2012).
[61]R. G. Barnes, P. Dantzer, H. Grabert, D. K. Ross, H. R. S. H. Vehoff, and H. Wipf, in *Hydrogen in Metals III: Properties and Applications*, Volume 73 of Topics in Applied Physics, edited by H. Wipf (Springer-Verlag, Germany, 1997).
[62]Z. Qi, J. Völkl, R. Lasser, and H. Wenzl, J. Phys. F: Met. Phys. **13**, 2053 (1983).
[63]R. Messer, A. Blessing, S. Dais, D. Höpfel, G. Majer, C. Schmidt, A. Seeger, and W. Zag, Z. Phys. Chem., Neue Folge **2**, 61 (1986).
[64]P. G. Sundell and G. Wahnström, Phys. Rev. Lett. **92**, 155901 (2004); Phys. Rev. B **70**, 224301 (2004).
[65]W. Kehr, in *Hydrogen in Metals I: Basic Properties*, Volume 28 of Topics in Applied Physics, 1st ed., edited by G. Alefeld and J. Völkl (Springer-Verlag, Germany, 1978), Chap. 8, pp. 202–204.
[66]C. P. Flynn and A. M. Stoneham, Phys. Rev. B **1**, 3966 (1970).
[67]J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992).
[68]P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[69]H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).
[70]*Handbook of Chemistry and Physics*, edited by D. R. Lide (CRC, Boca Raton, New York, 1997).
[71]G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996); Phys. Rev. B **54**, 11169 (1996).
[72]V. Ásgeirsson, H. Jónsson, and K. T. Wikfeldt, J. Chem. Phys. **121**, 1648 (2017).
[73]T. P. M. Goumans and J. Kästner, Angew. Chem., Int. Ed. **49**, 7350 (2010).
[74]G. H. Jóhannesson and H. Jónsson, J. Chem. Phys. **115**, 9644 (2001).
[75]T. Bligaard and H. Jónsson, Comput. Phys. Commun. **169**, 284 (2005).
[76]P. F. Bessarab, V. M. Uzdin, and H. Jónsson, Phys. Rev. B **85**, 184409 (2012).
[77]P. F. Bessarab, V. M. Uzdin, and H. Jónsson, Z. Phys. Chem. **227**, 1543 (2013).
[78]P. F. Bessarab, V. M. Uzdin, and H. Jónsson, Comput. Phys. Commun. **196**, 335 (2015).
[79]S. M. Vlasov, P. F. Bessarab, V. M. Uzdin, and H. Jónsson, Faraday Discuss. R. Soc. **195**, 93 (2016).
[80]S. M. Vlasov, P. F. Bessarab, V. M. Uzdin, and H. Jónsson, Nanosyst.: Phys., Chem., Math. **8**, 454 (2017).
[81]S. M. Vlasov, P. F. Bessarab, V. M. Uzdin, and H. Jónsson, Nanosyst.: Phys., Chem., Math. **8**, 746 (2017).
[82]K. L. Sebastian and A. Debnath, J. Phys.: Condens. Matter **18**, S283 (2006).
[83]H. Mökkönen, T. Ikonen, T. Ala-Nissilä, and H. Jónsson, J. Chem. Phys. **142**, 224906 (2015).
[84]H. Mökkönen, T. Ala-Nissilä, and H. Jónsson, J. Chem. Phys. **145**, 094901 (2016).

Article II

Nudged elastic band calculations with focus on the region around the highest saddle point

Ásgeirsson, V., Birgirson, B.O., and Jónsson, H.

To be submitted to Journal of Chemical Theory and Computation, 2021

Nudged Elastic Band Calculations with Focus on the Region Around the Highest Saddle Point

Vilhjálmur Ásgeirsson, Benedikt O. Birgisson, and Hannes Jónsson*

Science Institute and Faculty of Physical Sciences, University of Iceland VR-III, 107 Reykjavík, Iceland

E-mail: hj@hi.is

Abstract

A method is presented for focusing the computational effort of a climbing image nudged elastic band (CI-NEB) calculation of a minimum energy path to the region of the energy maximum. This improves the resolution of the path in the most critical region and thereby provides a better estimate of the climb direction without introducing an excessive number of images. This zooming in on the top of the energy barrier takes place after a CI-NEB calculation has reached a certain level of convergence and the parallel computational resources are then shifted automatically to a denser grid of images near the maximum. The two endpoints in this zoom CI-NEB (ZCI-NEB) calculation of a shorter segment of the path are not fixed but move in the direction of the atomic force perpendicular to the current path and thereby converge on the minimum energy path along with the other images. The method is applied to a benchmark test set involving structural rearrangements of a heptamer island on a solid surface and it is found to be robust and efficient. The Z-NEB method is then applied to two molecular reactions that have long and complex paths: The hydrolysis of ethyl acetate and a rearrangement of 1,5-hexadiene. The saddle points for these reactions are successfully located using significantly fewer images than is required for a standard CI-NEB calculation to converge, resulting in considerable reduction of computational effort.

1 Introduction

The minimum energy path (MEP) on an energy surface characterizing a rearrangement of atoms from one configuration to another, such as a chemical reaction, a diffusion event or a conformational change of a cluster, provides important information about the mechanism and possible intermediate states visited during the transition. Calculations of MEPs can be carried out using the nudged elastic band (NEB) method[1,2] where the path is discretized with a set of replicas of the system, referred to as images, and an optimization algorithm based on the atomic forces is used to shift the images iteratively to the MEP. At the MEP, perpendicular components of the force on the atoms vanishes. For a recent review of the method and its implementations, see Ref.[3] The energy maximum along the MEP is the most important along the path and the calculation needs to be converged particularly well in that region. A maximum along the MEP represents a first order saddle point on the energy surface and the highest maximum is the activation energy of the transition within the harmonic approximation of transition state theory.[4,5]

In the climbing image NEB (CI-NEB) method,[6] one of the images is driven to the point of maximum energy, but its success relies on having sufficient resolution of the path near the maximum in order to determine the right direction for the climb. If the whole path is represented at such high resolution, the computational effort can become excessive as the number of images becomes large. With too low resolution of the path near the climbing image, the estimate of the tangent to the path, which determines the climbing direction, can oscillate back and forth and prevent the calculation from reaching convergence.

After a CI-NEB calculation reaches a certain loose level of convergence, the overall shape of the MEP, such as the possible presence of intermediate local minima and the relative height of maxima can become evident, well before the point of maximum energy has been found sufficiently well to provide a reliable estimate of the activation energy. Further calculations of images located away from the region of the maximum are then unnecessary and the computational resources are better used to focus on the images near the maximum and introduce additional images there to improve the resolution where it is most needed. Since one of the advantages of the NEB method is efficient use of parallel computing in simultaneous calculations of the energy and atomic forces of each image, the computational resources assigned to the computation can be used in a more optimal way by shifting processors to additional images near the maximum at a point when the overall shape of the MEP has been established in the CI-NEB calculation.

The importance of representing the path with higher resolution near the energy maximum becomes particularly clear in reactions of molecules where a significant portion of the MEP

may involve rotations of molecular fragments that involve little change in energy.The energy curve representing the energy as a function of atomic displacements along the MEP can then have extended regions that are nearly flat while the energy barrier represents only a small segment of the path. This problem is less pronounced in calculations of MEPs for solid state transitions, where the CI-NEB method has mainly been used so far.

There have been several formulations of NEB calculations where focus is shifted to the barrier region after loose convergence to the MEP. In some cases, two of the intermediate images on opposite sides of the current maximum energy image have been chosen and the atomic coordinates of those images fixed while additional movable images are introduced in the intermediate region by interpolation. Then, a new NEB optimization is initiated. This procedure is then carried out repeatedly on an even interval of optimization steps (see for example[10]). This, however, can lead to problems, as illustrated below, since the fixed endpoint images are in such cases not sitting on the MEP. A better approach is to allow the new endpoint images to move to the MEP along with the intermediate images.[11] Algorithms for moving endpoint images have been presented in the context of diffusion in solids[13] and calculations of optimal tunneling paths.[14] An elaborate scheme for focusing CI-NEB calculation on the barrier region, the autoNEB method, has recently been presented and used in various contexts.[12] In this method, new images are added sequentially into the largest energy or geometrical gap along the path during the NEB optimization.

In the present article, a simple scheme for focusing a CI-NEB calculation on a region around the energy maximum along an MEP is presented. After the CI-NEB calculation has been carried out until a loose tolerance in the magnitude of the force acting on the climbing image is met, the method zooms in on the region near the energy maximum. It is referred to as zoom-NEB (Z-NEB) and it is tested using a standard benchmark involving rearrangements of a heptamer island on a solid surface.[15] Z-NEB is found to be robust and outperform CI-NEB with even distribution of images as well as autoNEB. An application to two challenging molecular reactions is described where density functional theory is used to evaluate the energy and atomic forces, namely the hydrolysis of ethyl acetate and a rearrangement reaction of 1,5-hexadiene. Again, the Z-NEB method is found to outperform CI-NEB and lead to considerable savings in computational effort.

The article is organized as follows: In section 2, the Z-NEB method is described and two-dimensional test problems used to demonstrate the concepts and implementation aspects. In section 3, the benchmark testing and comparison with autoNEB for the heptamer island is presented. Moreover, an application of the Z-NEB method to the two molecular reactions is described and the performance analyzed. The MEPs of these reactions are characterized by extended tails where the energy hardly changes and the standard CI-NEB method requires

large number of images in order to reach convergence. The article then concludes with a discussion of the results in section 4.

2 Methodology

The Z-NEB method is an automatic two phase CI-NEB procedure that can be used to accurately locate the point of highest energy along an MEP connecting a given reactant and product state, with fewer images than a standard CI-NEB requires for convergence. This is especially true for cases where extended regions of the MEP involve little change in the energy. In the first phase of Z-NEB, the objective is to obtain a path that approximates the MEP well enough to identify the region around the the point of highest energy on the MEP. In the second phase, referred to as ZCI-NEB, the images outside this region are discarded and additional images are inserted by interpolation in the region near the maximum in such a way that the total number of active images remains the same and the computational resources (cores) assigned to the calculation continue to be used efficiently. In Fig. 1 the Z-NEB method is illustrated in calculation of the MEP on the two-dimensional Müller-Brown surface[16] and comparison is made with the standard CI-NEB method.

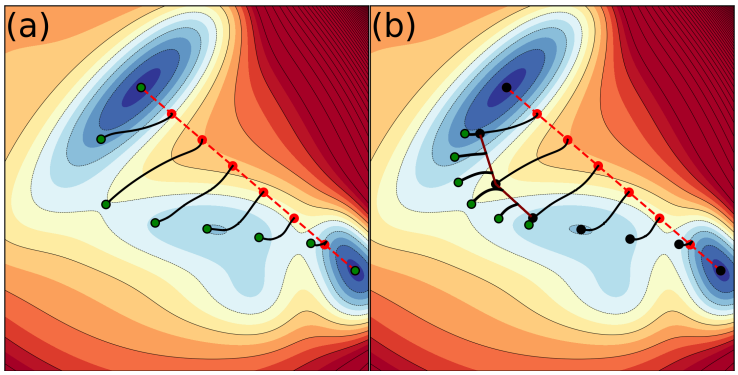


Figure 1: Illustration of the Z-NEB method and comparison with the standard CI-NEB method in calculations of the minimum energy path on the two-dimensional Müller-Brown surface.[16] The initial, linear interpolation path is shown by a red dashed line and the individual images by red circles. The displacement of the images during the iterative optimization is shown by black solid lines. The final position of the images converged to the minimum energy path is shown by green circles. Part (a) of the figure shows a successful calculation using CI-NEB with a total of 8 images, the two endpoints being fixed at the reactant and product state minima. In (b), an illustrative of the Z-NEB method is shown. In the first phase of Z-NEB, a loosely converged CI-NEB calculation is carried out and the final position of the images is shown by the black circles. A solid dark-red line indicates the region selected for introduction of additional images in the ZCI-NEB phase. It lies between the climbing image and its adjacent images. New set of images are introduced in this region by linear interpolation and images outside this region are not treated further. In the second phase, ZCI-NEB, the images are shifted to the minimum energy path, including the endpoint images which move in the direction of the atomic force perpendicular to the current path.

In the initial phase of Z-NEB, an initial path connecting the reactant and product energy minima is constructed, e.g., by using the image-dependent pair potential method.[17] The initial path is composed of a set of N system images with the end-point images constrained in the reactant and product energy minima. A CI-NEB calculation is then carried out, where the $N-2$ intermediate images are iteratively shifted towards the MEP but the calculation is stopped before convergence on the MEP is reached. The stopping criterion chosen here is based on the magnitude of the largest Cartesian component of the force acting on each of the atoms in the climbing image, i.e. $\max(|\mathbf{F}_{\mathrm{CI}}|) < \epsilon_{\mathrm{zoom}}$. In practice, the value of the tolerance, ϵ_{zoom}, needs to be chosen to be small enough so that the rough shape of the MEP has emerged, such as presence of all extrema but also large enough so that computational resources are not wasted on resolving irrelevant parts of the MEP to a higher precision, such as lower-lying saddle points or long tails of the MEP attributed to soft degrees of

freedom. Typically, a value of ϵ_{zoom} between 0.1 and 0.5 eV/Å is found to lead to reasonable performance.

Once this approximation of the MEP has been obtained, the region around the highest energy point on the MEP needs to be identified. Here, this region will be referred to as the Z-region. A piecewise-cubic polynomial is used to interpolate the energy along the current path.[?] In this interpolation scheme, both the energy of the images and atom force acting along the path are used in the interpolation. The inclusion of the tangential atom force in the interpolation leads to a more accurate energy estimate of the partially converged energy barrier than a simple (linear) interpolation of the energy of the images. Moreover, it can help determine which energy barrier in a multi-step reaction pathway is likely to be largest one on the MEP.

In the second phase, ZCI-NEB, the interpolated approximate MEP is used to select the region on the energy surface where the density of images is increased. The region of interest is defined by points on either side of the maximum where the energy has dropped to a certain fraction defined by a parameter γ where $\gamma \in (0, 1)$. The value of the energy, E^{γ}, defining this region is

$$E^{\gamma} = (1 - \gamma)(E^{\text{intp.}}_{\text{max}} - E^{\text{ref}}) + E^{\text{ref}} \tag{1}$$

where E^{ref} is the higher energy endpoint, $\max(E_1, E_{\text{N}})$, and $E^{\text{int}}_{\text{max}}$ is the energy of the maximum along the interpolated approximate MEP. The boundaries of the Z-region are then selected by locating the two intermediate images of the path that lie closest to the γ boundary-points. However, since the Z-region needs to bracket the point of maximum energy (i.e., the two boundary images of the Z-region need to be on opposite sides of the energy barrier), the CI cannot be selected as a boundary image of the Z-region. Therefore, if the CI is the nearest image to the point with energy E^{γ}, then a neighbor of the CI is chosen as a boundary image of the Z-region. Furthermore, if the approximate MEP is found to be strictly decreasing or increasing, one of the boundary images of the Z-region is chosen to lie between the endpoint minimum of higher energy and its neighboring image. This is important in order for the method to be able to find an energy maximum that is located close to one of the endpoint minima.

The selection of the γ-region ($\gamma = 0.2$) and the Z-region is illustrated in Fig. 2 for a Z-NEB calculation for the Müller-Brown energy surface. In this example, the Z-region corresponds to the path traced out by the CI and its two neighbors.

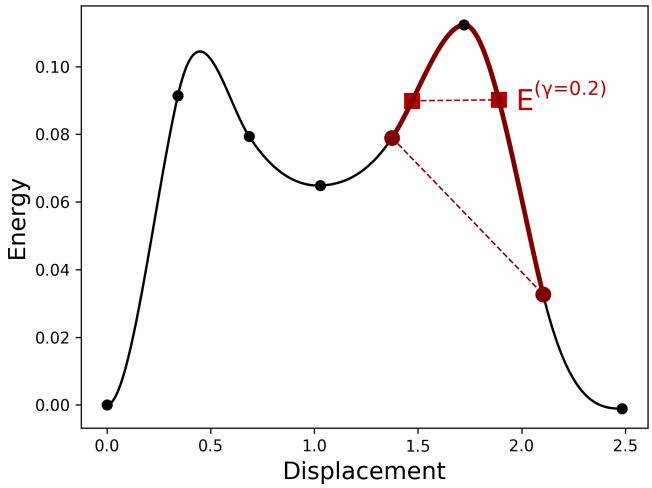


Figure 2: Energy along the approximate minimum energy path obtained in the CI-NEB calculation of Z-NEB on the Müller-Brown surface, as illustrated in Fig. 1(b). In ZCI-NEB, the neighborhood of the point of highest energy is selected according to $\gamma = 0.2$ and is highlighted by red boxes. The images closest to the boundaries of this region are then selected to define the Z-region, shown with dark-red color, where the density of images is increased.

Additional images are then constructed within the Z-region by either a piecewise linear or cubic interpolation of atomic configurations (see Sec. SI-1.1) and used to initiate a second CI-NEB calculation to find the point of maximum energy more accurately. The Z-NEB scheme, therefore, allows the saddle point region of the MEP to be resolved better than standard CI-NEB using a given number of images. Since the end-point images of the intermediate path in the ZCI-NEB calculation do not lie on the true MEP, it is important to optimize the position of these images along with the other images. In order to ensure efficient use of parallel computing, the ZCI-NEB calculation is carried out for $N - 2$ images, same as the number of movable images in the initial CI-NEB calculation, so the same central processors units continue to be used efficiently.

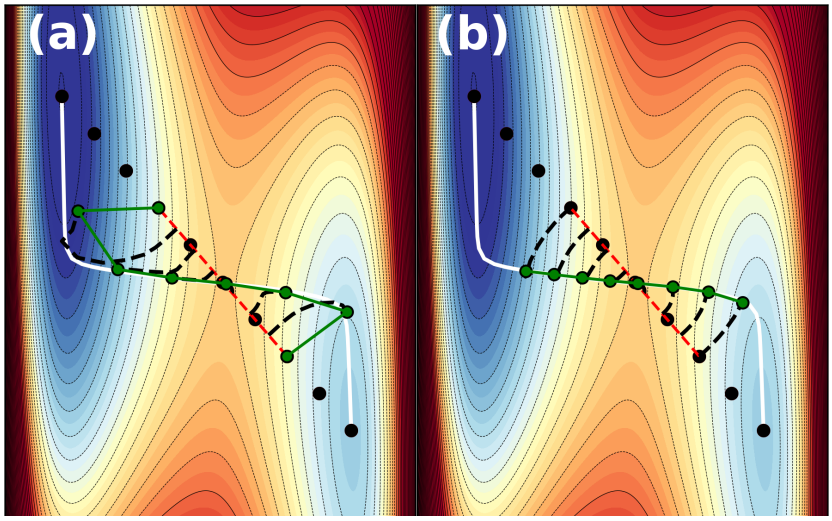


Figure 3: Illustration of different ways to handle endpoints in a CI-NEB calculation that does not include the local minima corresponding to the reactant and product states. This corresponds to the second phase of a Z-NEB calculation. In this example, the initial CI-NEB calculation is omitted and hence the ZCI-NEB starts from the linear interpolation. In (a), the endpoints are kept fixed even though they are not lying on the minimum energy path. This leads to images being pulled off the MEP. In (b), the end-points are iteratively moved according to the atomic force perpendicular to the current path. The red dashed line shows the region corresponding to $\gamma = 0.5$ where density of images is increased. The displacement of the images during the iterative optimization is shown by black dashed-lines. The converged position of the images is shown by green circles. The minimum energy path is shown by a white solid line.

The displacement of the two endpoint images to the MEP can be carried out by following the atomic force perpendicular to the current path,[11] see Fig. 3(a), or by making the end-point images follow the energy contour, in which case a force projection and a restraint needs to be added,[13,14] see Fig. 3(b). The latter option can be advantageous if an endpoint starts out far from the MEP, as the energy contour will prevent it from leaking down into an energy minimum. However, this calls for an additional 'stiffness' parameter (analogous to the spring-coefficient) in the calculation. If the value of the stiffness parameter is not chosen carefully the performance of the CI-NEB calculation can be affected. The first option is used in the calculations presented here. In any case, it is important to optimize the end-point images in ZCI-NEB. If the end-points are simply fixed during the optimization, the path becomes kinked and the neighboring images of the end-points may not converge on

the MEP, as shown in Fig. 3(c). This can drastically affect the convergence properties of the method and reduce the computational efficiency of the calculation, in particular if a small number of images is used. An example of that is given in the heptamer benchmark calculations in the next section.

2.1 Computational settings

For the heptamer island benchmark calculations, the interatomic interactions are described by using a simple Morse potential. The potential parameters and configurations set are given in Ref.[15] This benchmark has become used extensively to assess methods for finding MEPs and saddle points. The CI-NEB and Z-NEB calculations of the benchmark reactions are carried out with both the L-BFGS[20] (for a demonstration of various choices of Δ_{step}, see Sec. SI-1.2) and FIRE optimization methods.[21] The computational settings for these calculations are given in Sec. SI-4. To compare performance, the calculations are also carried out using the AutoNEB method as implemented in the ASE-3.17 simulation package.[22] A slight modifications to the implementation was made in order to be able to use the same convergence criterion in the two types of calculations, see Sec. SI-4.4. The AutoNEB calculations make use of the FIRE optimization method, since the NEB/L-BFGS method is not available in ASE.

For the two molecular reactions, the calculations are carried out using the ORCA software, where the B3LYP[23–25]+D3(BJ)[26]/def2-SVP level of theory is chosen for the density functional theory evaluation of the energy and atomic forces. The reactant and product configurations for the reactions are obtained from Ref.[27] and re-optimized on the aforementioned level of theory. In both the Z-NEB and CI-NEB calculations, the IDPP method[17] is used to construct the initial paths. The climbing image is activated from the first step of the calculation and the spring-coefficient is chosen to be 5.0 eV/Å^2 (0.051 Ha/Bohr^2). The L-BFGS optimization method is used with a maximum allowed step-size, Δ_{step}, of 0.2 Å. If the maximum component of a step exceeds Δ_{step}, the L-BFGS optimizer is restarted. In the Z-NEB calculations, $\gamma = 0.5$ and $\epsilon_{\text{zoom}} = 0.5$ eV/Å (0.01 Ha/Bohr) and a linear interpolation of Cartesian atomic coordinates is used to generate additional configurations. Since, the objective of the calculations is to locate the point of maximum energy along the MEP, the focus is on the atomic forces of the CI and the calculations are considered to be converged when largest Cartesian component of the force on any atom satisfies $\max(|\mathbf{F}_{\text{CI}}|) < 0.0005$ Ha/Bohr (0.026 eV/Å) and the $\text{RMS}(\mathbf{F}_{\text{CI}}) < 0.0003$ Ha/Bohr (0.015 eV/Å). The maximum energy configurations obtained is verified to be a first order saddle point on the energy surface by ensuring that the Hessian has only one negative eigenvalue. The atomic coordinates of the

reactant, saddle point and product configurations are given in Sec. SI-2 and SI-3.

3 Results

The Z-NEB method is first applied to a benchmark test that has been used widely to assess the performance of methods for finding MEPs and saddle points, the heptamer island transitions on a solid surface. The results are described in Sec. 3.1 and comparison is made with both the standard CI-NEB method and the AutoNEB method. In Sec. 3.2, the Z-NEB method is used in combination with DFT evaluation of the energy and atomic forces in calculations of two molecular reactions, hydrolysis reaction of ethyl acetate and rearrangement reaction of 1,5-hexadiene. For the latter example, the convergence rate using various number of images in the calculations is analyzed.

3.1 Benchmark test on heptamer island

To evaluate the performance of the Z-NEB method on a well defined benchmark, calculations are carried out for various rearrangement transitions of a heptamer island on a FCC(111) surface. This set includes 59 transitions, some of which differ only by symmetry. For comparison, both CI-NEB and AutoNEB calculations are carried out using the same convergence criteria. The results from the calculations is summarized in table. 1.

Table 1: Application of CI-NEB, Z-NEB and AutoNEB methods to the heptamer island benchmark set.[15] Both the L-BFGS and FIRE optimization method are used for CI-NEB and Z-NEB methods, but could not be used with AutoNEB because those calculations where carried out with ASE software and L-BFGS with NEB is not available there. The efficiency of the NEB calculations is measured by the number of energy and force evaluations (FCs) required for the calculations to reach convergence, i.e. $||\mathbf{F}_{\mathrm{CI}}|| < 0.001$ eV/Å. Standard deviation, σ, in the number of FCs for the 59 transitions is given, as well as minimum, maximum and median number of FCs. The unsuccessful calculations (with AutoNEB only) are omitted from the data on FCs.

Method	⟨FCs⟩	σ(FCs)	min(FC)	median(FC)	max(FC)	#failed
CI-NEB/L-BFGS	387	135	182	382	932	0
Z-NEB/L-BFGS	348	100	182	357	847	0
CI-NEB/FIRE	907	528	412	742	2337	0
Z-NEB/FIRE	876	498	477	662	2167	0
AutoNEB(9)/FIRE	965	678	463	748	4733	3
AutoNEB(12)/FIRE	953	487	494	744	2264	4

The CI-NEB calculations utilizing the L-BFGS optimization method successfully locate the saddle points of the benchmark set without any failure in an average of 387(±135) energy/gradient evaluations per transition. These results are in agreement with previously reported CI-NEB/L-BFGS calculations for this benchmark.[15] The Z-NEB method is found to be slightly more efficient than CI-NEB, by 10%, with an average of 348(±100) evaluations per transition when using $\gamma = 0.2$ and $\epsilon_{\mathrm{zoom}} = 0.2$ eV/Å. The MEPs for these transitions are mostly simple and without long tails (see Sec. SI-4.1) so it is reasonable that the advantage of Z-NEB over CI-NEB is small. If the FIRE optimization method is used instead of L-BFGS, the average evaluations per transition increases to 907(±528) and 876(±498) for CI-NEB and Z-NEB, respectively. Again, Z-NEB gives slight reduction in computational effort, but only marginal. This difference in the computational efficiency of L-BFGS and FIRE is also to be expected since the L-BFGS method uses both the NEB gradient and approximate second derivative information in the optimization[20] while FIRE is a dynamical optimization method that uses only the NEB gradient.[21]

For illustration purposes, Z-NEB calculations were also carried out where the endpoints in the second phase are fixed. Then, 9 of the reactions require over 2000 energy/force evaluations, see Sec. SI-4.3. It is clearly important to allow the endpoints to relax down to the MEP along with the other images. Fixing endpoints at locations that are not on the

MEP may often lead to problems, in particular if too few images are used to resolve the path.

For comparison, calculations were also carried out with the AutoNEB method which is designed to accomplish a similar goal as Z-NEB. The initial path is the same as for the Z-NEB and CI-NEB calculations, an IDPP path with $N_{\text{initial}} = 7$. Then, additional images are added, on-the-fly, into the largest energy (or geometrical gap) along the path after a fixed number of optimization steps. Here, an image is added every 7 steps. The addition of images is carried out until the total number of images in the path becomes $N = 9$ in the AutoNEB(9) calculation and $N = 12$ in the AutoNEB(12) calculation. At that point, the climbing image is activated and the saddle point targeted. To maintain the parallelization property of NEB in AutoNEB (analogous to Z-NEB), only N_{s} (where $\text{N}_{\text{s}} = N_{\text{initial}} - 2$) images surrounding the most recently added image are computed and moved in each optimization step. All other images remain fixed in the calculations. Therefore, AutoNEB(12) with $N_{\text{s}} = 5$ is expected to behave in a similar manner to that of Z-NEB($N = 7$), where 5 images are added into the higher energy regions of the path.

Surprisingly, both AutoNEB(9) and AutoNEB(12) exhibit failures, i.e. are unable to converge to the saddle point to the specified tolerance for 3 and 4 different transitions, respectively. Two of the failed calculations are attributed to the inability to locate a low energy maximum that is close to an endpoint. For the remaining failures, i.e., 1 failed calculation in AutoNEB(9) and 2 failed calculations in AutoNEB(12), the optimization reached the maximum number of allowed optimization steps and it appears the optimization has become erratic. For the remaining reactions, AutoNEB locates the saddle points in an average of 965($\pm$678) and 953($\pm$487) evaluations and hence proves to be less efficient than Z-NEB and even less efficient than the standard CI-NEB method in disagreement to the findings of Ref. [12]

3.2 Application to two molecular reactions

3.2.1 Hydrolysis of ethyl-acetate

The hydrolysis of ethyl-acetate to form acetic acid and ethanol is a rather complicated reaction with an MEP that has a rather high and narrow energy barrier. The path is long so a large number of images is required for a standard CI-NEB calculation to converge. When 7 images are used in total including the endpoints, the CI-NEB calculation does not converge, but the Z-NEB reaction does. The number of energy/force calculations is, however, quite large, 1345. This reaction can occur by either a concerted mechanism or in a step-wise manner. Here, the concerted reaction mechanism is obtained in the Z-NEB calculation. The key configurations of the reaction path and energy profiles are shown in Fig. 4. Note, that

the alternative and slightly lower energy step-wise mechanism can be found by insertion of an image into the initial path generation of NEB, see Sec. SI-2.2.

The Z-NEB calculation is started from an IDPP path with $N = 7$, constructed between the reactant and product minimum energy configurations. The highest energy image on this path is 7.1 eV above the reactant energy minimum. Then, a CI-NEB calculation is carried out and the path is iteratively shifted downhill in energy towards the MEP. This calculation is terminated, i.e., $\max(|\mathbf{F}_{\mathrm{CI}}|) < \epsilon_{\mathrm{zoom}}$, after 282 energy/force evaluations. At this point, the reaction is predicted to follow the single concerted mechanism with an approximate path length of 10.6 Å. The approximate MEP is already characterized by a relatively narrow energy barrier and with a height of 2.1 eV.

In the second Z-NEB phase, the two images on each side of the CI are automatically selected to bracket the energy barrier and become the boundary images of the Z-region. These images are shown as insets in Fig. 3. These three images, the CI and its two neighbors are therefore used to interpolate and construct a new intermediate path with 5 images. The length of the interpolated path is roughly 3.7 Å. Then, a second CI-NEB calculation is carried out, where all of the images, including the two end-point images, are optimized to the MEP.

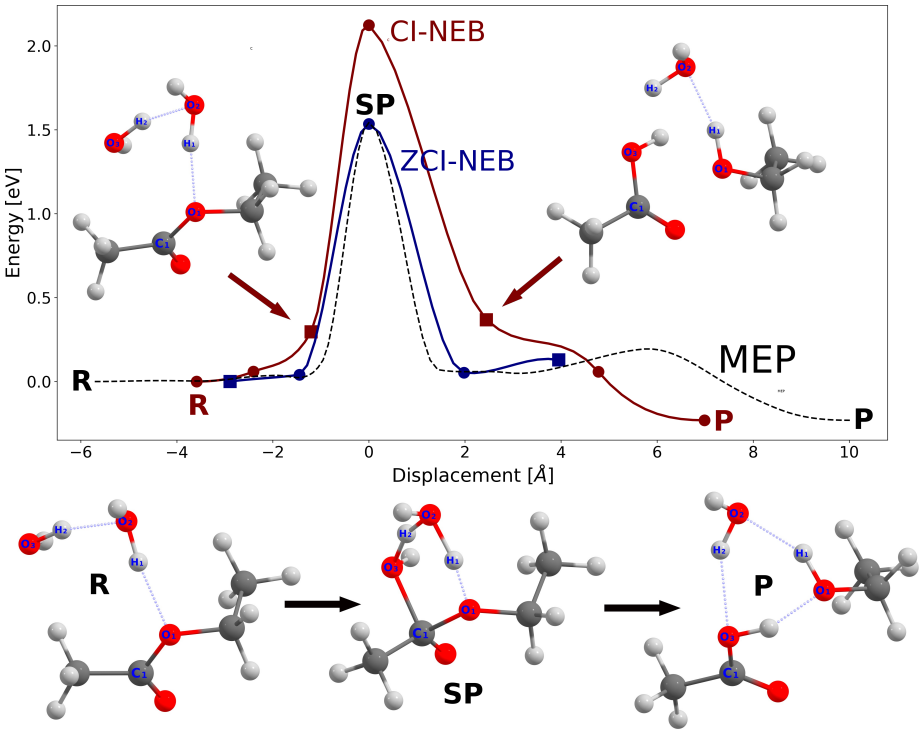


Figure 4: Illustration of a Z-NEB calculation of the hydrolysis of ethyl acetate to form acetic acid and ethanol. In the lower panel, the reactant (denoted by **R**), saddle point (denoted by **SP**) and product (denoted by **P**) configurations are shown. For clarity, atoms that are displaced most in the reaction are labeled as C_1, O_1–O_3 and H_1–H_2. The saddle point is characterized by a 6-membered pericyclic configuration. In the upper panel, the approximate MEP obtained after the initial CI-NEB calculation is shown by the red curve. The energy path obtained after the second phase ZCI-NEB calculation is shown by a blue curve. The boundary images obtained from the first phase of Z-NEB are shown by red and blue squares and the configurations are shown as insets. The full minimum energy path for the reaction is shown by a black dashed line.

The path obtained with Z-NEB predicts that the hydrolysis occurs by a single-step nucleophilic attack of a water molecule on the carbonyl group of the ester. Simultaneously, the C-O bond is ruptured and a double proton transfer occurs where the ethoxy group is protonated. The saddle point configuration is characterized by a 6-membered ring configuration formed by atoms C_1-O_3-H_2-O_2-H_1-O_1(-C_1) shown in the lower panel of Fig. 4. The energy barrier is estimated to be 1.55 eV and the path length of the partial MEP obtained in

Z-ZNEB(2) is found to be 6.8 Å. The point of maximum energy is verified to be a first order saddle point on the energy surface with a single negative eigenvalue of the Hessian matrix, corresponding to a magnitude of 1251 cm^{-1} for the imaginary frequency.

The large number of evaluations required to obtain the saddle point can be attributed to both the complicated reaction mechanism of this reaction and also to the relatively poor resolution of the saddle point region. In Z-NEB (and CI-NEB) calculations for complicated reactions such as this one, it may be advantageous to use more images to better resolve the path, because low resolution can drastically reduce computational efficiency.

3.2.2 Rearrangement of 1,5-hexadiene

As for the hydrolysis reaction studied in the previous section, the rearrangement of 1,5-hexadiene is a complicated reaction with simultaneous bond ruptures and formation of a σ-bond and two π-bonds. The energy barrier for the reaction is found to be 1.41 eV. The saddle point configuration has again 6-membered ring geometry and is verified to be of first order, with an imaginary frequency of magnitude 530 cm^{-1}. The reactant, saddle point and product configurations are shown in Fig. 5

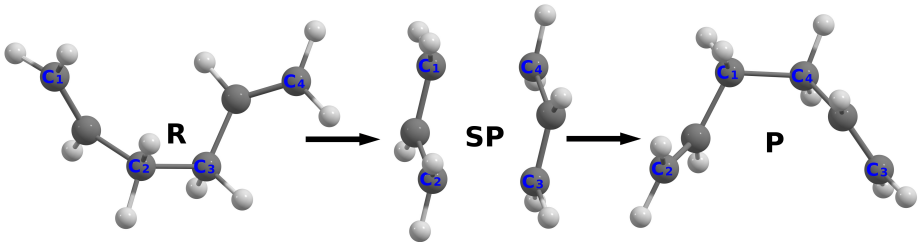


Figure 5: The reactant (denoted by **R**), saddle point (denoted by **SP**) and product (denoted by **P**) configurations of the rearrangement reaction of 1,5-hexadiene. For clarity, the four C atoms involved in bond formation/breaking are labeled from C_1 to C_4 in all configurations.

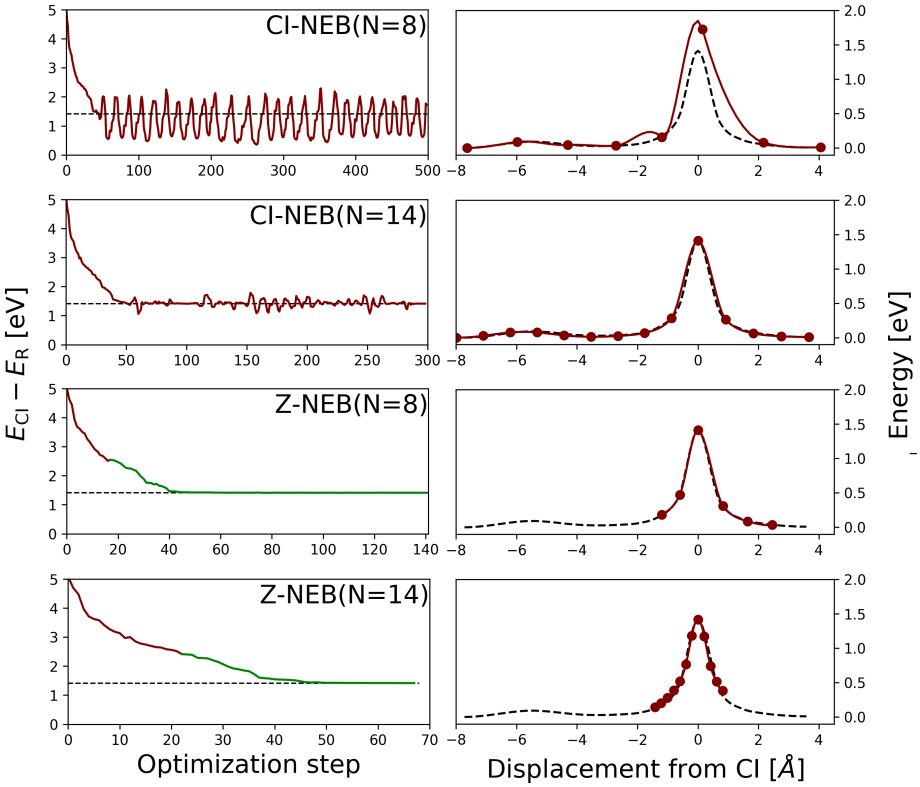


Figure 6: Results from four different calculations of the rearrangement of 1,5-hexadiene, namely CI-NEB and Z-NEB using 8 and 14 images. In the left-panel, the energy difference between the climbing image and the reactant energy minimum is shown as a function of optimization step. The maximum energy on the true minimum energy path is given by the black dashed line. In the right-panel, the energy profile of the reaction obtained from the last optimization step is shown. The minimum energy path is shown with the dashed black line. The CI-NEB calculation with 8 images is unable to converge in 500 optimization steps, while the Z-NEB calculation converges in 140 steps when 8 images are used and in 68 steps when 14 images are used.

To compare the convergence behavior of CI-NEB and Z-NEB calculations are carried out using both methods for $N = 8$ and $N = 14$. The results from these calculations are summarized in Fig. 6. The CI-NEB calculation with $N = 8$ is unable to converge to the saddle point in 500 optimization steps. In the first 50 optimization steps, the relative energy of the CI quickly drops from 5.0 eV (the maximum energy along the IDPP path) to 1.3 eV.

After this, the energy of the CI begins to exhibit large fluctuations around the maximum. The CI is unable to converge to the maximum on the MEP, while the other intermediate images already lie quite close to the MEP. It is evident, from Fig. 6, that the two neighboring images of the CI barely lie at the base of the energy barrier and hence a poor tangent estimate is obtained at the CI. This leads to large and inaccurate optimization steps which lead to fluctuations in the estimated maximum of the energy. By increasing the number of images to $N = 14$, CI-NEB is able to converge in 300 optimization steps (corresponding to 3986 energy/force evaluations). In this case, the CI exhibits smaller fluctuations during the optimization. When such a large number of images being used in the calculation, most of the images are simply placed on an uninteresting part of the MEP that corresponds to the overall motion of the allyl-group (which includes the C_4 atom, see Fig. 5).

In Z-NEB, the images are distributed into the region where the reaction actually occurs, i.e. the region in configuration space where the σ and π-bond are being broken and formed. It is this part of the path that gives rise to the energy barrier. For the Z-NEB calculation with $N = 8$, the first phase of Z-NEB successfully finishes in about 18 optimization steps (corresponding to 134 energy/force evaluations). The approximate MEP obtained in that way is shown in Sec. SI-3.2. The ZCI-NEB calculation converges to the point of maximum energy in additional 122 optimization steps (corresponding to 768 energy/force evaluations). In other words, Z-NEB with 8 images requires only roughly half the number of the optimization steps that CI-NEB with 14 images requires to converge and only 23% of the number of energy/gradient evaluations. It is clear from Fig. 6 that the resolution of the energy barrier is improved significantly by using Z-NEB. However, the long and flat tail in the curve showing energy as a function of optimization step indicates that there are still some small energy fluctuations of the CI that hinder it from converging easily to the maximum on the MEP. By increasing the number of images further to $N = 14$, Z-NEB converges in a total of 68 optimization steps, corresponding to 926 energy/force evaluations with a high resolution of the energy barrier. The optimal performance for Z-NEB as a function of N is obtained for N=10 where the calculation completes in 62 optimization steps, see Sec. SI-3.3.

4 Discussion

The method presented here makes CI-NEB calculations better suited for atomic rearrangements with long and complicated paths where part of the path does not involve large change in energy. Previously, CI-NEB has in particular been used in studies of transitions in and on the surface of solids. There, MEPs can be fairly short as the atoms are confined to a well defined space. In molecular reactions, however, part of an MEP can represent rotation or translation of a molecular fragment where the energy does not change appreciably and this leads to long energy tails. If an even distribution of images is used in such cases, the resolution of the path in the part where energy has increased significantly, the critical part of the path, can be low and this leads to poor estimate of the tangent to the path at the CI. While this can be overcome by simply adding more images to represent the path, this can increase the computational effort beyond available resources. The method presented here provides a more economical solution as the density of images is increased only in the most critical region once a reasonable accurate estimate has been obtained for the shape of the path.

The benchmark study of the heptamer island is carried out mainly because it is a well documented benchmark and it is important so show that the method is robust in that case. But this system is simple in the sense that the atoms are confined to well defined positions and the paths are relatively simple. As a result, the method does not give significant savings over CI-NEB, only about 10% in that case. It is interesting, however, to see that the AutoNEB method which was proposed a few years ago for the same purpose of the present Z-NEB method and performs poorly on this benchmark. First of all, AutoNEB does not reach convergence for a few of the transitions and secondly there is no savings in computational effort as compared with standard CI-NEB. This is surprizing and probably relates to the fact that new images are introduced repeatedly and the endpoints are fixed in place, often not even in vicinity to the MEP. The AutoNEB method is, furthermore, complex so it is difficult to ascertain where the culprit lies. One of the advantages of Z-NEB is its simplicity.

The application of Z-NEB to the two complex molecular reactions with MEPs that contain extended tails demonstrates the power of the method. For a number of images which is too low for CI-NEB to converge, Z-NEB can reach convergence. Since the computational effort in terms of core-hours is largely proportional to the number of images, fewer images used in the calculation translates directly to smaller number of core-hours. A more extensive test of the method for a large data base of molecular reactions including over 100 reactions is ongoing and will be reported elsewhere.

The Z-NEB method has been implemented and made available in the ORCA software.[28,29]

5 Data availability statement

The authors confirm that the data supporting the findings of this study are available within the article and/or its supplementary materials.

Acknowledgement

This work was supported by the Icelandic Research Fund and the University of Iceland Research Fund. V.Á. acknowledges a fellowship from the Doctoral Fund of the University of Iceland. The calculations were carried out at the Icelandic Research High Performance Computing (IRHPC) facility.

References

(1) Mills, G.; Jónsson, H.; Schenter, G. K. Reverisble work based transition state theory: application to H_2 dissociative adsorption. *Surf. Sci.*, **1995**, 324, 305.

(2) Jónsson, H.; Mills, G.; Jacobsen, K. W. Nudged elastic band method for finding minimum energy paths of transitions. 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.

(3) Ásgeirsson, V.; Jónsson, H. Exploring potential energy surfaces with saddle point searches. In *Handbook of Materials Modeling: Methods: Theory and Modeling*; Andreoni, W.; Yip, S., Eds.; Springer Cham, **2020**, pp.689–714.

(4) Wigner, E. The transition state method, *Trans. Faraday Soc.*, **1938**, 34, 29.

(5) Vineyard, G. H. Frequency factors and isotope effects in solid state rate processes, *J. Phys. Chem. Solids*, **1957**, 3, 121.

(6) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. Climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys*, **2000**, 113, 9901–9904

(7) McPherson, K. E.; Bartolotti, L. J.; Morehead, A. T.; Sargent, A. L. Utility of the Nudged Elastic Band Method in Identifying the Minimum Energy Path of an Elementary Organometallic Reaction Step. *Organometallics*, **2016**, 35, 1861–1865.

(8) Zimmerman, P. Reliable Transition State Searches Integrated with the Growing String Method. *J. Chem. theory and Comput.*, **2013**, 9, 3043–3050.

(9) Zimmerman, P. Growing string method with interpolation and optimization in internal coordinates: Method and examples. *J. Chem. Phys.*, **2013**, 138, 184102.

(10) Maragakis, P.; Andreev, S. A.; Brumer, Y.; Reichman, D. R.; Kaxiras, E. Adaptive nudged elastic band approach for transition state calculation. *J. Chem. Phys.*, **2002**, 117, 4651.

(11) Zhang, J.; Zhang, H.; Ye, H.; Zheng, Y., Free-end adaptive nudged elastic band method for locating transition states in minimum energy path calculation. *J. Chem. Phys.*, **2016**, 145, 094104.

(12) Kolsbjerg, E.L.; Groves, M.N.; Hammer, B. An automated nudged elastic band method. *J. Chem. Phys.*, **2016**, 145, 094107.

(13) Zhu, T.; Li, J.; Samanta, A.; Kim, H.G.; Suresh, S. Interfacial plasticity governs strain rate sensitivity and ductility in nanostructured metals. *Proc. Nat. Acad. Sci.*, **2007** 104, 3031-3036.

(14) Ásgeirsson, V.; Arnaldsson, A.; Jónsson, H. Efficient evaluation of atom tunneling combined with electronic structure calculations. *J. Chem. Phys.*, **2018**, 148, 102334.

(15) Chill, S. T.; Stevenson, J.; Ruehle, V.; Shang, C.; Xiao, P.; Farrell, J. D.; Wales, D. W.; Henkelman, G. Benchmarks for Characterization of Minima, Transition States, and Pathways in Atomic, Molecular, and Condensed Matter Systems. *J. Chem. Theory and Comput.*, **2014**, 10, 5476–5482.

(16) Müller, K.; Brown, L. D. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theor. Chim. Acta*, **1979**, 53, 75-93.

(17) Smidstrup, S.; Pedersen, A.; Stokbro, K.; Jónsson, H. Improved initial guess for minimum energy path calculations. *J. Chem. Phys.*, **2014**, 140, 214106.

(18) Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A. A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *J. Chem. Phys.*, **2004**, 120, 7877.

(19) Henkelman, G. and Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.*, **2000**, 113, 9978–9985.

(20) Liu, D. C.; Nocedal, J. On the limited memory BFGS method for large scale optimization. *Math. program.*, **1989**, 45, 503–528.

(21) Bitzek, E.; Koskinen, P.; Gähler, F.; Moseler, M; Gumbsch, P. Structural Relaxation Made Simple. *Phys. Rev. Letters*, **2006**, 97, 170201.

(22) Larsen, A. H.; et al. The atomic simulation environmentâĂŤa Python library for working with atoms. *J. Phys.: Condens. Matter*, **2017**, 29, 273002.

(23) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, **1988**, 37, 785–789.

(24) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, **1988**, 38, 3098–3100.

(25) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, **1993**, 98, 5648–5652.

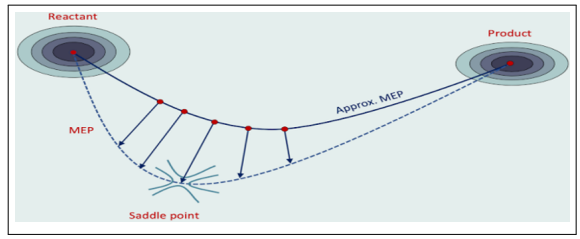
(26) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.*, **2011**, 32, 1456–1465.

(27) Birkholz, A.B; Schlegel, H.B. Using bonding to guide transition state optimization. *J. Comp. Chem.*, **2015**, 36(15), 1157-1166.

(28) Neese, F. The ORCA program system. *WIRES Comput. Mol. Sci.*, **2012**, 2, 73-78.

(29) Neese, F. Software update: the ORCA program system, version 4.0. *WIRES Comput. Mol. Sci.*, **2018**, 8, e1327.

Graphical TOC Entry



Supporting information: Nudged elastic band calculations with focus on the region around the highest saddle point

Vilhjálmur Ásgeirsson, Benedikt Orri Birgisson, and Hannes Jónsson*

Science Institute and Faculty of Physical Sciences, University of Iceland VR-III, 107 Reykjavík, Iceland

E-mail: hj@hi.is

SI-1 Methodology

SI-1.1 Configurational interpolation and construction of the intermediate path

For the interpolation of the Z-region and construction of the intermediate path, two different configurational interpolants are implemented, a piecewise linear and cubic polynomial. Following the discussion of Ref.,[2] a parameter s is defined for any sub-interval of the path, i.e., $s \in [0, d_{i,i+1}]$ where $d_{i,i+1}$ is the pairwise distance between images i and $i+1$. The set of atom coordinates for a new intermediate image k (located within the interval s) is then computed by either a cubic polynomial,

$$\mathbf{R}_k^j(s_k) = a_i^j s_k^3 + b_i^j s_k^2 + c_i^j s_k + d_i^j$$

or alternatively by a straight line,

$$\mathbf{R}_k^j(s_k) = c_i^j s_k + d_i^j$$

where $j = 1, ..., 3M$. The polynomial coefficients are expressed as,

$$
\begin{aligned}
a_i^j &= -\frac{2(\mathbf{R}_{i+1}^j - \mathbf{R}_i^j)}{(d_{i+1,i})^3} + \frac{2\tau_{i+1}^j + \tau_i^j}{(d_{i+1,i})^2} \\
b_i^j &= \frac{3(\mathbf{R}_{i+1}^j - \mathbf{R}_i^j)}{(d_{i+1,i})^2} - \frac{2\tau_{i+1}^j + \tau_i^j}{(d_{i+1,i})} \\
c_i^j &= -\tau_i^j \\
d_i^j &= \mathbf{R}_i^j
\end{aligned}
$$

SI-1.2 Alternative treatment of the end-points in ZCI-NEB

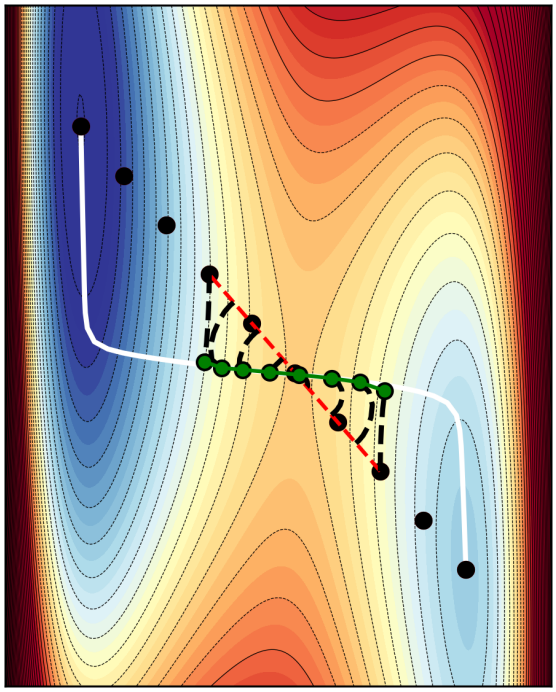


Figure 1: Illustration of an alternative method to handle the endpoints in a CI-NEB calculation that does not include the local minima corresponding to initial and final states. In this work, this corresponds to the ZCI-NEB calculation in Z-NEB. The first phase of Z-NEB is omitted so that ZCI-NEB starts from the linear interpolation. For this alternative treatment, the endpoints are relaxed downhill in energy in the direction normal to the path and simultaneously restraint to move along the energy levels of the end-point images from which the ZCI-NEB calculation started, see Ref.[1]. This renders the ZCI-NEB calculation more sensitive to the interplay of the spring-constant and the additional stiffness parameter introduced to the ZCI-NEB calculation. However, it may be beneficial to use these restraints when the typical treatment of the endpoint images results in the endpoints sliding downhill away from the Z-region. The red dashed line shows the region corresponding to $\gamma = 0.5$ where density of images is increased. The displacement of the images during the iterative optimization is shown by black dashed-lines. The converged position of the images is shown by green circles. The minimum energy path is shown by a white solid line.

SI-1.3 Interplay of the L-BFGS optimization method and the parameter for the maximum allowed step-size

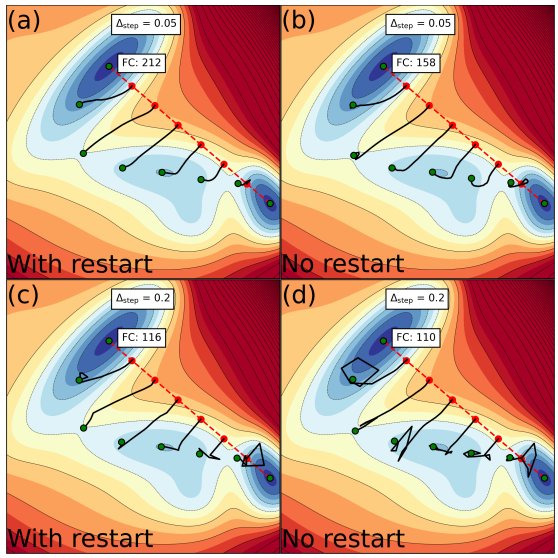


Figure 2: CI-NEB calculations carried out on the Müller-Brown potential using 8 images and the L-BFGS optimization method. In (a) and (b), the maximum displacement allowed per optimization step (Δ_{step}) is set as 0.05 Å. For (c-d), $\Delta_{\text{step}} = 0.2$ Å. The L-BFGS optimization is restarted if a step exceeding Δ_{step} is attempted in (a) and (c). In each figure, the number of energy and gradient evaluations (FC) required to reach convergence, i.e., $\max(|\mathbf{F}^{\perp}|) < 0.01$ eV/Å, is given.

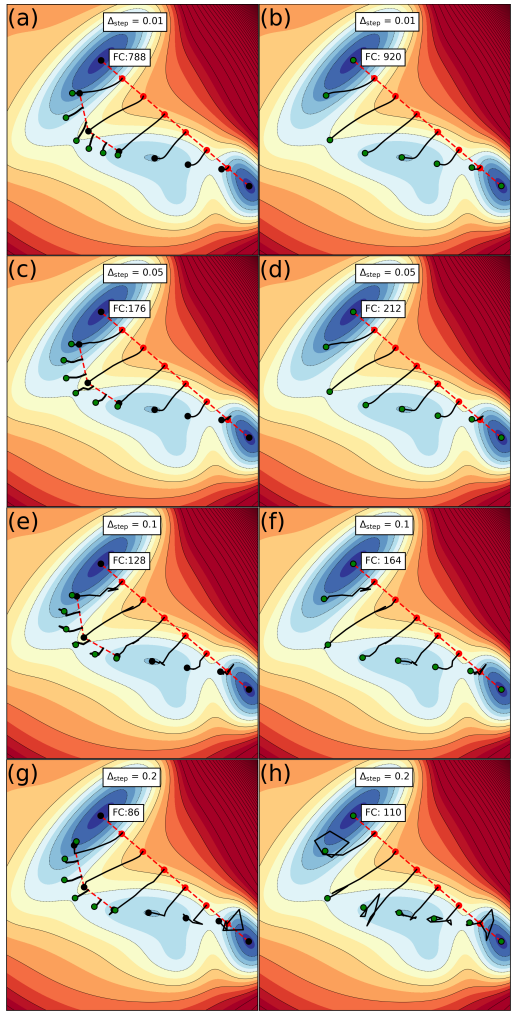


Figure 3: Z-NEB (a,c,e and g) and CI-NEB (b,d,f and h) calculations carried out on the Müller-Brown potential using 8 images and the L-BFGS optimization method. The Z-NEB calculations use $\epsilon_{\text{zoom}} = 0.05$ eV/Å and $\gamma = 0.2$. In (a) and (b), the maximum displacement allowed per optimization step (Δ_{step}) is set as 0.01 Å. For (c-d), (e-f), and (g-h) $\Delta_{\text{step}} = 0.05, 0.1$ and 0.2 Å respectively. The L-BFGS optimization is restarted if a step exceeding Δ_{step} is attempted. In each figure, the number of energy and gradient evaluations (FC) required to reach convergence, i.e., $\max(|\mathbf{F}^{\perp}|) < 0.01$ eV/Å, is given.

SI-2 Hydrolysis of ethyl acetate

SI-2.1 System configurations

Reactant configuration:

C	-0.68432250105073	1.91791629735588	-0.48054172511861
H	-1.53612906758603	1.39200879526999	-0.00763334155213
H	-0.83315205610761	1.87569899655704	-1.56995655798047
H	-0.65654289671130	2.95891955012630	-0.13660638980533
C	0.59484149902475	1.23380963422406	-0.09758994569369
O	0.70898707632263	0.01842577412506	-0.72081968850994
O	1.42971320488056	1.64286476248071	0.66560441333084
C	1.86656486059245	-0.78531096223494	-0.39728962813732
H	2.01585446928416	-1.42458217904235	-1.27989464656442
H	2.73153579083829	-0.11638244619790	-0.28559024443314
C	1.63342670654854	-1.60957858984604	0.85636403010327
H	1.53094654670859	-0.95504723523695	1.73479002918539
H	2.48861812064091	-2.28281240010149	1.02769214998189
H	0.72168653835481	-2.21730885946548	0.75169516817969
O	-3.08424213083654	0.07891431952033	0.77541797877554
H	-2.80541919254909	0.00277073086586	1.69817428405240
H	-2.61283636975558	-0.66039391525464	0.33602125645008
O	-1.46040769893080	-1.73547363087178	-0.57123759539055
H	-1.80835617702838	-2.01909111355727	-1.42736715321640
H	-0.76193572263965	-1.08404750871642	-0.78198540365711

Saddle-point configuration:

C	-0.85376525454881	2.17259427816116	-0.64840341507531
H	-1.54817165072453	2.82025146723338	-0.09436408154808
H	-1.39779128551170	1.65823202472105	-1.45087877074951
H	-0.05360359627671	2.78697180744737	-1.08536046930335
C	-0.22905403891002	1.19884078535470	0.31841812064083
O	0.41299075449631	0.02901904167650	-0.72574006028747
O	0.49509720633106	1.45986327771577	1.24210179809610
C	1.63414087217855	-0.55140150110077	-0.26949241051421
H	2.06891911027358	-1.07107161542927	-1.14060179300345

H	2.31345643687226	0.26913942957690	0.00909542035122
C	1.45435812909026	-1.51168509460286	0.90068966578459
H	1.18657511502431	-0.95624011455040	1.81068158624829
H	2.39459627290828	-2.05295767381760	1.09508871268615
H	0.67032001185536	-2.25516262703212	0.68574289246974
O	-1.58414788508972	0.28868225535086	0.81925084227495
H	-1.32292997742629	0.01692977095473	1.71298572444690
H	-1.69023570005660	-0.67959005065353	0.07258604167067
O	-1.44194491723344	-1.46403208429848	-0.78761116099528
H	-2.04422573007875	-1.34450094687760	-1.53468869091774
H	-0.46458387317338	-0.81388242982979	-0.92949995227502

Product configuration:

C	-0.59362983796967	2.09482551323932	0.42658340812862
H	-1.43563447085298	1.60952793130587	0.94622089828471
H	-1.00943511839187	2.62066966862460	-0.44639209391574
H	-0.10387242347304	2.80537324007282	1.10297243810319
C	0.39841969346684	1.04909415803316	-0.02197746531242
O	1.20958293000285	-1.93861689743216	-1.55620147722528
O	1.53249084457725	0.95601529136763	0.38483188281662
C	2.25738360894619	-2.46546696125699	-0.74454869786923
H	2.49205839604167	-3.49476230578345	-1.07643391644478
H	3.14479634810531	-1.84620204141558	-0.94842853233162
C	1.92313841148143	-2.44318414921422	0.74000608725730
H	1.81058080844863	-1.40694407085767	1.09181966369737
H	2.72667978385947	-2.92118282379958	1.32377590095782
H	0.98716210090805	-2.99216114439314	0.93836271614661
O	-0.12435470610020	0.20949915410849	-0.93193814727354
H	0.56716565342026	-0.47887801377582	-1.19841583943326
H	-1.28140019964160	-1.38907734173259	-0.62528680386988
O	-1.36190556780431	-2.35218763442974	-0.75863577530646
H	-1.99667911942588	-2.43749596262554	-1.48403611578522
H	0.37048932440153	-2.40494655003541	-1.34494898062478

SI-2.2 Importance of initial path generation in CI-NEB and Z-NEB

The hydrolysis of ethyl acetate reaction is used here to emphasize the importance of initial path generation in NEB calculations and variants thereof. It is shown how the chosen initial path can affect the final outcome of a CI-NEB calculation.

A CI-NEB calculation using $N = 10$ is carried out for the hydrolysis of ethyl acetate. The calculation is started from an initial path generated by the IDPP method[3] without any additional input, apart from the given reactant and product configurations. The barrier height for the reaction is found to be 36.1 kcal/mol and the SP is confirmed to be a first order SP with an imaginary frequency of -1265 cm^{-1}. The reaction mechanism obtained by the CI-NEB calculation occurs in a single concerted step and is visualized in the lower panel of Fig. 4.

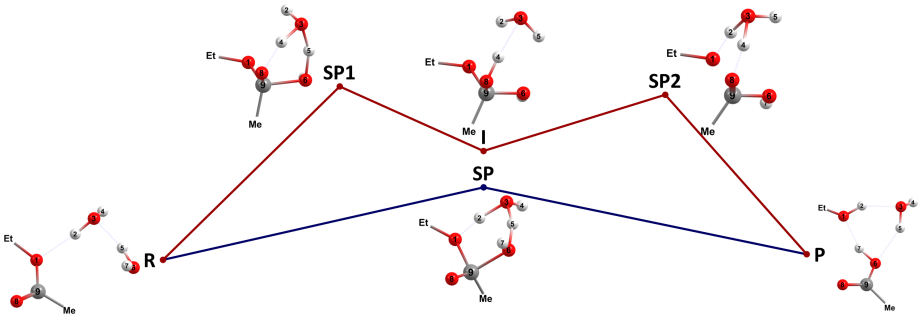


Figure 4: The reactant (**R**), product (**P**) and saddle point (**SP**) configurations are visualized for the concerted hydrolysis reaction of ethyl acetate in the lower panel. In the upper panel, the two saddle point configurations (**SP1** and **SP2**) and the inserted intermediate energy minimum (**I**) of the step-wise pathway are visualized.

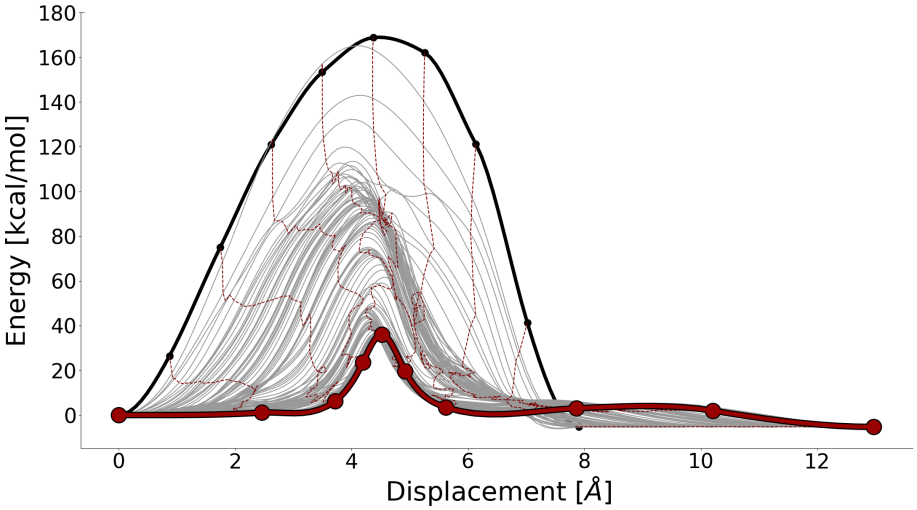


Figure 5: The optimization profile for a CI-NEB(N=10) calculation on the concerted hydrolysis of ethyl acetate. The black curve represents energy profile of the initial path. The minimum energy path is given by the red curve. Image position are denoted by circles.

However, for the hydrolysis reaction, there is an alternative and more favorable reaction pathway that connects the same given reactant and product states as the concerted mechanism. This alternative pathway proceeds in a step-wise fashion instead of concerted.[4] In order for CI-NEB to locate the SP for the step-wise reaction, an initial path local to the step-wise mechanism needs to be constructed. To obtain this path, the intermediate energy minimum along the step-wise mechanism is located *a priori* and inserted into the IDPP initial path generation. Then, with a suitable initial path, lying in vicinity to the MEP of the step-wise reaction, a CI-NEB calculation is started using $N = 18$. The reason for the increased N (compared to the concerted mechanism) is that the MEP for the step-wise mechanism is expected to be significantly longer. Hence, more intermediate images are required to resolve the path and the two barriers. The optimization profile for the step-wise mechanism is visualized in Fig. 6 and the configurations along the reaction coordinate in the upper panel of Fig. 4. The CI converges to the higher energy SP of the two and estimates the barrier height to be 32.2 kcal/mol, in agreement with the findings of Yamabe and coworkers.[4] Furthermore, the SP is confirmed to be of first order with an imaginary frequency of -1179 cm^{-1}.

This example shows that for systems with multiple MEPs, the choice of the initial path will determine to which MEP the calculation will converge. In particular, CI-NEB will

converge to the MEP lying closest to the given initial path. Also, this path is not necessarily the most energetically favorable MEP. Therefore, in mechanistic studies of complex reactions, alternative paths often have to to be explored with NEB, or variants thereof.

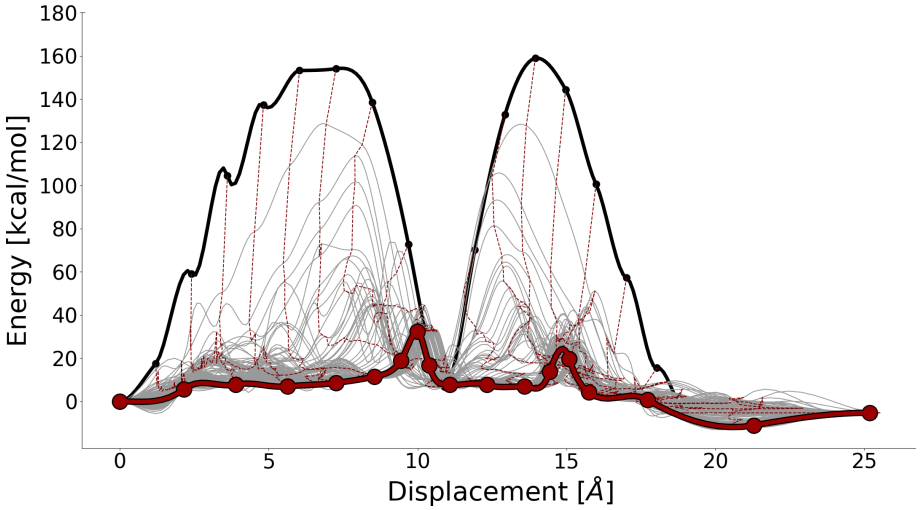


Figure 6: The optimization profile for a CI-NEB(N=18) calculation on the concerted hydrolysis of ethyl acetate. The black curve represents energy profile of the initial path. The minimum energy path is given by the red curve. Image position are denoted by circles.

SI-3 Rearrangement of 1,5-hexadiene

SI-3.1 System configurations

Reactant configuration:

C	-3.04347311082097	1.22446328604673	-0.06938989830847
C	-2.43415631414004	0.04415491858594	-0.21080080988922
H	-2.81058999834088	1.88987620626786	0.76909403563432
H	-3.79812662772203	1.56928213266814	-0.78225714132432
C	-1.38556185061648	-0.51534791817130	0.70829007519665
H	-2.69506857144252	-0.58707157530605	-1.07190091032989
C	-0.02962150947334	-0.75994399156094	0.00808235868131
H	-1.74075833950023	-1.48146736625732	1.11197540339779
H	-1.23963772371405	0.15877473815881	1.56881575380254
C	0.63631539271109	0.50308803695813	-0.45854284023481
H	-0.20183104654269	-1.42409011850252	-0.86092054108344
H	0.64231637223266	-1.30561763836492	0.69122174371637
C	1.84762449278914	0.91871328368038	-0.07860619134922
H	0.05346160738780	1.12275110316835	-1.15208227676047
H	2.46045868735481	0.33573922882586	0.61836595835187
H	2.27651853983777	1.85341567380283	-0.45117471950101

Saddle-point configuration:

C	-1.01343293590784	1.06307904091664	-0.64036370761165
C	-1.37259153113572	-0.27499582132995	-0.42620771059571
H	-1.03944832775078	1.75544370864743	0.20723624370911
H	-1.24714934790849	1.52982744696457	-1.60354490803281
C	-1.05335056219308	-0.92455193133719	0.77456347258365
H	-1.56419730325057	-0.89949254016543	-1.30855012902708
C	1.00735848077358	-1.06445995444015	0.63782529188753
H	-1.32117185245523	-1.97915185315702	0.89848185599093
H	-1.07549448035970	-0.34669767093096	1.70403801469849
C	1.37996988382316	0.26946883377680	0.42373112940174
H	1.00710094150537	-1.75301450801527	-0.21157971263068
H	1.25456644465003	-1.54014808883258	1.59227028509825

C	1.04830717642146	0.93171491165715	-0.76631922220994
H	1.61079026249630	0.88231512691534	1.30320803951315
H	1.05563763641820	0.36598114764846	-1.70193885427586
H	1.32310551487330	1.98468215168216	-0.88285008849915

Product configuration:

C	-1.46389288185209	1.16991521836219	-0.50426611910320
C	-2.20314035468432	-0.12225221642722	-0.30797588398807
H	-1.63292148423919	1.83197430948123	0.36278040364856
H	-1.87445880432901	1.69568063371530	-1.38602655116852
C	-2.90411484241656	-0.45611331366218	0.77819019379859
H	-2.13296911483792	-0.84057236757293	-1.13647447537690
C	1.45883505543428	-0.73624576693585	0.48407752181121
H	-3.41593870906557	-1.41928232609971	0.85765712521681
H	-2.99158645355137	0.22518582598140	1.63179573984434
C	0.75454144734777	0.39799304423885	0.47105616492082
H	1.57820125230637	-1.34554476461971	-0.41859855290233
H	1.94236501665236	-1.10040349201164	1.39500529570867
C	0.05251002165555	0.97716367655302	-0.72353071567133
H	0.65437433049865	0.97072901902244	1.40332590861694
H	0.22387372231808	0.33519084800645	-1.60516857408504
H	0.49220179876300	1.96328167196839	-0.96169748127056

SI-3.2 Energy profile obtained from CI-NEB of Z-NEB

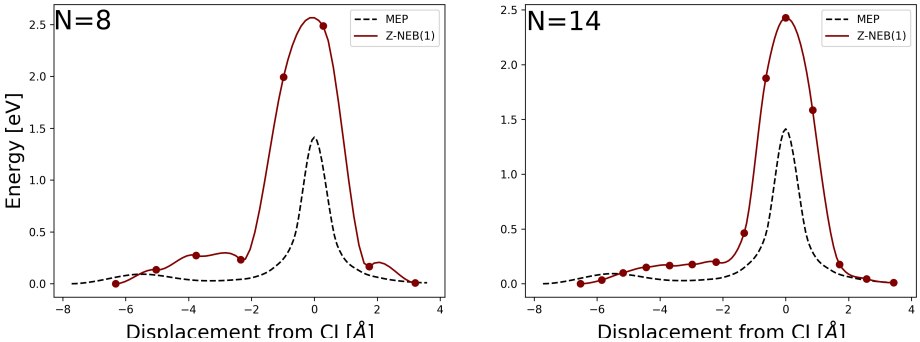


Figure 7: Energy profile (shown by a red curve) obtained from the converged initial phase of Z-NEB for the rearrangement reaction of 1,5-hexadiene. The calculations are carried out using $N = 8$ (left) and $N = 14$ (right). The MEP for the reaction is shown by a black dashed line.

SI-3.3 Computational efficiency as a function of N

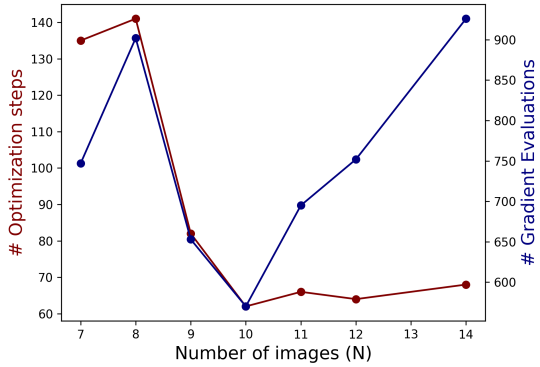


Figure 8: The computational efficiency of Z-NEB as a function of number of images. The Z-NEB calculations are carried out on the rearrangement reaction of 1,5-hexadiene. The computational efficiency is measured in the number of optimization steps (red) and by the number of energy and gradient evaluations (blue).

SI-4 Benchmark set: Pt heptamer island on Pt(111)

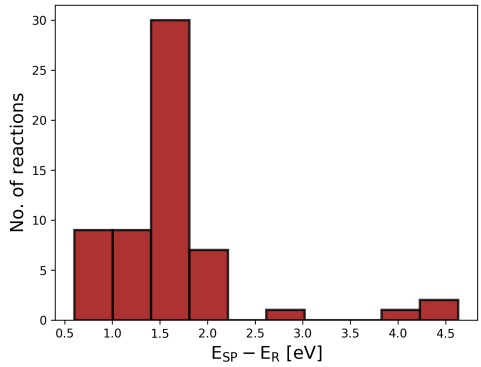


Figure 9: Distribution of potential energy barrier heights for the benchmark set of Pt-heptamer island diffusion on a Pt(111) surface.

The following computational settings are used for the Pt-heptamer island calculations:

- **CI-NEB**: 7 images (i.e., 5 intermediate images), initial path generated by IDPP. Climbing image is started at the beginning of the CI-NEB calculation, $k^{sp} = 5.0$ eV/Å^2
- **Z-NEB**: $\gamma = 0.2$, $\epsilon_{\text{zoom}} = 0.2$ eV/Å. Linear interpolation of images in the zoom-region. The end-point images are not restrained in ZCI-NEB. Other parameters are chosen to be the same as in CI-NEB.
- **AutoNEB**: The number of images to be propagated simultaneously is $N_{\text{simul}} = 5$. Total number of images is set to $N_{\text{max}} = 9$ or 12. A new image is introduced every $N_{\text{add}} = 7$ optimization steps. The geometric and energy resolution parameter is set to the default value of ASE-3.17, or $r_{\text{se}} = 0.5$.
- **L-BFGS**: $\Delta_{\text{step}} = 0.2$ Å. Memory is chosen to be 20. The memory is reset and a finite difference steepest-descent ($\Delta s = 0.001$Å) step is taken when the maximum component of a step exceeds Δ_{step}.
- **FIRE**: $\Delta_{\text{step}} = 0.2$ Å, $\Delta t = 0.1$ ($x10.18$) fs, $\alpha_{\text{start}} = 0.1$, $f^{\text{inc}} = 1.1$, $f^{\text{dec}} = 0.5$, $N = 5$, $f^{\alpha} = 0.99$, $\max(\Delta t) = 10\Delta t$.
- **Convergence crit. & thresholds**: $||\mathbf{F}_{\text{CI}}|| < 0.001$ eV/Å. Maximum allowed number of optimization steps is set to 1000.

SI-4.1 Data for CI-NEB calculations

Table 1: Results obtained by CI-NEB/LBFGS calculations on reactions 0–30 of the Pt-heptamer island diffusion benchmark set.

Reaction	Force calls	Opt. steps	E^{SP} (eV)
0	282	56	0.99
1	262	52	0.60
2	277	55	0.99
3	277	55	1.51
4	182	36	0.62
5	267	53	0.60
6	447	89	1.20
7	282	56	1.51
8	282	56	0.99
9	357	71	1.21
10	747	149	4.24
11	382	76	1.48
12	467	93	1.61
13	352	70	1.49
14	282	56	0.99
15	282	56	0.99
16	697	139	1.62
17	267	53	1.51
18	347	69	1.21
19	282	56	0.99
20	407	81	1.49
21	447	89	1.20
22	382	76	1.20
23	272	54	1.51
24	477	95	1.62
25	282	56	1.51
26	397	79	1.48
27	577	115	1.20
28	277	55	2.18
29	417	83	1.62
30	462	92	1.61

Table 2: Results obtained by CI-NEB/LBFGS calculations on reactions 31–58 of the Pt-heptamer island diffusion benchmark set.

Reaction	Force calls	Opt. steps	E^{SP} (eV)
31	577	115	1.62
32	217	43	4.10
33	267	53	1.51
34	382	76	1.48
35	407	81	1.48
36	517	103	1.62
37	422	84	1.62
38	382	76	1.48
39	402	80	1.63
40	382	76	1.49
41	397	79	1.49
42	262	52	2.18
43	262	52	2.18
44	932	186	2.87
45	677	135	4.63
46	397	79	1.20
47	447	89	1.49
48	402	80	1.62
49	322	64	2.16
50	437	87	1.49
51	452	90	1.20
52	387	77	1.48
53	452	90	1.21
54	262	52	2.18
55	252	50	2.18
56	317	63	2.16
57	407	81	1.62
58	402	80	1.63

Table 3: Results obtained by CI-NEB/FIRE calculations on reactions 0–30 of the Pt-heptamer island diffusion benchmark set.

Reaction	Force calls	Opt. steps	E^{SP} (eV)
0	592	118	0.99
1	677	135	0.60
2	592	118	0.99
3	567	113	1.51
4	412	82	0.62
5	732	146	0.60
6	757	151	1.20
7	567	113	1.51
8	592	118	0.99
9	747	149	1.21
10	847	169	4.24
11	847	169	1.48
12	2257	451	1.61
13	557	111	1.49
14	642	128	0.99
15	592	118	0.99
16	2012	402	1.62
17	567	113	1.51
18	747	149	1.21
19	592	118	0.99
20	567	113	1.49
21	757	151	1.20
22	742	148	1.20
23	592	118	1.51
24	2337	467	1.62
25	567	113	1.51
26	847	169	1.48
27	1342	268	1.20
28	562	112	2.18
29	797	159	1.62
30	2257	451	1.61

Table 4: Results obtained by CI-NEB/FIRE calculations on reactions 31–58 of the Pt-heptamer island diffusion benchmark set.

Reaction	Force calls	Opt. steps	E^{SP} (eV)
31	2012	402	1.62
32	652	130	4.10
33	582	116	1.51
34	857	171	1.48
35	857	171	1.48
36	2337	467	1.62
37	797	159	1.62
38	847	169	1.48
39	797	159	1.63
40	557	111	1.49
41	567	113	1.49
42	562	112	2.18
43	592	118	2.18
44	2122	424	2.87
45	1502	300	4.63
46	742	148	1.20
47	557	111	1.49
48	767	153	1.62
49	687	137	2.16
50	557	111	1.49
51	1342	268	1.20
52	947	189	1.48
53	1737	347	1.21
54	522	104	2.18
55	522	104	2.18
56	647	129	2.16
57	767	153	1.62
58	797	159	1.63

SI-4.1.1 CI-NEB/L-BFGS optimization profiles of selected reactions

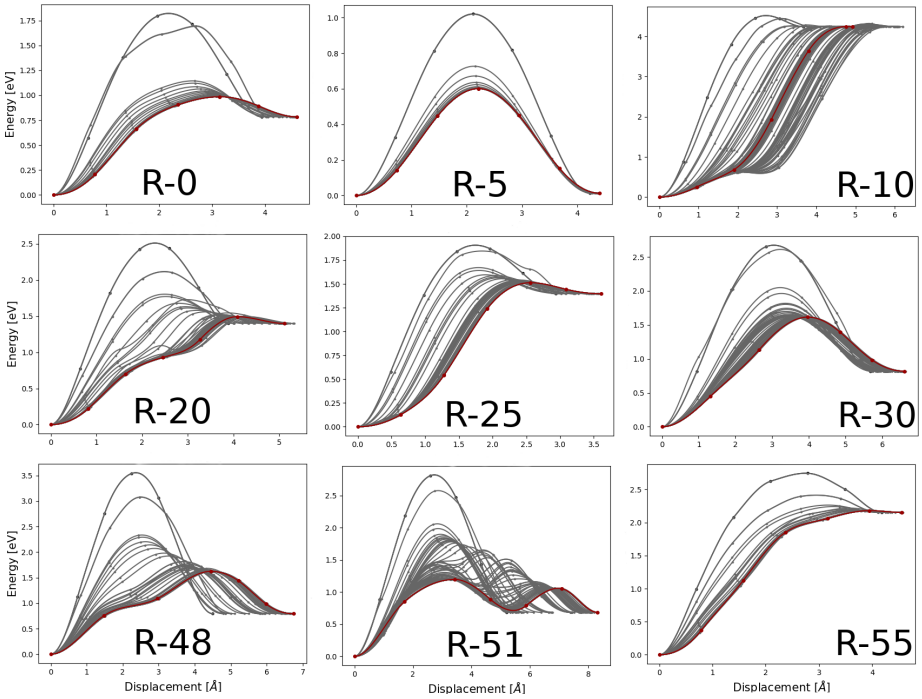


Figure 10: CI-NEB(N=7) optimization profiles for 9 selected reactions (numerated as 0, 5, 10, 20, 25, 30, 48, 51 and 55) of the Pt-heptamer island diffusion benchmark set. The gray energy profiles show how the MEP evolves as the CI-NEB optimization proceeds. The minimum energy path is given by the red curve.

SI-4.2 Data for Z-NEB calculations

Table 5: Results obtained by Z-NEB/L-BFGS calculations on reactions 0–30 of the Pt-heptamer island diffusion benchmark set.

Reaction	CI-NEB(1) FCs	CI-NEB(2) FCs	Total FCs	Opt. steps	E^{SP} (eV)
0	47	235	282	56	0.99
1	27	235	262	52	0.60
2	47	230	277	55	0.99
3	42	240	282	56	1.51
4	27	155	182	36	0.62
5	27	240	267	53	0.60
6	42	325	367	73	1.20
7	32	245	277	55	1.51
8	52	235	287	57	0.99
9	32	335	367	73	1.21
10	67	245	312	62	4.24
11	52	340	392	78	1.48
12	52	345	397	79	1.61
13	87	205	292	58	1.49
14	52	240	292	58	0.99
15	52	245	297	59	0.99
16	52	400	452	90	1.62
17	32	245	277	55	1.51
18	32	345	377	75	1.21
19	52	230	282	56	0.99
20	87	220	307	61	1.49
21	42	340	382	76	1.20
22	42	325	367	73	1.20
23	42	235	277	55	1.51
24	52	350	402	80	1.62
25	42	240	282	56	1.51
26	52	330	382	76	1.48
27	162	335	497	99	1.20
28	47	185	232	46	2.18
29	117	270	387	77	1.62
30	52	355	407	81	1.61

Table 6: Results obtained by Z-NEB/L-BFGS calculations on reactions 31–59 of the Pt-heptamer island diffusion benchmark set.

Reaction	CI-NEB(1) FCs	CI-NEB(2) FCs	Total FCs	Opt. steps	E^{SP} (eV)
31	52	375	427	85	1.62
32	32	190	222	44	4.10
33	42	240	282	56	1.51
34	52	350	402	80	1.48
35	52	330	382	76	1.48
36	52	360	412	82	1.62
37	102	300	402	80	1.62
38	52	330	382	76	1.48
39	117	270	387	77	1.63
40	87	200	287	57	1.49
41	87	215	302	60	1.49
42	47	185	232	46	2.18
43	47	210	257	51	2.18
44	27	820	847	169	2.87
45	52	400	452	90	4.63
46	42	330	372	74	1.20
47	87	215	302	60	1.49
48	117	275	392	78	1.62
49	107	250	357	71	2.16
50	87	200	287	57	1.49
51	192	370	562	112	1.20
52	52	340	392	78	1.48
53	187	275	462	92	1.21
54	47	185	232	46	2.18
55	47	185	232	46	2.18
56	117	215	332	66	2.16
57	117	270	387	77	1.62
58	117	280	397	79	1.63

Table 7: Results obtained by Z-NEB/FIRE calculations on reactions 0–30 of the Pt-heptamer island diffusion benchmark set.

Reaction	CI-NEB(1) FCs	CI-NEB(2) FCs	Total FCs	Opt. steps	E^{SP} (eV)
0	117	470	587	117	0.99
1	52	650	702	140	0.60
2	117	470	587	117	0.99
3	82	570	652	130	1.51
4	52	440	492	98	0.62
5	52	650	702	140	0.60
6	62	600	662	132	1.20
7	82	570	652	130	1.51
8	122	465	587	117	0.99
9	62	600	662	132	1.21
10	107	705	812	162	4.24
11	102	855	957	191	1.48
12	87	2075	2162	432	1.61
13	137	415	552	110	1.49
14	67	505	572	114	0.99
15	92	515	607	121	0.99
16	107	2060	2167	433	1.62
17	82	570	652	130	1.51
18	62	600	662	132	1.21
19	122	465	587	117	0.99
20	137	400	537	107	1.49
21	62	600	662	132	1.20
22	62	615	677	135	1.20
23	82	525	607	121	1.51
24	107	2060	2167	433	1.62
25	82	570	652	130	1.51
26	102	855	957	191	1.48
27	227	685	912	182	1.20
28	67	455	522	104	2.18
29	177	670	847	169	1.62
30	87	2075	2162	432	1.61

Table 8: Results obtained by Z-NEB/FIRE calculations on reactions 31–58 of the Pt-heptamer island diffusion benchmark set.

Reaction	CI-NEB(1) FCs	CI-NEB(2) FCs	Total FCs	Opt. steps	E^{SP} (eV)
31	107	2060	2167	433	1.62
32	97	380	477	95	4.10
33	82	525	607	121	1.51
34	102	890	992	198	1.48
35	102	850	952	190	1.48
36	107	2060	2167	433	1.62
37	177	670	847	169	1.62
38	102	850	952	190	1.48
39	157	660	817	163	1.63
40	137	375	512	102	1.49
41	137	400	537	107	1.49
42	67	455	522	104	2.18
43	67	450	517	103	2.18
44	57	1955	2012	402	2.87
45	97	1355	1452	290	4.63
46	62	615	677	135	1.20
47	137	415	552	110	1.49
48	177	670	847	169	1.62
49	167	440	607	121	2.16
50	137	375	512	102	1.49
51	227	685	912	182	1.20
52	102	890	992	198	1.48
53	277	700	977	195	1.21
54	67	455	522	104	2.18
55	67	455	522	104	2.18
56	167	480	647	129	2.16
57	177	670	847	169	1.62
58	157	660	817	163	1.63

SI-4.3 Data for Z-NEB calculations with fixed end-points

Table 9: Results obtained by Z-NEB/L-BFGS calculations with fixed end-points in ZCI-NEB. The calculations are carried out on reactions 0–30 of the Pt-heptamer island diffusion benchmark set.

Reaction	CI-NEB(1) FCs	CI-NEB(2) FCs	Total FCs	Opt. steps	E^{SP} (eV)
0	47	152	199	39	0.99
1	27	140	167	33	0.60
2	47	149	196	38	0.99
3	42	365	407	81	1.51
4	27	92	119	23	0.62
5	27	140	167	33	0.60
6	42	3002	3044	608	3.07
7	32	230	262	52	1.51
8	52	146	198	39	0.99
9	32	2258	2290	457	3.22
10	67	266	333	66	4.24
11	52	218	270	53	1.48
12	52	2159	2211	441	1.62
13	87	125	212	42	1.49
14	52	140	192	38	0.99
15	52	143	195	38	0.99
16	52	1118	1170	233	1.62
17	32	269	301	59	1.51
18	32	3002	3034	606	4.20
19	52	143	195	38	0.99
20	87	200	287	57	1.49
21	42	2651	2693	538	1.45
22	42	3002	3044	608	5.90
23	42	365	407	81	1.51
24	52	2483	2535	506	1.62
25	42	353	395	78	1.51
26	52	221	273	54	1.48
27	162	224	386	76	1.20
28	47	110	157	31	2.18
29	117	155	272	54	1.62
30	52	2039	2091	417	1.61

Table 10: Results obtained by Z-NEB/L-BFGS calculations with fixed end-points in ZCI-NEB. The calculations are carried out on reactions 31–58 of the Pt-heptamer island diffusion benchmark set.

Reaction	CI-NEB(1) FCs	CI-NEB(2) FCs	Total FCs	Opt. steps	E^{SP} (eV)
31	52	476	528	105	1.62
32	32	122	154	30	4.10
33	42	284	326	64	1.51
34	52	221	273	54	1.48
35	52	239	291	57	1.48
36	52	1433	1485	296	1.60
37	102	233	335	66	1.62
38	52	236	288	57	1.48
39	117	158	275	54	1.63
40	87	128	215	42	1.49
41	87	182	269	53	1.49
42	47	113	160	31	2.18
43	47	119	166	32	2.18
44	27	3002	3029	605	3.64
45	52	287	339	67	4.63
46	42	1163	1205	240	3.21
47	87	128	215	42	1.49
48	117	158	275	54	1.62
49	107	347	454	90	2.16
50	87	131	218	43	1.49
51	192	191	383	76	1.20
52	52	230	282	56	1.48
53	187	227	414	82	1.21
54	47	116	163	32	2.18
55	47	113	160	31	2.18
56	117	212	329	65	2.16
57	117	158	275	54	1.62
58	117	155	272	54	1.63

SI-4.4 Data for AutoNEB calculations (performed using ASE-3.17)

For AutoNEB calculations[5] presented in this section, the ASE-3.17 simulation package[6] was slightly modified, i.e., the convergence criterion of the optimizer class was changed from the absolute maximum force acting on an atom to the norm of the atom force acting on the highest energy image, i.e. the CI. Otherwise, calculations from the in-house software package and ASE-3.17 would not have been directly comparable.

SI-4.4.1 AutoNEB(5,9) calculations

Table 11: Results obtained by AutoNEB(5,9)/FIRE calculations on reactions 0–30 of the Pt-heptamer island diffusion benchmark set. The calculations are carried out using the slightly modified ASE-3.17. Two calculations were observed to raise an error and are represented by '-1' in all columns.

Reaction	Force calls	Opt. steps	E^{SP} (eV)
0	588	113	0.99
1	713	138	0.60
2	588	113	0.99
3	473	90	1.51
4	463	88	0.62
5	713	138	0.60
6	733	142	1.20
7	598	115	1.51
8	788	153	0.99
9	728	141	1.21
10	-1	-1	-1.00
11	903	176	1.55
12	2053	406	1.61
13	2786	1014	9.86
14	528	101	0.99
15	558	107	0.99
16	1933	382	1.62
17	628	121	1.51
18	1198	235	1.21
19	533	102	0.99
20	553	106	1.49
21	863	168	1.20
22	863	168	1.20
23	593	114	1.51
24	1938	383	1.62
25	473	90	1.51
26	818	159	1.48
27	1628	321	1.43
28	563	108	2.18
29	848	165	1.62
30	2053	406	1.61

Table 12: Results obtained by AutoNEB(5,9)/FIRE calculations on reactions 31–58 of the Pt-heptamer island diffusion benchmark set. The calculations are carried out using the slightly modified ASE-3.17. Two calculations were observed to raise an error and are represented by '-1' in all columns.

Reaction	Force calls	Opt. steps	E^{SP} (eV)
31	1888	373	1.62
32	-1	-1	-1.00
33	593	114	1.51
34	843	164	1.54
35	943	184	1.48
36	1823	360	1.62
37	848	165	1.62
38	788	153	1.48
39	803	156	1.70
40	638	123	1.49
41	648	125	1.49
42	623	120	2.18
43	608	117	2.18
44	1933	382	2.87
45	1423	280	4.63
46	718	139	1.20
47	763	148	1.49
48	818	159	1.62
49	588	113	2.16
50	673	130	1.49
51	968	189	1.20
52	923	180	1.55
53	4733	942	2.94
54	613	118	2.18
55	573	110	2.18
56	713	138	2.16
57	773	150	1.70
58	848	165	1.63

SI-4.4.2 AutoNEB(5,12) calculations

Table 13: Results obtained by AutoNEB(5,12)/FIRE calculations on reactions 0–30 of the Pt-heptamer island diffusion benchmark set. The calculations are carried out using the slightly modified ASE-3.17. Two calculations were observed to raise an error and are represented by '-1' in all columns.

Reaction	Force calls	Opt. steps	E^{SP} (eV)
0	639	119	0.99
1	689	129	0.60
2	639	119	0.99
3	689	129	1.51
4	494	90	0.62
5	689	129	0.60
6	744	140	1.20
7	649	121	1.51
8	639	119	0.99
9	779	147	1.21
10	-1	-1	-1.00
11	1009	193	1.48
12	2029	397	1.61
13	579	107	1.49
14	674	126	0.99
15	679	127	0.99
16	2199	431	1.62
17	719	135	1.51
18	924	176	1.21
19	699	131	0.99
20	579	107	1.49
21	744	140	1.20
22	814	154	1.20
23	689	129	1.51
24	2264	444	1.62
25	689	129	1.51
26	1069	205	1.48
27	5219	1035	1.64
28	564	104	2.18
29	1109	213	1.62
30	2029	397	1.61

Table 14: Results obtained by AutoNEB(5,12)/FIRE calculations on reactions 31–58 of the Pt-heptamer island diffusion benchmark set. The calculations are carried out using the slightly modified ASE-3.17. Two calculations were observed to raise an error and are represented by '-1' in all columns.

Reaction	Force calls	Opt. steps	E^{SP} (eV)
31	1944	380	1.62
32	-1	-1	-1.00
33	744	140	1.51
34	929	177	1.48
35	799	151	1.48
36	2044	400	1.62
37	1674	326	1.62
38	989	189	1.48
39	1079	207	1.70
40	579	107	1.49
41	579	107	1.49
42	624	116	2.18
43	639	119	2.18
44	2014	394	2.87
45	1474	286	4.63
46	734	138	1.20
47	579	107	1.49
48	1149	221	1.62
49	744	140	2.16
50	579	107	1.49
51	1024	196	1.20
52	929	177	1.48
53	5219	1035	7.94
54	564	104	2.18
55	564	104	2.18
56	1059	203	2.16
57	894	170	1.70
58	899	171	1.63

SI-5 Addressing the previously reported 'instability of CI-NEB' and the elastic band springs

In the present study, a CI-NEB calculation (using an unmodified version of ASE-3.17) is carried out on the O-adatom diffusion on Ag-decorated Pt(211) step edge, in an attempt to observe the long-detour reported by Bjørk Hammer and coworkers in the paper "An automated nudged elastic band method".[5] The calculation is carried out using $k^{\mathrm{sp}} = 1.0$ eV/Å^2 and $N = 10$ as in the original article. The CI-NEB optimization appears to be unstable. By examination of the calculation this behavior appears to be caused by large atom forces (because of overlapping of the adatom and the edge) along the initial. Here, the IDPP method was used with default settings used by ASE.

To fix this problem, the time-step of the dynamical optimization method for the initial IDPP calculation is reduced from the default 0.2 to 0.01 and the default convergence criterion lowered by an order of magnitude. This allows IDPP to generate a much more reasonable initial path, void of overlapping of atoms. Then, CI-NEB is able to converge without problems and the resulting path is reasonable and without any long detours, see Fig. 11. .

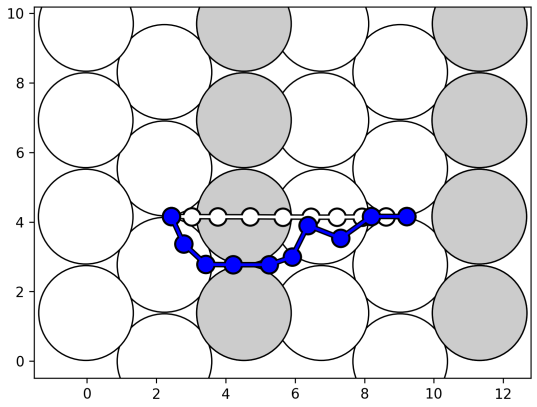


Figure 11: CI-NEB calculation of O-adatom diffusion on a Ag-decorated Pt(211) surface using CI-NEB. The initial path is denoted by white and the converged MEP by blue, respectively. The path obtained here is not in agreement to the path reported by Bjørk Hammer and coworkers and shows no sign of long detours. Note that in the CI-NEB calculation, the two upper-most layers are free to move. However, for clarity, the movement of surface atoms is omitted.

SI-5.1 The combination of CI and real-springs

In the paper by Bjørk Hammer and coworkers[5] a new formulation for the spring force is introduced, namely the elastic band. The full spring force is used in this formulation, i.e., the component of the spring force acting perpendicular to the path is included. The elastic band formulation is used with the CI variant of NEB. It is suggested that because of the use of CI, the perpendicular spring force will not affect the convergence of CI-EB calculations.[5]

However, it is important to keep in mind that even though the CI is not affected by the spring forces, the neighboring images of CI will be and are therefore subjected to corner-cutting[7] where they tend to get pushed off the MEP. The corner-cutting of the neighboring images will lead to an inaccurate estimate of the tangent at the CI and can possibly hinder convergence. To illustrate this concept, a CI-NEB calculation including the spring force acting perpendicular to the path is carried out on a simple two-dimensional PES, see Fig. 12. Here, the CI exhibits oscillating behavior and is unable to converge to the saddle point.

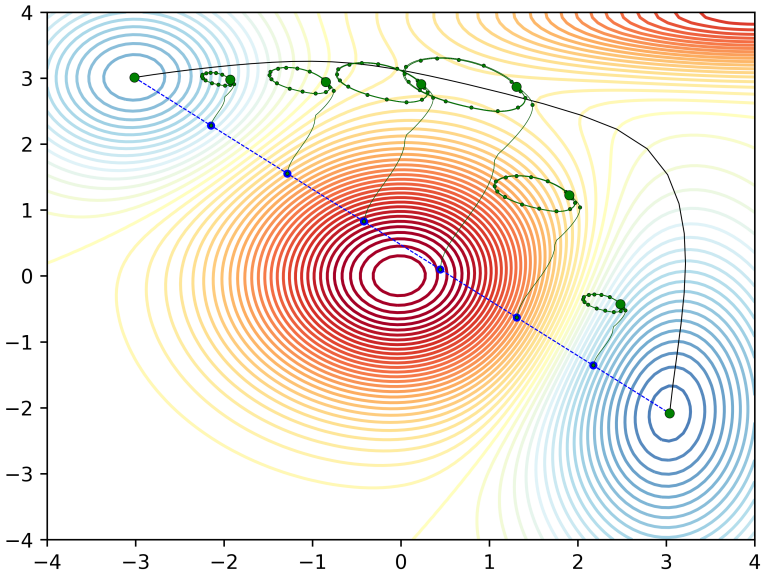


Figure 12: CI-NEB(N=8) calculation with a spring force acting perpendicular to the path on a two-dimensional PES. The calculation is started from a path constructed by linear interpolation (denoted by the blue line). The optimization path of the intermediate images is shown by a solid green line and at evenly spaced intervals the images are shown by green circles. The oscillating behaviour of the CI (and other images) is clear in this example.

References

(1) Ásgeirsson, V.; Arnaldsson, A.; Jónsson, H., Efficient evaluation of atom tunneling combined with electronic structure calculations. *J. Chem. Phys.*, **2018**, 148(10), 102334.

(2) Henkelman, G. and Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.*, **2000**, 113(22), 9978–9985.

(3) Smidstrup, S.; Pedersen, A.; Stokbro, K.; Jónsson, H. Improved initial guess for minimum energy path calculations. *J. Chem. Phys.*, **2014**, 140(21), 214106.

(4) Yamabe, S.; Tsuchida, N.; Hayashida, Y. Reaction Paths of the Water-Assisted Neutral Hydrolysis of Ethyl Acetate. *J. Phys. Chem. A*, **2005**, 109(32), 7216–7224.

(5) Kolsbjerg, E.L.; Groves, M.N.; Hammer, B. An automated nudged elastic band method. *J. Chem. Phys.*, **2016**, 145(9), 094107.

(6) Larsen, A.H.; Mortensen, J.J.; Blomqvist, J., Castelli, I.E.; Christensen, R.; Dulak, M.; Friis, J.; Groves, M.N.; Hammer, B.; Hargus, C. and Hermes, E.D. The atomic simulation environment – a Python library for working with atoms. *J. Phys. Condens. Matter*, **2017**, 29(27), 273002.

(7) Jónsson, H.; Mills, G.; Jacobsen, K. W. Nudged elastic band method for finding minimum energy paths of transitions. 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.

Article III

Nudged elastic band method for molecular reactions using energy-weighted springs combined with eigenvector following

Ásgeirsson, V., Birgirson, B.O., Bjornsson, R., Becker, U.,Neese, F., Riplinger, C. and Jónsson, H.

Accepted in Journal of Chemical Theory and Computation, 2021

Nudged elastic band method for molecular reactions using energy-weighted springs combined with eigenvector following

Vilhjálmur Ásgeirsson,[†] Benedikt Orri Birgisson,[†] Ragnar Bjornsson,[‡] Ute Becker,[¶] Frank Neese,[¶] Christoph Riplinger,[§] and Hannes Jónsson[*,†]

[†]*Science Institute and Faculty of Physical Sciences, University of Iceland VR-III, 107 Reykjavík, Iceland*

[‡]*Max-Planck-Institute for Chemical Energy Conversion, Mülheim an der Ruhr 45470, Germany*

[¶]*Max-Planck-Institute for Kohlenforschung, Mülheim an der Ruhr 45470, Germany*

[§]*FaccTs GmbH, Rolandstrasse 67, 50677 Cologne, Germany*

E-mail: hj@hi.is

Abstract

The climbing image nudged elastic band method (CI-NEB) is used to identify reaction coordinates and to find saddle points representing transition states of reactions. It can make efficient use of parallel computing as the calculations of the discretization points, the so-called images, can be carried out simultaneously. In typical implementations, the images are distributed evenly along the path by connecting adjacent images with equally stiff springs. But, for systems with high degree of flexibility, this can lead to poor resolution near the saddle point. By making the spring constants increase with energy, the resolution near the saddle point is improved. To assess the performance of this energy-weighted CI-NEB method, calculations are carried out for a benchmark set of 121 molecular reactions. The performance of the method is analyzed with respect to the input parameters. Energy-weighted springs are found to greatly improve performance and result in successful location of the saddle points in less than a thousand energy and force evaluations on average (about a hundred per image) using the same set of parameter values for all the reactions. Even better performance is obtained by stopping the calculation before full convergence and complete the saddle point search using an eigenvector following method starting from the location of the climbing image. This combination of methods, referred to as NEB-TS, turns out to be robust and highly efficient as it reduces the average number of energy and force evaluations down to a third, to 305. An efficient and flexible implementation of these methods has been made available in the ORCA software.

1 Introduction

Atomic scale calculations of chemical reactions typically involve identifying the transition mechanism and estimating the reaction rate. This is typically done by analyzing the relevant part of the potential energy surface (PES) characterizing the system where each state corresponds to a local minimum. The minimum energy path (MEP) connecting two local minima represents a natural choice for a reaction coordinate. At each point on the MEP, the energy is at a minimum with respect to all orthogonal directions and a maximum in the energy along an MEP corresponds to a first order saddle point (SP) on the energy surface. An estimate of the activation energy of a transition can, within the harmonic approximation to transition state theory,[1,2] be obtained from the (highest) energy maximum along the MEP, and the tangent to the path at that point characterizes the reaction mechanism as it shows the atomic displacements that represent the 'bottle-neck' for the transition. A local minimum on the energy surface can be found by a number of available minimization methods given that an initial point can be specified that is close enough. It is, however, significantly more challenging to locate a SP and the development of such methods remains an active research area. Methods for locating SPs can be divided into two categories: single ended and double ended methods. In the double ended methods, both the reactant state and the product state of the transition are specified. The task is then to use a discrete representation of a path and search for a reaction path (RP) that connects the two reactant and product state energy minima. In the single ended methods, only a single configuration of the atoms is specified, and a climb up the energy surface is carried out to converge on a nearby SP. The starting point of such calculations can be located anywhere on the energy surface. However, optimal performance and rapid convergence is only achieved if the starting point is close enough to the SP. The possibility exists, however, that the calculation identifies an SP that does not correspond to the transition of interest and several searches may need to be carried out with large accumulated computational effort. The single ended methods have the advantage that the dimensionality of the optimization problem is significantly smaller than that of the double-ended methods.

The nudged elastic band (NEB) method[3,4] belongs to the first category and is frequently used in studies of reactions and diffusion events in and on the surfaces of solids,[5] but has also been extended for calculations of, e.g. magnetic transitions,[6] paths on free-energy surfaces[7] and tunneling paths.[8] In an NEB calculation, the path between the reactant and product state minima is discretized by a set of replicas of the system, referred to as images. Given some initial path, an iterative optimization algorithm is used to bring the intermediate images to the nearest RP. Typically, the initial path for an NEB calculation is generated using a

linear interpolation in Cartesian coordinates between the endpoint images, but this can lead to the placement of an image in a high energy region where atoms are too close to each other, causing problems for electronic structure calculations. One way to avoid this is to define some minimum distance between atoms in the interpolated images. A better method for generating an initial path involves interpolation of distances between pairs of atoms and then generating a path that matches these distances as closely as possible in a least squares sense.[9] This method is referred to as image-dependent pair potential (IDPP) method and it requires only minimal computational effort as it does not involve any evaluation of the energy or force of the system. Typically, the IDPP yields an initial path that lies closer to the RP than a linear interpolation in Cartesian coordinates, and the electronic structure calculations of the images then require fewer iterations to converge on the RP. Alternatively to IDPP, it is possible to construct the initial path by linear interpolation in internal coordinates[10] or by using a geodesic interpolation between the reactant and product states.[11] Peters *et al.*[12] approached the problem of entering high energy regions in a different way and used a method where the path connecting the reactant and product states is grown by adding images sequentially. In this method, referred to as the growing string (GS) method, the images are added one after another starting from each of the two endpoints and electronic structure calculations carried out to optimize their position. The two path fragments grow until they connect. However, because of the sequential addition of images, the advantage of parallel computing in NEB calculations is lost.

In NEB calculations, only the components of the atomic forces acting perpendicular to the path modify the shape and location of the path. A projection of the atomic forces is, therefore, required based on an estimate of the local tangent to the path, so an accurate approximation of the tangent plays a crucial part in NEB calculations.[13] The distribution of images along the path is controlled using harmonic springs between adjacent images, but only the component of the spring force acting tangentially to the path is included. Without such control of the distribution, the images tend to slide down to the local energy minima at the fixed endpoints.[4] In commonly used implementations of the NEB method, the spring constant for all the springs is chosen to be the same, leading to an even distribution of the images along the path. But, different spring constants can be used for different pairs of images to yield any desired distribution. In order to obtain an accurate estimate of the SP, one of the images, typically the highest energy image, is forced to 'climb' upwards along the path to converge on the point of highest energy. This is referred to as 'climbing image' NEB (CI-NEB). The accuracy of the tangent estimate at the climbing image is a particularly important aspect of CI-NEB calculations.

If the system has large flexibility, i.e. degrees of freedom for which the energy does not

change much, and if the part of the RP that corresponds to the energy barrier is correspondingly small, then the resolution of the path near the climbing image may be too small for the tangent estimate to be accurate enough. The RPs for gas-phase molecular reactions are often found to have long tails corresponding, for example, to rotation and/or translation of molecules or molecular fragments. There have been numerous variants of the NEB method that attempt to resolve this issue and improve efficiency by focusing the computational effort on the barrier region.[14–18] This is often accomplished by adding new intermediate images in the barrier region or moving images from a tail region of the path to the barrier region.

A simple way to achieve improved resolution in the barrier region is to make the springs stiffer in regions of higher energy, i.e. energy-weighted (EW) springs, resulting in higher density of images near the SP. Such a procedure was already proposed some time ago,[19] but has not become commonly used. This method is applied here and its efficiency compared with the standard, equal spring implementation in calculations of various molecular reactions.

Single ended methods for locating SPs, where the product of the transition is not specified, include eigenvector-following (EF)[20–24] and minimum mode following (MMF)[25–29] methods. In these methods, a particular eigenvector (or a mode) of the Hessian matrix is selected to represent the reaction coordinate and the system is pushed along this mode uphill in energy, while the energy is minimized along all orthogonal directions. Therefore, the single ended methods can be started from almost anywhere on the energy surface, in the convex region near an energy minimum or even in a region with multiple negative eigenvalues. Typically, the eigenvector corresponding to the lowest eigenvalue, referred to as the minimum mode, is selected as the reaction coordinate. However, in cases where the minimum mode does not correspond to the correct reaction coordinate, e.g. in regions on the energy surface far from the SP, a different eigenvector may need to be selected and followed. Otherwise, the EF calculation may converge to an undesired SP. After finding an SP, the corresponding reactant and product states can be found by sliding down the steepest-descent path to the local minimum on either side of the SP.

The modified Newton-Raphson[20] and EF[22] partitioned rational function optimization (P-RFO)[21] methods require the use of the full Hessian matrix. However, the evaluation and diagonalization of the Hessian matrix is computationally demanding and can even become computationally prohibitive for large systems, especially if analytical second derivatives of the energy are not available. In practical implementations of the EF methods, the Hessian matrix is often computed only at the beginning of a calculation or at a fixed interval of optimization steps. For steps where the Hessian matrix is not evaluated, Hessian update methods[24,30] are used to approximate it. In this case, only a single energy and gradient evaluation is required in each iteration. In MMF calculations, however, the full evaluation

and storage of the Hessian matrix is avoided. Instead, only the minimum mode is computed by e.g. the dimer method.[23,25,26] The computation of the minimum mode typically requires a few energy and gradient evaluations per displacement of the system. This renders the EF P-RFO method, where the Hessian matrix is approximated, more efficient than the MMF methods when used on energy surfaces described by computationally demanding electronic structure methods. However, the MMF method may prove to be more robust than EF P-RFO since the minimum mode of the Hessian matrix is always computed explicitly.[26]

Saddle point searches using the EF P-RFO or MMF methods can typically be initiated from any point on the PES but if the starting point is not close enough to the SP of interest, the calculation may perform poorly or converge on a SP characterizing a different reaction than the reaction of interest. On the other hand, if EF or MMF calculations are started from a point on the surface that lies within the basin of attraction of the desired SP, then rapid convergence to the desired SP can be achieved. In fact, with an accurate Hessian estimate and a starting configuration in close proximity to the SP, EF P-RFO is quadratically convergent.

Considerable human (and computational) effort is often required to generate an initial configuration that is close enough to the desired SP and this becomes more challenging the more complex the transition is. Various interpolation methods[31,32] or relaxed surface scans (coordinate-driving) are often used to generate an initial configuration for a SP search.[24] Another consideration for EF is that important features of the reaction path may be missed, such as possible intermediate states and other higher energy SP that may lie between the reactant and product states.[33] It is, therefore, advantageous to combine the double and single ended methods in a complementary manner,[12,13] where the double ended methods are used to find an approximate RP that identifies the important features of the MEP and provides a good initial guess configuration (and reaction coordinate) for the relevant, highest energy, first order SP characterizing the transition of interest. Then, the SP can be located more accurately using the computationally efficient EF P-RFO or MMF method.

Heyden et al.[26] used the GS method to generate reasonable initial guess configurations for subsequent MMF and also EF P-RFO calculations and compared the efficiency of the two single ended methods using a medium-sized benchmark set of 25 molecular reactions. In the MMF calculations, the improved dimer method[26] was used to compute the minimum mode. Heyden and co-workers found that the EF P-RFO method needs on average 25-30% fewer optimization cycles than the improved dimer method, while the latter was shown to be more robust. More recently, Zimmerman devised an integrated scheme of GS and EF where evenly spaced intermediate images are introduced sequentially and optimized with a large tolerance, starting from each end point until the two segments connect.[34] The highest energy image on this path is forced to climb up towards the SP to provide an initial guess for

a subsequent EF optimization. This GS-CI-EF method has been applied to a benchmark set of 105 main-group molecular reactions and was found to be both reliable and efficient.[34,35] But, because of the sequential introduction of the images, it does not allow for optimal use of parallel computing and it is not easy to enhance the distribution of images according to the energy (as in EW-NEB) during the growth-phase of the GS.

In the present study, the performance of the EW-CI-NEB method is assessed and compared with the commonly used equal spring version of CI-NEB. Moreover, to swiftly and accurately locate first order saddle points connecting a given reactant and product states, an integrated scheme of CI-NEB and EF, referred to as NEB-TS, is presented. In NEB-TS, the EW-CI-NEB method is started from an IDPP path and parallel calculations of the images are used to converge on an approximate RP that connects the given reactant and products. Then, the location of the CI and the associated path tangent are used to initiate an EF calculation. A benchmark set of 121 main-group molecular reactions, adopted from Refs. 34 and 36 is used to evaluate the performance of the CI-NEB method and variants thereof. Furthermore, SPs are found for three challenging reactions taken from recent bioinorganic chemistry and heterogeneous catalysis studies to illustrate the applicability of the NEB-TS method to larger and more complex systems, where locating the desired SP was found to be particularly challenging.[37–39]

The article is organized as follows: In section 2, a description of the EW-CI-NEB and NEB-TS methods is given. Then, in section 3, the benchmark set of main-group molecular reactions is introduced. In section 4, the results of calculations on the benchmark and the three bioinorganic and heterogeneous catalysis reactions are presented. Finally, the findings are summarized in section 5.

2 Methodology

A concise description of the CI-NEB method[3,4] is given in section 2.1. For a more detailed review of CI-NEB, the reader is referred to Ref. 5. In Section 2.2, the new integrated scheme of CI-NEB and EF, referred to as NEB-TS, is presented. The computational settings of the CI-NEB and NEB-TS calculations are given in section 2.3.

2.1 Nudged elastic band method

The objective of NEB is to locate an RP connecting two local energy minima on the potential energy surface (PES) representing the reactant and product states. A discrete representation of path is used by generating $N_{\mathrm{im}} - 2$ intermediate images (i.e. molecular configurations) in between the reactant and product configurations. As for any numerical method based on discretization of continuous variables, convergence to the path is reached when $N_{\mathrm{im}} - 2$ is large enough. An initial path is first generated using some interpolation scheme and then an iterative algorithm is used to shift the intermediate images to the RP. An effective NEB force acting on the atoms is used in combination with a gradient-based optimization method, while keeping the end-point configurations fixed. The effective force is the sum of contributions from the component of the atomic force ($\mathbf{F}$) that is perpendicular to the path, ($\mathbf{F}^{\perp}$), and a spring force, ($\mathbf{F}^{\mathrm{sp}}$), that only acts along the path. The perpendicular force is

$$\mathbf{F}_i^{\perp} = \mathbf{F}_i - (\mathbf{F}_i \cdot \hat{\tau}_i)\hat{\tau}_i \tag{1}$$

where $\hat{\tau}_i$ is the local unit tangent to the path at image i. The 'upwind tangent' scheme is used,[13] where the unit tangent is taken as the normalized line segment between an image and its higher energy neighboring image. If the image is at an energy extremum along the path, the unit tangent is the normalized energy weighted average of the two line segments connecting the image to its two neighbors. The spring force acting tangential to the path is used to control the distribution of images along the path. The spring force is

$$\mathbf{F}_i^{\mathrm{sp},\parallel} = (k_i^{\mathrm{sp}}|\mathbf{R}_{i+1} - \mathbf{R}_i| - k_{i-1}^{\mathrm{sp}}|\mathbf{R}_i - \mathbf{R}_{i-1}|)\hat{\tau}_i \tag{2}$$

where $\mathbf{R}_i$ is the 3M-dimensional vector of atom coordinates for image i. Segment i of the path is characterized by $|\mathbf{R}_{i+1} - \mathbf{R_i}|$. The stiffness of the spring interaction for this segment of the path is given by the spring constant, k_i^{sp}.

The highest energy point along an RP gives an estimate for the activation energy for the transition. This point is a first order SP on the energy surface and characterizes the

bottleneck of the transition between the reactant and product states. To locate this point more accurately in NEB calculations, the highest energy image is treated differently.[19] It is pushed uphill on the energy surface according to the component of the atom force acting parallel to the path and shifted downhill by the force acting perpendicular to the path. This image is referred to as the climbing image (CI) and the method as climbing image NEB (CI-NEB). The spring forces do not apply to the CI. In practice, it may be beneficial to initiate the CI from a loosely converged NEB calculation. In other words, convert the highest energy image to CI when a given criterion, e.g. $\max(|\mathbf{F}_{\mathrm{HEI}}^{\perp}|) < \epsilon_{\mathrm{CI}}$ is met. This can also be done at the start of the NEB calculation. The CI is allowed to shift from one image to another during the optimization, e.g. if a higher energy image than the current CI emerges. This ensures that the CI will converge to the highest energy maximum along the path. Since the RP is primarily needed to obtain a good tangent estimate for the CI, significantly tighter convergence thresholds are typically applied to the CI than to the other images.[5]

Any desired distribution of images along the path can be obtained through the choice of the spring constants. In typical CI-NEB calculations, the spring constant is chosen to have the same value for all segments of the path, leading to an even distribution of images along the path. But, since it is most important to have good resolution of the path near the point of maximum energy, the highest SP on the path, one can choose the spring constants to be larger where the energy is high. Such a scheme of energy-weighted NEB (EW-NEB) springs has been proposed,[19] where the spring constant for segment i is given by

$$k_i^{\mathrm{sp}} = \begin{cases} (1-\alpha_i)k_{\mathrm{u}} + \alpha_i k_{\mathrm{l}}, & \text{if } E_i > E_{\mathrm{ref}} \\ k_l, & \text{otherwise} \end{cases} \tag{3}$$

$$\alpha_i = \frac{E_{\mathrm{max}} - E_i}{E_{\mathrm{max}} - E_{\mathrm{ref}}} \tag{4}$$

k_{u} and k_{l} are upper- and lower-bound values for the spring constant. E_{max} is the current estimate of the maximum energy along the path, E_i is the higher energy image of the pair of images connected by line segment i and E_{ref} is a reference energy chosen to be equal to the energy of either the reactant or product. Essentially, the spring constant is scaled in such a way that stiffer springs act between images located in higher energy regions, leading to higher density of images near the SP. At the same time, other less important parts of the RP become less well resolved. To illustrate the different image distributions obtained by the EW-CI-NEB and CI-NEB methods, calculations using the two methods are carried out on the ene-reaction of 1-propylene and ethylene. Results of the calculations are shown in Fig. 1. In EW-CI-NEB, most of the intermediate images are placed in the region at and around the energy maximum, thus greatly improving the resolution in this important

region. In CI-NEB, the images are equidistant and half of the images are utilized to resolve the chemically irrelevant potential energy tail of the RP corresponding to the rotation and translation of the ethylene fragment. The evolution of the inter-image distances during the two calculations are shown in section SI-2.3. To begin with, all of the images are placed in an equidistant manner along the initial path. In CI-NEB, the inter-image distance then slightly increase as the path lengthens during the optimization. The inter-image distance between different pairs of images can vary slightly in intermediate steps of the optimization. In EW-CI-NEB, the distance between a pair of images in the higher energy regions of the RP gradually becomes shorter, while the distance between images in the lower energy regions become longer. Hence, the final path obtained by the CI-NEB and EW-CI-NEB methods are approximately the same but with different image distributions. The different image distribution of EW-CI-NEB provides a more accurate estimate of the path tangent near the SP while not using an excessive number of images to represent the path and thereby prevents the computational cost of the calculation from becoming prohibitively large.

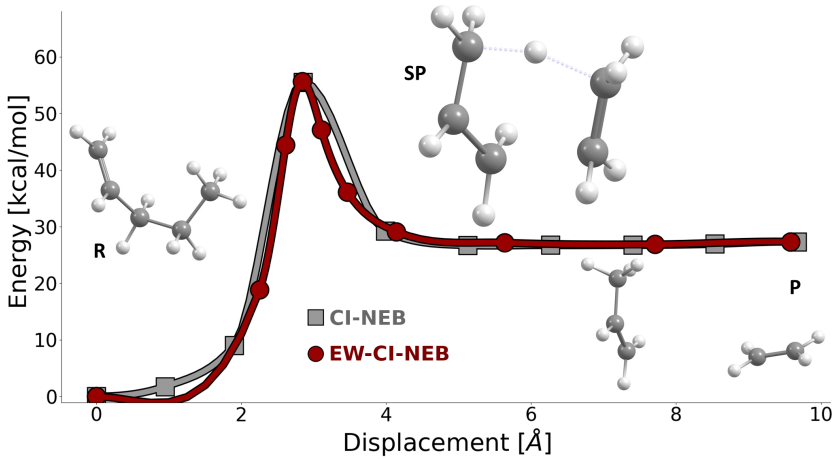


Figure 1: Comparison of CI-NEB and EW-CI-NEB calculations of the ene-reaction of 1-propylene and ethylene. The image position along the final converged reaction path is given by gray boxes for CI-NEB and red circles for EW-CI-NEB. The solid curves represent piecewise cubic interpolation between the images using the tangential atom force and energy. In the case of CI-NEB, the images are distributed evenly along the path, while in EW-CI-NEB the density of images is higher where the energy is higher which greatly improves the resolution of the path around the energy maximum. In the CI-NEB calculation a fixed spring constant of $k^{\mathrm{sp}} = 0.1\,\mathrm{E_H/a_0^2}$ is used, while an energy dependent spring constant scaled from 0.01 to $0.1\,\mathrm{E_H/a_0^2}$ is used in the EW-CI-NEB calculation. The reactant (denoted by **R**), saddle point (denoted by **SP**) and product (denoted by **P**) configurations are shown as insets.

An initial path connecting the given reactant and product configurations is needed to start a CI-NEB calculation. It is typically generated by linear interpolation in Cartesian coordinates or by using the IDPP method. Then, the intermediate images are iteratively shifted towards the RP using the effective CI-NEB force and a gradient-based optimization method. Often the velocity projection optimization (VPO) method[4] or the more elaborate FIRE version of that approach is used.[40]

More efficient optimization is provided by the limited memory BFGS (L-BFGS) method,[41] it is a quasi-Newton method where an approximation to the inverse Hessian matrix is constructed at each optimization step based on a memory of the energy and gradient obtained in the previous N steps. The product of the inverse Hessian matrix at the current position and the effective NEB force is then used to determine both the length and the direction of the

next displacement of the atoms. In NEB/L-BFGS, the path is treated as a single $3M(N_{\text{im}}-2)$ dimensional point, instead of treating each movable image individually (i.e. $N_{\text{im}}-2$ sets of $3M$ coordinates). This has been reported to stabilize NEB/L-BFGS calculation.[42] With a global definition of the Hessian matrix, the off-diagonal inter-image interactions get included in the Hessian matrix. This, however, also makes the L-BFGS optimization more sensitive to the choice of spring constants.

If the largest absolute component of an optimization step exceeds a user-defined constant trust-radius, Δ_{step}, the step is scaled in such a way that this component of the step equals Δ_{step}. This serves to prevent the optimization from taking unreasonably large steps. Moreover, if the step is scaled back, the memory of the L-BFGS is erased, which may serve to fix the Hessian matrix if it has reached an ill-conditioned state.[43]

2.2 NEB-TS method

The NEB-TS method automatically combines EW-CI-NEB with an EF optimization (or 'TS search') in a simple manner to efficiently locate an SP for a transition between a given reactant and product energy minima. An illustration of an NEB-TS calculation for the two-dimensional Müller-Brown energy surface[44] is shown in Fig. 2.

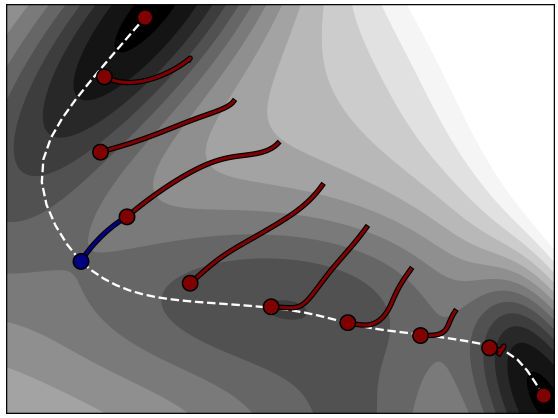


Figure 2: An illustration of the NEB-TS method on the two-dimensional Müller-Brown energy surface. The calculation starts from a linear interpolation using $N_{\text{im}} = 10$. The red curves show how each intermediate image moves during the NEB path optimization which is terminated at positions marked with filled circles, when $\max(|\mathbf{F}_{\text{CI}}|) \leq \epsilon_{\text{max}}^{\text{TS}}$. The subsequent TS search starts from the climbing image and is shown by the blue curve. The minimum energy path is shown by a dashed white line. The NEB-TS method accurately locates the higher energy saddle point along the minimum energy path.

In NEB-TS, the EW-CI-NEB method is used to get close to an RP in order to generate a good initial guess for the TS search. The EW-CI-NEB calculation is carried out until the atom force acting on the CI drops below a user-defined threshold; $\max(|\mathbf{F}|) < \epsilon_{\max}^{\mathrm{TS}}$ and $\mathrm{RMS}(\mathbf{F}) < \epsilon_{\mathrm{RMS}}^{\mathrm{TS}}$. The TS search is then automatically started from the location of the CI. The search is carried out using the EF partitioned rational function optimization (P-RFO) method[21,22] in redundant internal coordinates, as already implemented in the ORCA software.[45,46] In EF P-RFO, an eigenvector (or a mode) of the initial Hessian matrix is selected and used to represent the desired reaction coordinate. This mode is then followed uphill on the energy surface to a first order SP. While, the energy is minimized along all other modes. To achieve this, two shift parameters are used to ensure that only the eigenvalue of the Hessian matrix corresponding to the selected reaction coordinate is negative and that all other eigenvalues are positive.

The initial Hessian matrix in a TS search needs to be of correct form and with a mode to be followed that aligns with the reaction coordinate (i.e. the unstable mode) corresponding to the desired SP. Therefore, to obtain an accurate reaction coordinate, the initial Hessian matrix is most often computed analytically or numerically at the beginning of a EF P-RFO calculation. In NEB-TS, the initial Hessian matrix used by the TS search can either be taken as the exact Hessian matrix (e.g. when the analytical Hessian is available and readily computed) or by constructing a model Hessian using the empirical Almlöf method.[47] The eigenvector selected as the reaction coordinate is obtained by computing the maximum overlap with the path tangent at the CI of the EW-CI-NEB calculation. The tangent at the CI already provides an excellent approximation to the unstable mode at the SP. Also, if the Almlöf scheme is used, then the model Hessian needs to be modified accordingly. The curvature of the selected mode is shifted to correspond to the curvature of the path estimated from the two neighboring images using a finite difference approximation. The Hessian matrix is then updated after each displacement of the system using the Bofill update method.[30] Therefore, in EF PRFO, only a single energy and force evaluation is needed for each displacement of the system.

For comparison, a simpler method is also proposed where the NEB calculation is skipped. In this method, the IDPP interpolation method is used to construct an initial path that connects the reactant and product energy minima configurations, the energy and force is evaluated along this path and the highest energy image is chosen to serve as the initial guess configuration for the TS search. Therefore, only N_{im} energy and force calculations are carried out prior to the TS search calculation. This method is similar to the frequently used combination of interpolation methods and TS searches.[31] The less accurate initial configuration supplied by the IDPP interpolation method, compared to EW-CI-NEB, and the selection

of a reaction coordinate from the IDPP tangent may lead to inadequate results for some systems, especially when dealing with more complex reaction coordinates.

2.3 Computational settings

2.3.1 CI-NEB and NEB-TS calculations

All NEB calculations in this study are carried out using the NEB module recently implemented in the ORCA (4.2) suite of programs[45,46] with the exception of the three bioinorganic and heterogeneous catalysis reactions. In those calculations the development version of the ORCA software was used. The reactant and product configurations used as input for the calculations have been tightly optimized to a local energy minimum, i.e. with $\max(|\mathbf{F}|) < 1{\cdot}10^{-4}$ and $\mathrm{RMS}(\mathbf{F}) < 3 \cdot 10^{-5}$ $\mathrm{E_H/a_0}$. The center-of-geometry for the reactant and product configurations is then automatically aligned and the root-mean-square deviation between the two configurations minimized. This is necessary to remove rotational and translational degrees of freedom from the reaction coordinate and is found to greatly impact the success and efficiency of the CI-NEB calculations, see section SI-4.4. Furthermore, during the NEB optimization, the root-mean-square deviation between adjacent images along the path is minimized, starting from the reactant state.[48,49] All initial paths for the NEB calculations are constructed using the IDPP method. A comparison between the IDPP method and linear interpolation in Cartesian coordinates reveals that IDPP yields better initial paths with lower energy configurations, see section SI-2.1.

A variable number of images is tested and used to represent the path, where the number of images ranges from $N_{\mathrm{im}} = 7$ to 34. The CI variant of NEB is used in all calculations. It is found that it is optimal to activate the CI when $\max(|\mathbf{F}_{\mathrm{HEI}}|) < 2 \cdot 10^{-2}$ $\mathrm{E_H/a_0}$, see section SI-4.5. Both CI-NEB and EW-CI-NEB calculations are carried out using fixed and energy-weighted spring constants, respectively. For the fixed spring constants, both $k^{\mathrm{sp}} = 0.01$ and 0.1 $\mathrm{E_H/a_0^2}$ are used and compared to the EW springs, where the lower- and upper-boundaries for the spring constant are selected to be $k_l^{\mathrm{sp}} = 0.01$ and $k_u^{\mathrm{sp}} = 0.1$ $\mathrm{E_H/a_0^2}$. The choice of the spring constants is found to greatly affect the performance of the CI-NEB and EW-CI-NEB methods using a small set of molecular reactions, see section SI-4.1. The reference energy value for the energy weighting, E_{ref}, is always chosen to be equal to the energy of the higher energy reactant or product energy minima.

The L-BFGS method is found to be significantly more efficient than VPO (see Section SI-4.3) and is used in all (EW-)CI-NEB and NEB-TS calculations with a memory of 20 steps.[43] The constant trust-radius, Δ_{step}, is chosen to be 0.2 $\mathrm{a_0}$ which is found to be an excellent choice for the interplay of robustness and efficiency, see section SI-4.7. The step

size of the initial finite difference step in is chosen to be 0.002 a_0. For the TS search of NEB-TS a trust-radius of 0.1 a_0 is used.

2.3.2 Convergence criteria and thresholds

The CI-NEB and NEB-TS calculations are considered to be converged when $\max(|\mathbf{F}_{\mathrm{CI/TS}}|) < 5 \cdot 10^{-4}$ and $\mathrm{RMS}(\mathbf{F}_{\mathrm{CI/TS}}) < 3 \cdot 10^{-4}$ E_H/a_0. Also, for the TS search, there are three additional convergence criteria and thresholds gauged (as implemented in ORCA), namely the energy difference and the maximum and root-mean-square change in step size (ΔS) between two subsequent optimization steps, $\Delta E < 1 \cdot 10^{-5}$ E_H, $\max(\Delta S) < 4 \cdot 10^{-3}$ and $\mathrm{RMS}(\Delta S) < 2 \cdot 10^{-4}$ a_0, respectively. In all calculations, the maximum allowed number of total optimization steps is 500. If a calculation is unable to reach the aforementioned convergence thresholds, within the allowed number of optimization steps, it is defined as failed or a non-convergent calculation.

In NEB-TS, the initial EW-CI-NEB calculation is halted when the absolute value of the atom force acting on the CI drops below the TS activation thresholds, $\epsilon^{TS}_{\mathrm{max}}$ and $\epsilon^{TS}_{\mathrm{RMS}}$. Two sets of threshold values are used in the calculations of the benchmark set of reactions, $\epsilon^{TS}_{\mathrm{max}} = 0.01$ and 0.002 E_H/a_0. In all cases, $\epsilon^{TS}_{\mathrm{RMS}} = 0.5\epsilon^{TS}_{\mathrm{max}}$. As expected, preliminary results of NEB-TS calculations on a small test-set of molecular reactions shows that it is likely to be more efficient to activate the TS search already at $\epsilon^{TS}_{\mathrm{max}} = 0.01$ E_H/a_0, see section SI-4.6.

To assess the chosen convergence criteria, exact Hessian calculations were carried out on SPs obtained by EW-CI-NEB calculations (with $N_{\mathrm{im}} = 10$) on the benchmark set of main-group molecular reactions (see section 3). About 80% of the configurations are confirmed to be first order SPs, i.e. the exact Hessian matrix yields a single negative eigenvalue. About 17% of the SPs found in this way are confirmed to be of second order with two negative eigenvalues. For the second order SPs, one of the negative eigenvalues is typically found to be dominant, where the other eigenvalue has a significantly smaller magnitude. This is typical for molecular reactions, where the mode corresponding to the second negative eigenvalue represents motion along a soft degree of freedom. For the remaining systems, three negative eigenvalues are obtained. For the second and higher order SPs, a first order SP can typically be obtained by further EW-CI-NEB optimization. In these cases, it often sufficed to simply converge the other intermediate images to the same degree as the CI to obtain a first order SP. However, for a few of the more difficult saddle points, additional measures had to be taken, e.g. using stricter convergence criteria. The additional cost required to refine the saddle points is not taken into account in the performance numbers reported in Sec. 4.1. To construct a reference set of first order saddle points, the set of first order saddle points obtained from EW-CI-NEB is further optimized using EF P-RFO

with analytical Hessian matrix computations at the beginning and in every five steps, until $\max(|\mathbf{F}_{\mathrm{TS}}|) < 2.5 \cdot 10^{-4}$ and $\mathrm{RMS}(\mathbf{F}_{\mathrm{TS}}) < 1.5 \cdot 10^{-4}$ $\mathrm{E_H/a_0}$. The reference set is used to asses the SPs obtained by the NEB-TS calculations, in Sec. 4.2.

2.3.3 Electronic structure calculations

The electronic structure calculations are performed using the B3LYP[50–52] density functional in the spin-unrestricted formalism along with the D3 dispersion correction employing Becke-Johnson damping,[53,54] with the exception of the GaP(110) system, where instead the D2 dispersion correction[55] is used, as in the original publication.[38]

In order to accelerate the electronic structure calculations, the resolution of identity (RI) approximation is used for the Coulomb integrals[56] and the COSX numerical integration scheme[57,58] with a grid-size 'GRIDX4' (ORCA keyword) is used for the evaluation of the Hartree-Fock exchange integrals. For the self-consistent field (SCF) calculations a combination of the KDIIS algorithm[59,60] and a quasi-Newton scheme (SOSCF)[47,61] is used for most of the systems. However, it turned out to be necessary to turn off the SOSCF for some systems to avoid convergence problems. All calculations use an integration grid of type 'GRID4' (ORCA keyword). The def2-SVP basis set[62,63] is used in the electronic structure computations for the main-group reactions and the Pt-complex.[39] For the GaP(110) system, the def2-SVP basis set is used only for the Ga atoms and 6-31G(d,p)[64,65] for all other atoms. The def2-TZVP basis set[66] is used for the Fe-complex.[37] For the Ga and Pt atoms the Stuttgart-Dresden effective core potentials (SDD ECPs) are used.[67,68] In all calculations, the def2/J auxiliary basis sets are used.[69,70] For the Fe-complex, the broken symmetry DFT approach is used, to converge to a broken-symmetry solution with an $\mathrm{M}_S = \frac{5}{2}$ Fe(III) ion that is antiferromagnetically coupled to an $\mathrm{M}_S = \frac{1}{2}$ superoxo radical, to give a spin state of $\mathrm{M}_S = 2$. To account for solvation effects, the implicit conductor-like polarizable continuum model (CPCM)[71] is used for the Fe and GaP systems. For the Fe-complex, the CPCM is used with a dielectric constant of 4, in order to mimic the protein environment. For the GaP system, the CPCM model is used for modeling the water environment, using the Gaussian charge scheme with a scaled vdW cavity.[72,73]

2.3.4 Parallelization of CI-NEB and NEB-TS

In CI-NEB calculations, the energy and atom forces of the intermediate images can be calculated in a parallel manner for each computational cycle, e.g. by using $N_{\mathrm{im}} - 2$ central processor units (CPUs). Moreover, the electronic structure computation of each image can also make use of additional CPUs. In a NEB-TS calculation, all CPUs used previously in the

EW-CI-NEB calculation are redirected to parallelize the electronic structure computations in the TS search. In the (EW-)CI-NEB and NEB-TS calculations presented in this study, 1 or 2 CPUs are used for the computation of each intermediate image.

3 Benchmark set of main-group molecular reactions

A benchmark set of 121 unique main-group molecular reactions is used to evaluate the performance of both the EW-CI-NEB and NEB-TS methods, see sections 4.1 and 4.2, respectively. Additionally, to illustrate the transferability of the NEB-TS method to more complex reactions relevant to catalysis and bioinorganic chemistry, calculations on three additional reactions are presented[37–39] in section 4.3.

The benchmark set of main-group molecular reactions combines two recently published benchmark sets, created by Birkholz and Schlegel[36] and by Zimmerman.[34] The combined set includes 42 unique reactant configurations containing from 3 to 53 atoms. The systems contain elements: H, B, C, N, O, F, Mg, P, S, Cl and Br. Note that all of the systems have been optimized here and include dispersion correction which was not included in the original work.[34,36] Furthermore, for several systems, either the reactant or product configurations have been slightly modified from the original publications, e.g. to resolve inconsistent atom indexing between the reactant and product geometries, see section SI-1.2.

This benchmark set of reactions is diverse in terms of reaction type and shape of the energy surface, as is seen in Fig. 3 (see also sections SI-1.3 and SI-2.2). The reactions, included in the benchmark set, are primarily single elementary step reactions but there are also a few multi-step pathways, with two or more energy barriers along the RP. The reactant, product and saddle point configurations are accessible in a .xyz format as a supplementary material and are also visualized in section SI-1.1.

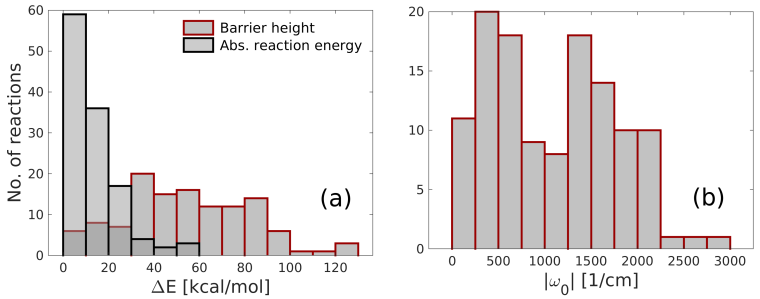


Figure 3: Characterization of the benchmark set of 121 molecular reactions used to assess the performance of the EW-CI-NEB and NEB-TS methods. The energy barrier heights and absolute value of the reaction potential energy is shown in (a) and in (b) the magnitude of the imaginary vibrational frequency at the first order saddle points is shown.

4 Results and discussion

In section 4.1, the performance of CI-NEB and EW-CI-NEB methods is evaluated and compared using the benchmark set of molecular reactions introduced in section 3. In the (EW-)CI-NEB calculations, the objective is to obtain an accurate approximation to the highest energy first order saddle point that connects the given reactant and product states. The formation reaction of 1,2-hexadiene is selected from the benchmark set to compare and illustrate the different image distributions and convergence properties of CI-NEB and EW-CI-NEB. In section 4.2, the results of performance tests of NEB-TS are presented. The reliability of the method is investigated by comparison of the obtained SPs to the set of reference SPs (defined in section 2.3.2). Furthermore, in order to demonstrate the importance of using an accurate initial guess for the SP in NEB-TS, a rearrangement reaction of allyl-vinyl-ether is chosen as an illustrative example. In section 4.2, the NEB-TS method is applied to three exemplary reactions taken from recently reported computational chemistry studies, where in some of these studies the search for the desired SP was found to be problematic using a combination of coordinate-dragging and EF P-RFO.[37–39]

4.1 Comparison of CI-NEB and EW-CI-NEB

To evaluate the performance of the CI-NEB and the EW-CI-NEB method, six sets of calculations are carried out for all 121 reactions of the benchmark set. The sets differ in the type and value of the spring constant and number of images used, see Table. 1. All other

(EW-)CI-NEB parameters are the same, as specified in section 2.3.

Table 1: Parameter values used in six sets of (EW-)CI-NEB calculations carried out on the set of 121 main-group molecular reactions. All other parameters are as specified in section 2.3.

Set	Method	N_{im}	k^{sp} [$\mathrm{E_H}/a_0^2$]
1	CI-NEB	7	0.1
2	CI-NEB	10	0.01
3	CI-NEB	10	0.1
4	CI-NEB	14	0.1
5	EW-CI-NEB	7	0.01–to–0.1
6	EW-CI-NEB	10	0.01–to–0.1

The performance of a set of calculations is specified by both the convergence ratio and the computational efficiency. The convergence ratio is given by the number of calculations where the magnitude of the atom forces acting on the CI drop below the given convergence thresholds within the maximum number of optimization steps (see section 2.3.2). The computational efficiency is measured by the average number and standard deviation of energy and force evaluations required for convergent CI-NEB calculations. The performance of the six sets of calculations is summarized in Fig. 4.

4.1.1 Fixed spring constant CI-NEB

The CI-NEB method with $N_{\mathrm{im}} = 7$ (i.e. five intermediate images) and $k^{\mathrm{sp}} = 0.1$ $\mathrm{E_H/a_0^2}$ (set 1) has a convergence ratio of 74% and uses on average 874 ± 665 energy/force evaluations (174 evaluations per movable image). It is quite common to use 5 intermediate images in CI-NEB calculations if the RP is simple. By increasing N_{im} to 10 (set 3) the convergence ratio increases to 82% and the energy/force evaluations increase to 1108 ± 758 (138 evaluations per movable image). Analysis of the non-convergent calculations in sets 1 and 3 reveals that the problem is typically caused by poor resolution near the CI and hence an inaccurate tangent estimate. This causes the location of the CI to oscillate and can lead to unreasonable optimization steps being taken. Therefore, the resolution around the energy barrier needs to be improved by increasing the number of images. When the number of images is further increased to $N_{\mathrm{im}} = 14$ (set 4) the convergence ratio increases to 93% while the energy/force evaluations become 1824 ± 1395 (152 evaluations per movable image). CI-NEB optimization profiles, using variable number of images, for two selected reactions of the benchmark are shown in section SI-2.5. The computational effort in the CI-NEB calculations necessarily increases when more images are added. However, as for any numerical method that is based

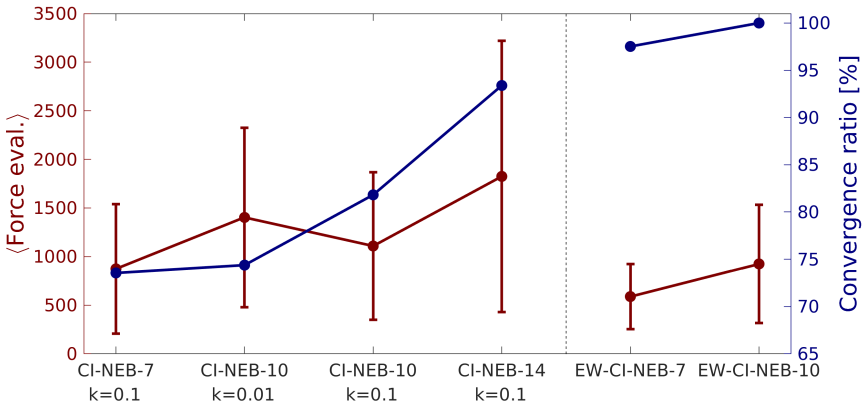


Figure 4: Results from four sets of CI-NEB calculations and two sets of EW-CI-NEB calculations (defined in Table 1) for the benchmark set of 121 molecular reactions. The notation (EW-)CI-NEB-X gives the total number of images used in the calculations, where $X = 7$, 10 and 14. For the CI-NEB calculations, the spring constant k^{sp} used in the CI-NEB calculations is specified on the x-axis and is in units of $\mathrm{E_H/a_0^2}$. For the EW-CI-NEB calculation, the spring constant is scaled from $k^{\mathrm{sp}} = 0.01$ to 0.1 $\mathrm{E_H/a_0^2}$. The average number of energy/force evaluations in convergent CI-NEB calculations is given by the red vertical axis on the left along with the standard deviation shown as an error bar. The convergence ratio is given by the blue vertical axis on the right. A calculation is convergent if the magnitude of the atom force acting on the climbing image drops below the user-supplied convergence tolerance within 500 optimization steps, see section 2.3.2. Note that non-convergent calculations are omitted from the calculation of averages.

on a discretization of a continuous variable, the number of discretization points needs to be large enough.

The L-BFGS optimization method is expected to be sensitive to the value of the spring constant since the off-diagonal elements of the approximate Hessian matrix arise from the spring interactions acting between images.[42] In sets 1, 3 and 4 the spring constant was chosen to be $k^{\mathrm{sp}} = 0.1$ $\mathrm{E_H/a_0^2}$. This value is larger than the commonly used value for the spring constant, which is closer to $k^{\mathrm{sp}} = 0.01$ $\mathrm{E_H/a_0^2}$. In fact, the set of CI-NEB calculations with $N_{\mathrm{im}} = 10$ and $k^{\mathrm{sp}} = 0.01$ $\mathrm{E_H/a_0^2}$ (set 2) is found to have a convergence ratio of 74% using 1403 ± 922 energy/force evaluations (175 evaluations per movable image), significantly less robust and computationally more demanding than the CI-NEB calculations using stiffer springs. In fact, $k^{\mathrm{sp}} = 0.1$ $\mathrm{E_H/a_0^2}$, is found to be the optimal value for the spring constant, based on more extensive tests covering a large range, but on a much smaller test-set, see section SI-4.1.

The problem of poor resolution of the energy maximum observed here for CI-NEB calculations of molecular reactions is less severe for calculations of diffusion events in solids or reactions on surfaces of solids. This is because the RPs of molecular reactions are often characterized by, irrelevant, long segments of small energy variation. In such cases, an even distribution of images does not represent the optimal distribution of images, since the majority of the images are being used to resolve these unimportant segments of the path rather than the region of the energy barrier.

4.1.2 Energy weighted CI-NEB

A more suitable distribution of images can be obtained by using energy-weighted spring constants. In this scheme, the spring constant for each line-segment of the path is adaptively scaled according to the relative energy of the connected images such that stiffer springs act between images in higher energy regions along the reaction path. This leads to an accumulation of images near the energy maxima, as is illustrated in Fig. 1.

Calculations using the EW-CI-NEB method with $N_{\mathrm{im}} = 7$ and the spring constants scaled from $k_{\mathrm{l}} = 0.01$ to $k_{\mathrm{u}} = 0.1\ \mathrm{E_H/a_0^2}$ (set 5), yield a convergence ratio of 98% and the average number of energy/force evaluations of 589±333 (117 evaluations per movable image). If the number of images is increased to $N_{\mathrm{im}} = 10$ (set 6), a 100% convergence ratio is obtained with an average of 924±608 energy/force evaluations (115 evaluations per movable image). In other words, EW-CI-NEB when used with large enough N_{im} is able to converge the CI below the prescribed convergence thresholds and yield highly accurate approximations to the relevant SPs for all 121 reactions of the benchmark set. About 80% of the saddle points obtained by EW-CI-NEB are already of first order. Clearly, the EW-CI-NEB method (sets 5 and 6) shows significant improvements in performance over the standard CI-NEB method with equal distribution of images (sets 1–4). As EW-CI-NEB leads to an increased density of images in the region of the energy barrier, a smaller number of images can be used in EW-CI-NEB to locate the SP compared to CI-NEB. In fact, EW-CI-NEB appears to be less sensitive to the choice of N_{im}, see section SI-4.2. The convergence behavior of CI-NEB and EW-CI-NEB is shown in section SI-2.6 for two example reactions from the benchmark set.

To further demonstrate the difference in the image distribution obtained by CI-NEB and EW-CI-NEB, the relative deviation of the inter-image distance between CI and its two neighbors from the even distribution of the RPs is evaluated and reported in section SI-2.4. For the CI-NEB calculations using $N_{\mathrm{im}} = 10$ and $k^{\mathrm{sp}} = 0.1\ \mathrm{E}_H/\mathrm{a_0}$, 92% of the convergent CI-NEB calculations show a less than 2% deviation from the even distribution and 99% of the systems are within a 5% deviation. This is to be expected since CI-NEB attempts to obtain an even distribution of images and the selected spring constants are relatively stiff

(compared to typical values used for k^{sp}). For CI-NEB, a large deviation from the ideal even distribution can be a sign of instability and kinked paths. For EW-CI-NEB, using $N_{im} = 10$, the inter-image distances in higher energy regions of the path become shorter as the images are shifted up along the path due to stiffer springs in those regions. The inter-image distance between CI and its closer neighboring image is less than 1/2 of the even distribution for 92% of the EW-CI-NEB paths and less than 1/3 for 64% of the paths. This shows how EW-CI-NEB effectively introduces a denser distribution of images around the CI and thereby improves the approximation to the path tangent at the CI.

4.1.3 Comparison of CI-NEB and EW-CI-NEB for an example reaction

To further examine the difference in convergence behavior between CI-NEB and EW-CI-NEB, the reaction of 1,3-butadiene and ethylene to form 1,2-hexadiene is chosen from the benchmark set of molecular reactions. CI-NEB calculations using $N_{im} = 10, 14, 18$ and 34 with $k^{sp} = 0.1\ E_H/a_0^2$ are compared to EW-CI-NEB calculations using $N_{im} = 7$ and 10 with k^{sp} scaled from 0.01 to 0.1 E_H/a_0^2. The resulting paths, computational efficiency and key configurations (reactant, product and saddle points) are shown in Figs. 5 and 6. In this case, the computational efficiency is measured by the number of optimization steps rather than the number of energy/force evaluations.

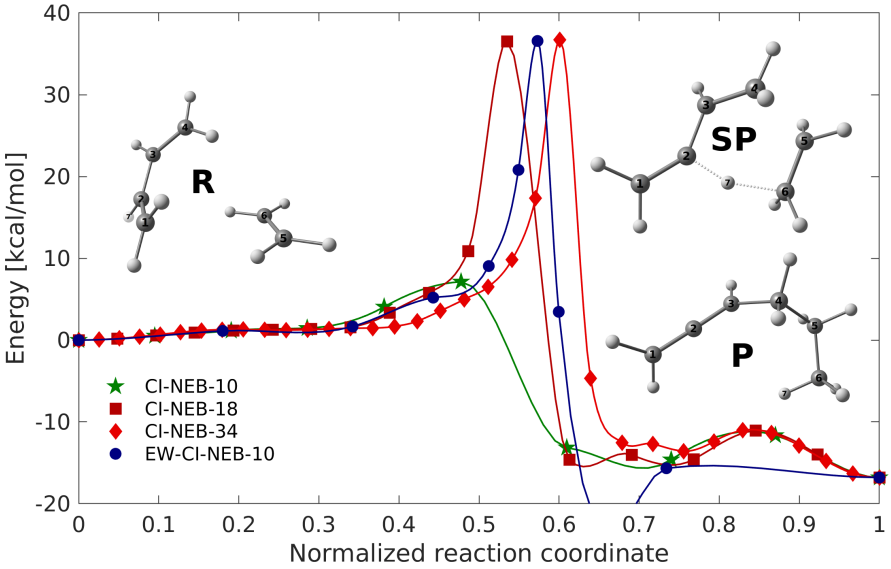


Figure 5: Energy curves (or profiles) for the reaction path of ethylene and 1,3-butadiene to form 1,2-hexadiene. The CI-NEB calculations use fixed spring constant of $k^{sp} = 0.1\ E_H/a_0^2$ and the number of images is $N_{im} = 10$, 18 or 34. For comparison, the energy weighted CI-NEB calculations use k^{sp} scaled from 0.01 to 0.1 E_H/a_0^2 and $N_{im} = 10$. Convergence is defined only in terms of the atomic forces acting on the climbing image, so the intermediate images are not converged to the same degree. The large energy barrier is associated with H-transfer from 1,3-butadiene to ethylene and a C-C bond formation. The second energy barrier corresponds to a rearrangement to yield the specified product state. Note that the energy curve for CI-NEB-10 is not converged and is unable to resolve the large energy barrier. The reactant (denoted by **R**), saddle point (denoted by **SP**) and product (denoted by **P**) configurations are shown as insets.

The resulting reaction path for the formation of 1,2-hexadiene (see Fig. 5) is characterized by a long, flat energy tail that corresponds e.g. to the rotation of the ethylene fragment into the necessary orientation needed for the reaction to occur. This is typical for calculations of molecular association reactions and can for some reactions be reduced by improving the alignment of the initial reactant and product configurations. The reaction then proceeds by H-transfer from 1,3-butadiene to ethylene and C-C bond formation with an energy barrier of 36.7 kcal/mol. Finally, to yield the product state, the reaction is followed by a isomerization with a barrier of 5.8 kcal/mol, relative to the product state energy minimum.

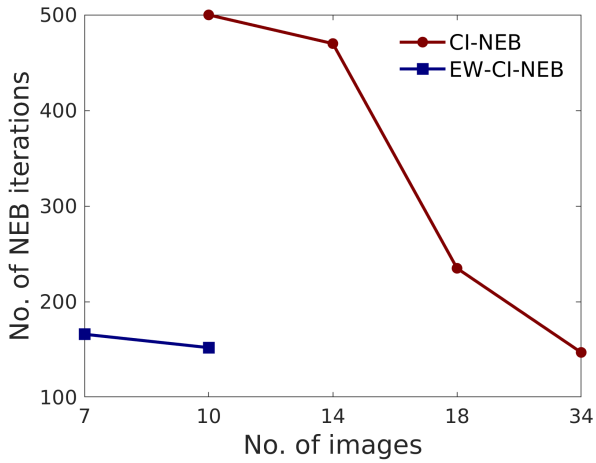


Figure 6: Efficiency of CI-NEB and EW-CI-NEB calculations of the reaction of ethylene and 1,3-butadiene to form 1,2-hexadiene. The efficiency is measured in terms of NEB iterations (or optimization steps) required to reach convergence. The CI-NEB calculations are carried out with a fixed spring constant of $k^{sp} = 0.1\ E_H/a_0^2$ and $N_{im} = 10$, 14, 18 or 34. The energy-weighted CI-NEB calculations scale the k^{sp} from 0.01 to 0.1 E_H/a_0^2 and are carried out using $N_{im} = 7$ and 10. Note that CI-NEB-10 with $k^{sp} = 0.1\ E_H/a_0^2$ failed to converge within 500 iterations.

The CI-NEB calculation with fixed spring constant and $N_{im} = 10$ is unable to converge within the 500 allowed optimization steps. Moreover, the energy profile of the path from the last optimization step of this CI-NEB calculation does not give a good estimate of the large energy barrier. The CI-NEB calculation with $N_{im} = 14$ is able to converge in 469 optimization steps. A closer inspection of the calculation reveals that this large number of steps is due to an oscillatory behavior of the CI. This behavior is attributed to the poor resolution of the energy barrier. By increasing the number of images even further to $N_{im} = 18$ and to 34, the oscillations are quenched and a large reduction in the number of optimization steps is achieved, to 235 and 147, respectively. Interestingly, for $N_{im} = 18$, the resolution of the energy barrier is still quite poor as the images adjacent to the CI are located at the base of the energy peak, see Fig. 5. The EW-CI-NEB calculations with $N_{im} = 7$ and 10 offer a better resolution of the energy barrier and achieve comparable performance to CI-NEB with $N_{im} = 34$ where the number of optimization steps required is 166 and 152, respectively. However, the EW-CI-NEB calculations are unable to resolve other features of the path such as the second energy barrier with this low number of images. In practice it may

be advantageous to partition multi-step reactions into multiple (EW-)CI-NEB calculations, with one (EW-)CI-NEB calculation per energy maxima.

4.2 Performance of NEB-TS

The performance of the NEB-TS is assessed using four sets of different method parameters. The results are compared to the IDPP-TS method and summarized in Figs. 7 and 8. Two different TS activation thresholds for stopping the EW-CI-NEB calculation and starting the TS search are used, i.e. $\epsilon_{\mathrm{max}}^{\mathrm{TS}} = 0.01$ and $0.002\ \mathrm{E_H/a_0}$. Also, the initial Hessian matrix is either computed analytically or constructed using the empirical Almlöf scheme. Other parameters of the NEB-TS method are as specified in section 2.3.

The convergence ratio of a set of NEB-TS calculations is determined by the number of calculations that converge on a SP within 500 optimization steps in total, i.e. sum of steps used in the EW-CI-NEB and the subsequent TS search. The set of SPs obtained by the NEB-TS calculations are further analyzed according to the absolute energy difference from the reference set of SPs defined in section 2.3.2. The computational efficiency is measured by the average number and standard deviation of energy/force evaluations required by convergent calculations.

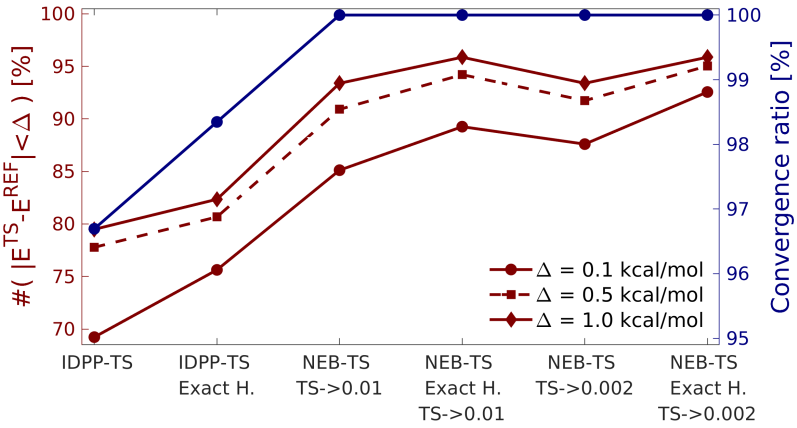


Figure 7: Convergence ratio and accuracy of results obtained by NEB-TS and IDPP-TS calculations for the benchmark set of 121 molecular reactions. The calculations differ in the degree to which the climbing image of the EW-CI-NEB calculation is converged before the TS search is started, i.e. $\epsilon_{\max}^{\mathrm{TS}}$, is 0.01 and 0.002 $\mathrm{E_H/a_0}$. In IDPP-TS, the initial EW-CI-NEB phase is skipped. The calculations also use either an approximation to the initial Hessian matrix constructed using the empirical Almlöf scheme or computed analytically (labeled on the x-axis). All calculations use 10 images. The convergence ratio is given by the blue line (right vertical axis). A calculation is considered to be converged if the magnitude of the atomic force drops below $\max(|\mathbf{F}_{\mathrm{TS}}|) < 5 \cdot 10^{-4}$ $\mathrm{E_H/a_0}$ and $\mathrm{RMS}(\mathbf{F}_{\mathrm{TS}}) < 3 \cdot 10^{-4}$ $\mathrm{E_H/a_0}$, within 500 optimization steps. The fraction of NEB-TS and IDPP-TS calculations that yield a saddle point estimate within Δ of the reference set is shown in red (left vertical axis) for three different values of Δ, or $\Delta = 0.1, 0.5$ and 1.0 kcal/mol.

All four sets of the NEB-TS calculations are found to have a 100% convergence ratio, regardless of when EW-CI-NEB is stopped and the TS search is started and also irrespective of whether the exact or an empirical Hessian matrix is used at the beginning of the TS search. In section SI-2.6, the convergence behavior of CI-NEB, EW-CI-NEB and NEB-TS is shown for two example reactions of the benchmark set.

The NEB-TS method with the earlier EW-CI-NEB stopping criterion, i.e. $\epsilon_{\max}^{\mathrm{TS}} = 0.01$ $\mathrm{E_H/a_0}$, and the empirical Almlöf initial Hessian matrix is found to require only 305 ± 140 energy/force evaluations. The average number of optimization cycles is 77. If the exact Hessian is computed and used at the beginning of the TS search, the efficiency improves only marginally as the average is reduced to 300 energy/force evaluations. About 92% of the saddle points obtained from these calculations are confirmed to be first order saddle points by carrying out a vibrational frequency analysis. When a higher order saddle point is

obtained, the other imaginary frequencies are found to be small, or $|\omega| < 75$ cm^{-1}. Majority of the computational effort of the NEB-TS calculations is in the initial EW-CI-NEB phase, or about 260 energy/force evaluation. This corresponds to roughly 32 parallel computational cycles over the images of EW-CI-NEB and 45 cycles in the TS part.

By converging the EW-CI-NEB to a tighter threshold, $\epsilon_{\max}^{\mathrm{TS}} = 0.002$ E$_{\mathrm{H}}$/a$_0$, the CI is brought even closer to the desired SP prior to the TS search. For this set of NEB-TS calculations, the average number of energy/force evaluations increases to 534 ± 237 and 95 cycles, where an average of 501 energy/force evaluations (about 62 cycles) is used in EW-CI-NEB and an average of 33 evaluations in the TS part. As expected, the total computational effort is significantly increased, while the computational effort of the individual TS searches is reduced.

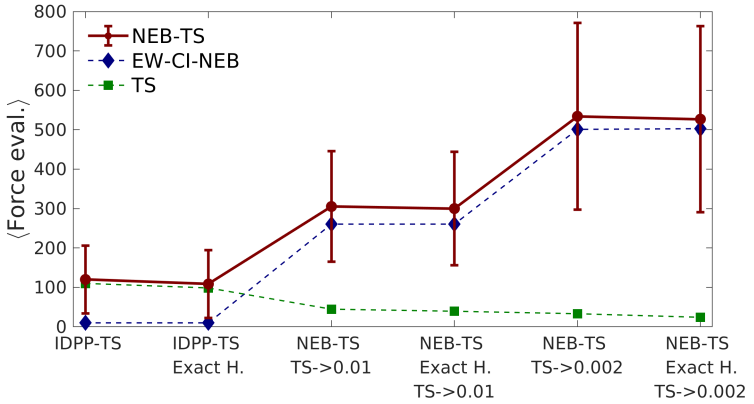


Figure 8: The computational efficiency of NEB-TS and IDPP-TS calculations of the benchmark set of 121 molecular reactions. The NEB-TS calculations differ in the degree to which the climbing image of the initial EW-CI-NEB calculation is converged before the TS search is started, i.e. $\epsilon_{\max}^{\mathrm{TS}}$, is 0.01 or 0.002 E$_{\mathrm{H}}$/a$_0$. In IDPP-TS, the initial EW-CI-NEB phase is skipped. The calculations also use either an initial Hessian matrix that is constructed using the empirical Almlöf scheme or computed analytically (labeled on the x-axis). All calculations use 10 images. The average number of energy/force evaluations required by convergent NEB-TS (and IDPP-TS) calculations to locate saddle points is shown with red solid line along with the standard deviation shown as an error bar. The dashed blue and green dashed lines show how the computational effort of the NEB-TS calculations is divided into the initial EW-CI-NEB phase and the subsequent TS search. Non-convergent calculations are omitted from the calculation of the averages.

For comparison, results of the IDPP-TS calculations are also reported, where the EW-CI-NEB is skipped and the highest energy image along an initial path generated by the IDPP

method is used as the initial guess configuration for the TS search. The TS search is then inevitably started from configurations that are located further away from the desired SP and often in regions on the energy surface where the Hessian is characterized by a larger number of negative eigenvalues as compared to the initial configurations generated by NEB-TS, as shown in Fig. 9.

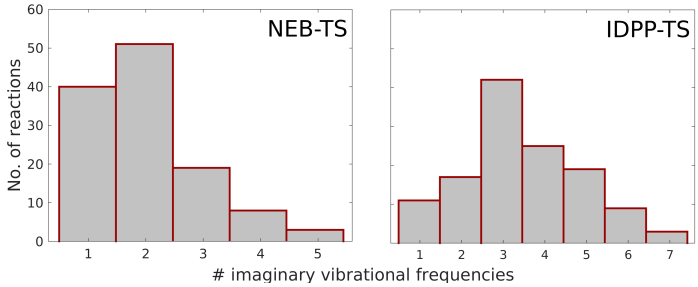


Figure 9: The number of negative eigenvalues (or imaginary vibrational frequencies) of the Hessian matrix at the starting configuration of the TS search in NEB-TS (left) and in IDPP-TS (right). In NEB-TS, the starting configuration of the TS search is taken as the climbing image configuration along the initial, partially converged EW-CI-NEB calculation using a TS activation threshold $\epsilon_{\mathrm{max}}^{\mathrm{TS}} < 0.01$ $\mathrm{E_H/a_0}$. In IDPP-TS, the starting configuration of the TS search is the highest energy image along an IDPP interpolated path connecting the given reactant and product energy minima.

The average number of energy/force evaluations for the two sets of IDPP-TS calculations is found to be about 120 and 109, respectively, i.e. about a factor of three larger than the TS search of NEB-TS (i.e. excluding the EW-CI-NEB part). Surprisingly, the convergence ratio of IDPP-TS is found to be 97% and 98%, using the model and exact initial Hessian, respectively. The high convergence ratio of IDPP-TS can be attributed to both the quality of the initial path generated by the IDPP method and hence the starting point used for the TS search (see section SI-2.1) and also to the robustness of the TS search method, i.e. EF P-RFO in internal coordinates. For comparison, the convergence ratio of a TS search started from the highest energy image along a linear interpolation in Cartesian coordinates is 69% (data is not shown here).

However, the quality of the initial guess configuration and the initial Hessian (and hence the reaction coordinate selected in the TS search) can determine whether the IDPP-TS and NEB-TS calculations are able to locate a SP that corresponds to the given reaction. To demonstrate this, the absolute energy difference of the SPs obtained by the different NEB-TS and IDPP-TS calculations and the reference set of SPs is examined, see section SI-3.1.

The fraction of SPs obtained by the calculations that deviate from the reference SPs by more than a selected value, Δ, are shown in Fig. 7.

The SPs obtained by IDPP-TS can deviate in energy from the reference SPs by up to roughly 80 kcal/mol. Around 31% of the SPs obtained by IDPP-TS show a larger energy deviation than 0.1 kcal/mol and 22% of the SPs a larger deviation than 0.5 kcal/mol. In NEB-TS, where the CI of EW-CI-NEB is converged to $\epsilon_{\max}^{\mathrm{TS}} = 0.01$ $\mathrm{E_H/a_0}$ before starting the TS search, the distribution of energy differences is found to span a range of 0 to 30 kcal/mol. The fraction of SPs that deviate by more than 0.1 and 0.5 kcal/mol is found to decrease to 15% and 9%, respectively. If the exact Hessian matrix is used to initialize the TS calculations (of IDPP-TS or NEB-TS), instead of the model Almlöf Hessian, the ratios can be further decreased by roughly 4–7%. The best agreement to the reference set of SPs is obtained for NEB-TS using $\epsilon_{\max}^{\mathrm{TS}} = 0.002$ $\mathrm{E_H/a_0}$ and an exact initial Hessian matrix. For this set of calculations, about 7% and 6% of the SPs show a larger energy deviation than 0.1 and 0.5 kcal/mol compared to the reference set, respectively.

Calculations that exhibit large energy differences are likely to have converged to different saddle points that may correspond to a different reaction than is specified by the initial and final state configurations. Further analysis of a few selected calculations reveals other possibilities for discrepancy, see section SI-3.2. First is the existence of multiple maxima along the RP in a complex reaction mechanism where the TS search ends up converging to a lower energy SP along the RP. This can occur in a multi-step transition where the EW-CI-NEB calculation is not converged well enough for the CI to be placed near the highest SP. Another reason is that an alternative RP for the same reactant and product state is found. Often, these shortcomings can be remedied by lowering the convergence of the EW-CI-NEB calculation and hence improving the initial configuration that is supplied as input to the TS search. A detailed analysis of one such example reaction, a rearrangement of allyl-vinyl-ether to 1-pentene-5-one, is given in the following subsection.

4.2.1 Dependency of the TS-search on the initial guess

To demonstrate the dependence of the NEB-TS method on the quality of the initial configuration, tangent and curvature information provided by the EW-CI-NEB calculation, the rearrangement of allyl-vinyl ether to form 1-pentene-5-one is chosen as an example. Calculations are carried out for various values of the TS activation threshold, $\epsilon_{\max}^{\mathrm{TS}}$, ranging from 0.02 $\mathrm{E_H/a_0}$ to 0.002 $\mathrm{E_H/a_0}$. Other parameters are kept the same as in the previous subsection. The results are summarized in Figs. 10 and 11. Two different SPs are found depending on the value of $\epsilon_{\max}^{\mathrm{TS}}$. For a low threshold, $\epsilon_{\max}^{\mathrm{TS}} \leq 0.012$ $\mathrm{E_H/a_0}$, a SP is obtained that is in agreement to the one found by converged EW-CI-NEB calculation and is confirmed to

be connected to the given reactant and product state by internal reaction coordinate (IRC) analysis,[74] as is shown in Fig. 11. The correct SP corresponds to rearrangement to form the given 1-pentene-5-one product state. When the threshold value is larger, $\epsilon_{\max}^{\mathrm{TS}} \geq 0.014\ \mathrm{E_H/a_0}$ the NEB-TS calculation converges to a different incorrect SP that appears to be connected to conformers of the allyl-vinyl ether reactant state. This incorrect SP is approximately 27 kcal/mol higher in energy than the correct SP, which is found to give a barrier height of about 32 kcal/mol.

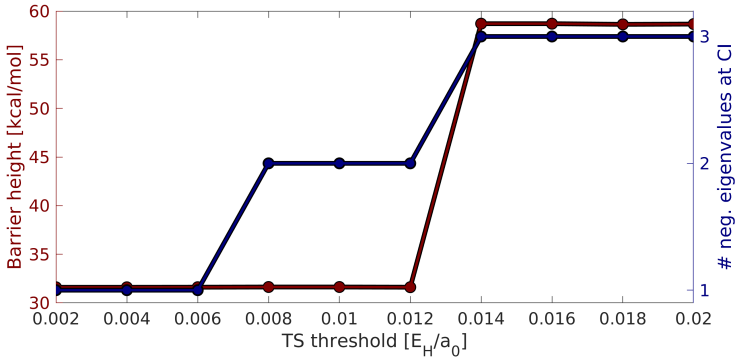


Figure 10: Results of NEB-TS calculations for various values of the activation threshold, ϵ_{TS}, when the EW-CI-NEB calculation is carried out for a rearrangement of allyl-vinyl ether to form 1-pentene-5-one. The red curve shows the energy of the saddle point obtained with respect to the energy of the reactant energy minimum (left vertical axis) The blue curve shows the number of negative eigenvalues obtained from an analytical calculation of the Hessian matrix at the climbing image configuration when the EW-CI-NEB calculation is terminated (right vertical axis). Two different saddle points are obtained depending on the value of ϵ_{TS}. IRC analysis shows that the lower energy saddle point corresponds to the correct reaction and yields both the reactant and product energy minima. While, the higher energy saddle point leads to a different conformer of the allyl-vinyl-ether, i.e. the saddle point is not connected to the given product state, see Fig. 11. This is most likely because for NEB-TS calculations with $\epsilon_{\max}^{\mathrm{TS}} \geq 0.014\ \mathrm{E_H/a_0}$, the incorrect eigenvector is selected and followed, the TS activation threshold is lowered further, the EW-CI-NEB calculation is converged to a tighter threshold and the TS search starts from a better initial configuration, where the right eigenvector (i.e. the mode that corresponds to the desired reaction coordinate) is selected and followed to the correct saddle point.

Calculation of the analytical Hessian matrix for the CI configuration obtained at the termination of the EW-CI-NEB calculations reveals that for $\epsilon_{\max}^{\mathrm{TS}} \geq 0.014\ \mathrm{E_H/a_0}$ there are three negative eigenvalues and the eigenvector corresponding to the lowest eigenvalue does

not correspond to the correct reaction coordinate of the allyl-vinyl-ether rearrangement. At this point, the initial configuration for the TS search and tangent estimate provided by EW-CI-NEB are not sufficiently accurate for the TS search to identify the correct SP. Instead, the TS search selects and most likely follows the wrong eigenvector leading to a SP that does not lead to the given 1-pentene-5-one energy minimum. The situation is not fixed by analytically computing the initial Hessian matrix, instead of using the Almlöf model Hessian matrix.

By lowering $\epsilon_{\text{max}}^{\text{TS}}$ to 0.012 $\text{E}_\text{H}/\text{a}_0$, the number of negative eigenvalues at the CI configuration is reduced to two. At this point, initial configuration provided by the CI is sufficiently close to the SP such that the selected eigenvector to be followed by the TS search corresponds to the correct reaction coordinate and the TS search converges the correct SP, i.e. the SP connected to the given reactant and product states. By further lowering the activation threshold to $\epsilon_{\text{max}}^{\text{TS}} = 0.006$ $\text{E}_\text{H}/\text{a}_0$ gives an even higher quality CI configurations with a single negative eigenvalue in the Hessian matrix and the TS search continues to identify the correct SP. The absolute values of the negative eigenvalues obtained from the Hessian matrix of the CI in the NEB-TS calculations presented in Fig. 10 are given in SI-3.3. This example shows that a key aspect of a successful NEB-TS calculation is good input from the EW-CI-NEB calculation to the TS search. The required TS activation threshold value is expected to be system dependent. However, the results obtained here for the 121 molecular reaction test set indicate that a value of $\epsilon_{\text{max}}^{\text{TS}} = 0.01$ $\text{E}_\text{H}/\text{a}_0$ gives reliable results.

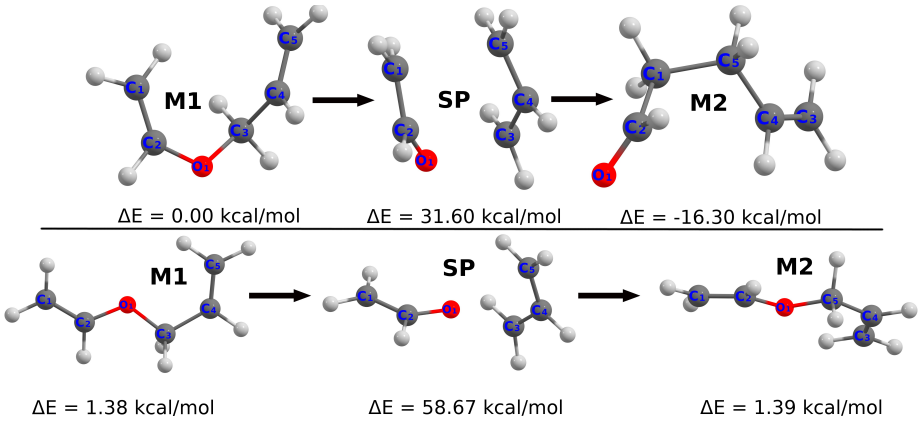


Figure 11: The two energy minima (denoted by **M1** and **M2**) obtained from IRC calculations carried out on the saddle points (denoted by **SP**) of the NEB-TS calculations presented in Fig. 10, using $\epsilon_{TS} = 0.002$ E_H/a_0 (shown in the upper panel) and $\epsilon_{TS} = 0.02$ E_H/a_0 (shown in the lower panel). The energy difference of the **M1**, **M2** and **SP** configurations and the reactant energy minimum are given. In the upper panel, for NEB-TS($\epsilon_{TS} = 0.002$ E_H/a_0) the minimum **M1** corresponds to the correct given reactant energy minimum and **M2** to the product energy minimum. For NEB-TS($\epsilon_{TS} = 0.02$ E_H/a_0) both **M1** and **M2** are observed to be conformers of the reactant energy minimum.

4.3 Applications: Bioinorganic and heterogenous catalysis reactions

In the previous sections, the performance of EW-CI-NEB and NEB-TS methods is evaluated for a benchmark set of 121 reactions of main-group molecules. With a small enough activation threshold for starting the TS search of NEB-TS, the method is found to be highly robust, as well as efficient. In this section, the NEB-TS method is applied to three larger systems that have recently been discussed in the literature. Some of these reactions were found to be computationally challenging using a combination of coordinate-dragging and single ended SP searches. Two of the reactions are taken from bioinorganic chemistry[37,39] and the third one from heterogeneous catalysis research.[38]

The first reaction considered is the platination of guanine by a monoaqueous nedaplatin complex, see Fig. 12.[39] The reactant state configuration is characterized by a square-planar geometry where a hydroxide, two ammonia molecules and a glycolate group are bound to a Pt(II) center with a guanine molecule in its vicinity. In the product, a water molecule is expelled from the complex and the guanine coordinates to the Pt(II).

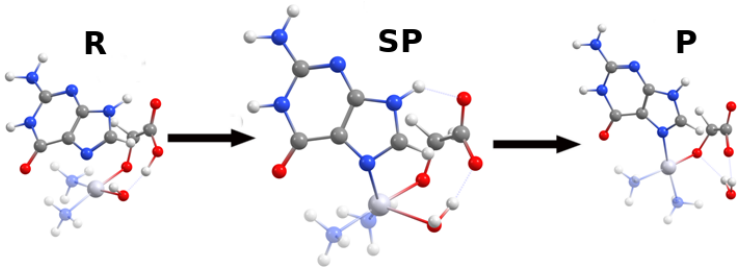


Figure 12: The reactant (denoted by **R**), saddle point (denoted by **SP**) and product state (denoted by **P**) configurations are shown for platination of the guanine nucleobase. For clarity, atoms that do not actively participate in the reaction are made transparent. The elements included are: Pt (silver), O (red), N (blue), C (gray) and H (white). Note that in the reactant state, the OH group is negatively charged while the COOH group is neutral. As the proton jumps from the COOH group to the hydroxyl group, H_2O is formed and expelled leaving a negatively charged carboxylate.

The second reaction involves a decarboxylation reaction of α-ketoglutarate (modelled as pyruvate) bound to the Fe(III)-(O_2)-active-site of the Fe(II) and α-ketoglutarate (αKG)-dependent dioxygenase enzyme, see Fig. 13. In the reactant state configuration, the O_2^- radical ion is bound and antiferromagnetically coupled to a high spin Fe(III) center in an octahedral geometry, bound to two His residues (modelled as imidazole groups), acetate and a bidentate α-ketoglutarate (modelled as pyruvate). In the product, CO_2 is expelled from the the substrate and a Fe(II) trigonal bipyramidal peroxosuccinate intermediate is formed.

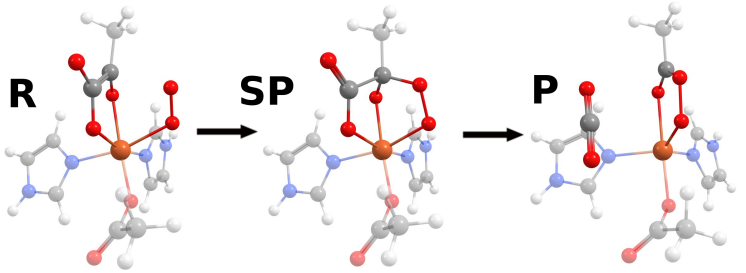


Figure 13: The reactant (denoted by **R**), saddle point (denoted by **SP**) and product state (denoted by **P**) configurations are shown for decarboxylation of bound α-ketoglutarate (modelled as pyruvate) to the Fe(III)-O_2-active-site complex of the Fe(II) and α-ketoglutarate (αKG)-dependent dioxygenase enzyme. For clarity, atoms that do not actively participate in the reaction are made transparent. The elements included are: Fe (orange), O (red), N (blue), C (gray) and H (white).

The third reaction is a hydride ion transfer from a GaP(110) surface to pyridine,[38] see Fig. 14. Here, both the hydride and pyridine molecule are initially adsorbed on the surface, along with water and partially dissociated water molecules (OH groups) that are included to incorporate explicit solvation effects in tandem with an implicit solvation model. It is important to note that reactions where solvation molecules are explicitly included are often characterized by many soft degrees of freedom and large flexibility. This can lead to inaccurate estimates of the activation entropy. Such reactions can, furthermore, have an ensemble of (inconsequential) saddle points and an analysis of the reaction rate may then require the use of full transition state theory where the free energy difference between the initial and transition state is evaluated, with variational optimization of the location and shape of the transition state.[3]

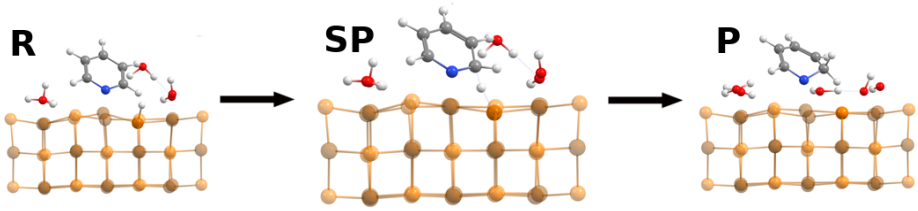


Figure 14: The reactant (denoted by **R**), saddle point (denoted by **SP**) and product state (denoted by **P**) configurations are shown for a hydride transfer from a GaP(110) surface to an adsorbed pyridine. For clarity, atoms that do not actively participate in the reaction are made transparent. The elements included are: Ga (beige), P (orange), O (red), N (blue), C (gray) and H (white).

In the cited publications, the identification of the SPs was carried out with a combination of one- and two-dimensional relaxed surface scans followed by a TS search. Some of the calculations turned out to be problematic and computationally demanding,[77,78] where the reaction coordinates are relatively complex. All three reactions proceed in a single elementary step characterized by multiple bond breaking and formation. NEB-TS calculations for these three reactions were carried out using the parameter values given in section 2.3. The NEB-TS calculations converged without any problems and the results are summarized in Table 2. The NEB-TS method was able to identify the first order SPs for all three reactions using on average 298 energy/force evaluations, showing comparable efficiency to the NEB-TS calculations on the large benchmark set of main-group molecular reactions. This demonstrates the usefulness of the NEB-TS method for challenging real-world chemical systems and structurally complex SPs.

Table 2: Summary of results for NEB-TS calculations on three reactions; (i) Pt of guanine nucleobase, (ii) decarboxylation of bound α-ketoglutarate to the Fe(III)-O_2-active-site complex of the Fe(II) and α-ketoglutarate dependent dioxygenase enzyme and (iii) hydride transfer fom GaP(110) surface to an adsorbed pyridine molecule.

Reaction	# Force eval.	Barrier (kcal/mol)	ω_0 [1/cm]
Pt of guanine	176	32.19	-179.5
Decarboxylation of α-ketoglutarate	304	6.98	-486.9
H^- transfer from GaP(110)	415	41.02	-1388.5

5 Conclusions

In the first part of this study, two different variants of the CI-NEB method are evaluated, on a diverse benchmark set of 121 molecular reactions involving main-group elements,[34,36] namely the standard formulation of CI-NEB[13,19] and energy-weighted (EW) CI-NEB.[19] An important aspect of the CI-NEB calculations is the efficient use of parallel computing as the energy and atom forces of all the images along the path can be calculated simultaneously. The (EW-)CI-NEB calculations were first optimized with respect to method parameters. Moreover, problems that are often encountered in CI-NEB calculations started from a linear interpolation in Cartesian coordinates are resolved for this benchmark set of reactions by using the IDPP method[9] to generate the initial path.

In CI-NEB, all of the spring constants are chosen to be equal, resulting in an even distribution of images along the reaction path. In EW-CI-NEB calculations, however, the stiffness of the springs acting between images is increased for images located in higher energy regions. This causes the images to accumulate in the more important regions of the path, i.e. around the energy maxima. The energy-weighting is of particular importance for calculations of flexible molecular reactions, where the reaction paths are often observed to be long and may be characterized by small change in the energy. In such situation, the energy-weighted springs focus the computational effort away from the unimportant parts of the reaction path to the energy barrier. This results in improved resolution and tangent estimate at the CI which is often found to be crucial for reaching convergence.

The performance of the EW-CI-NEB method on the benchmark set of main-group molecular reactions is found to be significantly better both in terms of reliability and reduction of computational effort as compared to the standard CI-NEB method. In fact, the convergence ratio increases from 82% to 100% and the computational effort is reduced by 17-34% (depending on the value of the spring constant selected) when EW-CI-NEB is used rather than CI-NEB, using $N_{\mathrm{im}} = 10$. The EW-CI-NEB method is also shown to be less sensitive

to the number of images, where EW-CI-NEB calculations, with $N_{\mathrm{im}} = 7$ images, has a convergence ratio of 98%, while CI-NEB using $N_{\mathrm{im}} = 14$ has a lower convergence ratio of 93% and requires on average a factor of roughly three more energy/force evaluations per reaction.

The combination of double ended and single ended methods is likely to represent the most efficient way to locate saddle points reliably for a given reaction,.[12,13,26] For this reason, the NEB-TS and IDPP-TS methods are introduced. In the NEB-TS method, the EW-CI-NEB method is used to obtain a partial convergence to the reaction path connecting the given reactant and product states. Then, a subsequent eigenvector-following calculation (TS search)[21,22] is automatically started from the current position of the climbing image. This allows the TS search to be started from a good initial configuration. Moreover, the reaction path tangent at the climbing image is used to identify the correct reaction coordinate to be followed. The initial Hessian matrix at the climbing image can either be computed analytically or constructed using the empirical Almlöf estimate[47] of the Hessian matrix, revised to include information about the curvature along the path. The NEB-TS method is found to be remarkably robust and efficient, giving 100% convergence ratio and requiring on average only 305 energy/force evaluations. By using the exact Hessian matrix at the beginning of the TS search, the efficiency is only slightly improved. Most of the computational effort of NEB-TS calculations is spent in the initial EW-CI-NEB calculation to a starting configuration that lies close enough to the saddle point of the given transition. The degree to which the climbing image needs to be converged, prior to the single ended SP search, can therefore be adjusted for specific systems to further reduce the computational effort. With the energy-weighted scheme it is possible to use fewer images than with CI-NEB calculations. In the IDPP-TS method, the EW-CI-NEB calculation is skipped. Instead, the eigenvector-following calculation is initialized using only the IDPP interpolated path. The IDPP-TS method is found to have a surprisingly high convergence ratio of 97% and an average of 120 energy/force evaluations per reaction.

If a poor initial configuration is used, the TS search may converge to different, incorrect, saddle point that does not connect to the given reactant and product states. The absolute energy difference between saddle points obtained by NEB-TS and IDPP-TS and a reference set of tightly converged first order saddle points shows that by improving the initial configuration (and initial Hessian matrix) for the TS search (i.e. by using EW-CI-NEB) the ratio of SPs with large energy differences can be largely reduced.

The NEB-TS method is also applied to three complex reactions from bioinorganic chemistry and heterogeneous catalysis. In all cases, the NEB-TS calculations converged without any parameter adjustments and the saddle points were obtained in an average of 298 energy/force evaluations.

All methods addressed in this article (along with other variants of the NEB method) have been implemented in the ORCA suite of programs.[45,46] The implementation is versatile and makes efficient use of parallel computing. The interface is user-friendly while also offering a great variety of customizable features for advanced calculations. A complete description of the ORCA/NEB implementation and a list of features is given in the ORCA manual.

6 Supplementary material

The molecular configurations included in the benchmark set of 121 main-group molecular reactions are available in a .xyz file format as a supplementary material. Additional information about the set of molecular reactions, along with further analysis on the performance of the CI-NEB, EW-CI-NEB, IDPP-TS and NEB-TS methods can be found in the Supporting Information.

Acknowledgement

This work was supported by the Icelandic Research Fund and by the University of Iceland Research Fund. Moreover, V.Á. acknowledges The Doctoral Grants of the University of Iceland Research Fund. The calculations were carried out at the Icelandic Research High Performance Computing (IRHPC) center.

References

(1) Wigner, E. The transition state method, *Trans. Faraday Soc.*, **1938**, 34, pp. 29.

(2) Vineyard, G. H. Frequency factors and isotope effects in solid state rate processes, *J. Phys. Chem. Solids*, **1957**, 3, pp. 121.

(3) Mills, G.; Jónsson, H.; Schenter, G. K. Reverisble work based transition state theory: application to H_2 dissociative adsorption. *Surf. Sci.*, **1995**, 324, pp. 305.

(4) Jónsson, H.; Mills, G.; Jacobsen, K. W. Nudged elastic band method for finding minimum energy paths of transitions. 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.

(5) Ásgeirsson, V.; Jónsson, H. Exploring potential energy surfaces with saddle point searches. In *Handbook of Materials Modeling: Methods: Theory and Modeling*; Andreoni, W.; Yip, S., Eds.; Springer Cham, **2020**, pp. 689–714.

(6) Bessarab, P.F.; Uzdin, V.M.; Jónsson, H. Method for finding mechanism and activation energy of magnetic transitions, applied to skyrmion and antivortex annihilation. *Comput. Phys. Commun.*, **2015**, 196, pp. 335-347.

(7) Bohner, M.U.; Zeman, J.; Smiatek, J.; Arnold, A.; Kästner, J. Nudged-elastic band used to find reaction coordinates based on the free energy. *J. Chem. Phys.*, **2014**, 140, pp. 074109.

(8) Ásgeirsson, V.; Arnaldsson, A.; Jónsson, H. Efficient evaluation of atom tunneling combined with electronic structure calculations. *J. Chem. Phys.*, **2018**, 148, pp. 102334.

(9) Smidstrup, S.; Pedersen, A.; Stokbro, K.; Jónsson, H. Improved initial guess for minimum energy path calculations. *J. Chem. Phys.*, **2014**, 140, pp. 214106.

(10) Kästner, J.; Carr, J.M.; Keal, T.W.; Thiel, W.; Wander, A.; Sherwood, P. *J. Phys. Chem. A*, **2009**, 113, pp. 11856–11865.

(11) Zhu, X.; Thompson, K. C.; MartÃŋnez, T. J. Geodesic interpolation for reaction pathways. *J. Chem. Phys.*, **2019**, 150, pp. 164103.

(12) Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A. A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *J. Chem. Phys.*, **2004**, 120, pp. 7877.

(13) Henkelman, G. and Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.*, **2000**, 113, pp. 9978–9985.

(14) Zhu, T.; Li, J.; Samanta, A.; Kim, H.G.; Suresh, S. Interfacial plasticity governs strain rate sensitivity and ductility in nanostructured metals. *Proc. Nat. Acad. Sci.*, **2007** 104, pp. 3031–3036.

(15) Goumans, T.P.M.; Catlow, C.R.A.; Brown, W.A.; Kästner, J.; Sherwood, P. An embedded cluster study of the formation of water on interstellar dust grains. *Phys. Chem. Chem. Phys.*, **2009**, 11, pp. 5431–5436.

(16) Zarkevich, N.A.; Johnson, D.D. Nudged-elastic band method with two climbing images: Finding transition states in complex energy landscapes. *J. Chem. Phys.*, **2015**, 142, pp. 024106.

(17) Zhang, J.; Zhang, H.; Ye, H.; Zheng, Y., Free-end adaptive nudged elastic band method for locating transition states in minimum energy path calculation. *J. Chem. Phys.*, **2016**, 145, pp. 094104.

(18) Kolsbjerg, E.L.; Groves, M.N.; Hammer, B. An automated nudged elastic band method. *J. Chem. Phys.*, **2016**, 145, pp. 094107.

(19) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. Climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys*, **2000**, 113, pp. 9901–9904

(20) Cerjan, C. J.; Miller, W. H. On finding transition states. *J. Chem. Phys.*, **1981**, 75, pp. 2800–2806.

(21) Banerjee, A.; Adams, N.; Simons, J.; Shepard, R. Search for stationary points on surfaces. *J. Phys. Chem.*, **1985**, 89, pp. 52–57.

(22) Baker, J. An algorithm for the location of transition states. *J. Comput. Chem.*, **1986**, 7, pp. 385–395.

(23) Olsen, R. A.; Kroes, G. J.; Henkelman, G.; Arnaldsson, A.; JÃşnsson, H. Comparison of methods for finding saddle points without knowledge of the final states. *J. of Chem. Phys.*, **2004**, 121, pp. 9776âĂŞ9792.

(24) Schlegel, H. B. Geometry optimization. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **2011**, 1, pp. 790-80

(25) Henkelman, G. and Jónsson, H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J. Chem. Phys.*, **1999**, 111, pp. 7010–7022.

(26) Heyden, A.; Bell, A. T.; Keil, F. J. Efficient methods for finding transition states in chemical reactions: Comparison of improved dimer method and partitioned rational function optimization method. *J. Chem. Phys.*, **2005**, 123, 224101.

(27) Kästner, J.; Sherwood, P. Superlinearly converging dimer method for transition state search. *J. Chem. Phys.*,**2008**, 128, pp. 014106

(28) Gutiérrez, M. P.; Argaéz, C.; Jónsson, H. Improved Minimum Mode Following Method for Finding First Order Saddle Points. *J. Chem. Theo. Comput.*, **2017**, 13, pp. 125–134.

(29) Koistinen, O. P.; Ásgeirsson, V.; Vehtari, A.; Jónsson, H. Minimum Mode Saddle Point Searches Using Gaussian Process Regression with Inverse-Distance Covariance Function. *J. Chem. Theory Comput.*, 16, pp. 499–509.

(30) Bofill, J. M. Updated Hessian matrix and the restricted step method for locating transition structures. *J. Comput. Chem.*, **1994**, 15, pp. 1–11.

(31) Halgren, T. A.; Lipscomb, W. N. The synchronous-transit method for determining reaction pathways and locating molecular transition states. *Chem. Phys. Lett.*, **1977**, 49, pp. 225–232.

(32) Chunyang P.; Schlegel H. B. Combining Synchronous Transit and Quasi-Newton Methods to Find Transition States. *Isr. J. Chem.*, **1993**, 33, pp. 449–454.

(33) McPherson, K. E.; Bartolotti, L. J.; Morehead, A. T.; Sargent, A. L. Utility of the Nudged Elastic Band Method in Identifying the Minimum Energy Path of an Elementary Organometallic Reaction Step. *Organometallics*, **2016**, 35, pp. 1861–1865.

(34) Zimmerman, P. Reliable Transition State Searches Integrated with the Growing String Method. *J. Chem. Theory Comput.*, **2013**, 9, pp. 3043–3050.

(35) Zimmerman, P. Growing string method with interpolation and optimization in internal coordinates: Method and examples. *J. Chem. Phys.*, **2013**, 138, pp. 184102.

(36) Birkholz, A. B.; Schlegel, H. B. Using bonding to guide transition state optimization. *J. Comput. Chem.*, **2015**, 36, pp. 1157–1166.

(37) Ye, S.; Riplinger, C.; Hansen, A.; Krebs, C.; Bollinger, J. M.; Neese, F. Electronic Structure Analysis of the Oxygen-Activation Mechanism by Fe II - and α-Ketoglutarate (αKG)-Dependent Dioxygenases. *Chem. Eur. J.*, **2012**, 18, pp. 6555–6567.

(38) Lessio, M.; Dieterich, J. M.; Carter, E. A. Hydride Transfer at the GaP(110)/Solution Interface: Mechanistic Implications for CO_2 Reduction Catalyzed by Pyridine. *J. Phys. Chem. C*, **2017**, 121, pp. 17321–17331.

(39) Alberto, M. E.; Butera, V.; Russo, N. Which One among the Pt-Containing Anticancer Drugs More Easily Forms Monoadducts with G and A DNA Bases? A Comparative Study among Oxaliplatin, Nedaplatin, and Carboplatin. *Inorg. Chem.*, **2011**, 50, pp. 6965–6971.

(40) Bitzek, E.; Koskinen, P.; Gähler, F.; Moseler, M; Gumbsch, P. Structural Relaxation Made Simple. *Phys. Rev. Letters*, **2006**, 97, pp. 170201.

(41) Liu, D. C.; Nocedal, J. On the limited memory BFGS method for large scale optimization. *Math. program.*, **1989**, 45, pp. 503–528.

(42) Sheppard, D.; Terrell, R.; Henkelman, G. Optimization methods for finding minimum energy paths. *J. Chem. Phys.*, **2008**, 128, pp. 134106.

(43) Chill, S. T.; Stevenson, J.; Ruehle, V.; Shang, C.; Xiao, P.; Farrell, J. D.; Wales, D. W.; Henkelman, G. Benchmarks for Characterization of Minima, Transition States, and Pathways in Atomic, Molecular, and Condensed Matter Systems. *J. Chem. Theory Comput.*, **2014**, 10, pp. 5476–5482.

(44) Müller, K.; Brown, L. D. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theor. Chim. Acta*, **1979**, 53, pp. 75–93.

(45) Neese, F. The ORCA program system. *WIRES Comput. Mol. Sci.*, **2012**, 2, pp. 73–78.

(46) Neese, F. Software update: the ORCA program system, version 4.0. *WIRES Comput. Mol. Sci.*, **2018**, 8, e1327.

(47) Fischer, T. H.; Almlöf, J. General methods for geometry and wave function optimization. *J. Chem. Phys.*, **1992**, 96, pp. 9768–9774.

(48) Coutsias, E. A.; Seok, C.; Dill, K. A. Using quaternions to calculate RMSD. *J. Comput. Chem.*, **2004**, 25, pp. 1849–1857.

(49) Melander, M.; Laasonen, K.; Jónsson, H. Removing External Degrees of Freedom from Transition-State Search Methods using Quaternions. *J. Chem. Theory Comput.*, **2015**, 11, pp. 1055–1062.

(50) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, **1988**, 37, pp. 785–789.

(51) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, **1988**, 38, pp. 3098–3100.

(52) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, **1993**, 98, pp. 5648–5652.

(53) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, **2010**, 132, pp. 154104.

(54) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.*, **2011**, 32, pp. 1456–1465.

(55) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, **2006**, 27, pp. 1787–1799.

(56) Vahtras, O.; Almlöf, J.; Feyereisen, M. W. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.*, **1993**, 213, pp. 514–518.

(57) Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree-Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree-Fock exchange. *Chem. Phys.*, **2009**, 356, pp. 98–109.

(58) Izsák, R. and Neese, F. An overlap fitted chain of spheres exchange method. *J. Chem. Phys.*, **2011**, 135, pp. 144105.

(59) Pulay, P. Improved SCF convergence acceleration. *J.Comput. Chem.*, **1982**, 3, pp. 556–560.

(60) Kollmar, C. The role of energy denominators in self-consistent field (SCF) calculations for open shell systems. *J. Chem. Phys.*, **1996**, 105, pp. 8204–8212.

(61) Neese, F., Approximate second-order SCF convergence for spin unrestricted wavefunctions. *Chem. Phys. Lett.*, *2000*, 325(1-3), pp. 93–99.

(62) Schäfer, A.; Horn, H.; Ahlrichs, R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.*, **1992**, 97, pp. 2571–2577.

(63) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, **2005**, 7, pp. 3297.

(64) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.*, **1982**, 77, pp. 3654–3665.

(65) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta*, **1973**, 28, pp. 213–222.

(66) Schäfer, A.; Huber, C.; Ahlrichs, R., Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys.*, **1994**, 100, pp. 5829–5835

(67) Bergner, A.; Dolg, M.; Küchle, W.; Stoll, H.; Preuß, H. Ab initio energy-adjusted pseudopotentials for elements of groups 13-17. *Mol. Phys.*, **1993**, 80, pp. 1431–1441.

(68) Leininger, T.; Berning, A.; Nicklass, A.; Stoll, H.; Werner, H.-J.; Flad, H.-J. Spin-orbit interaction in heavy group 13 atoms and TlAr. *Chem. Phys.*, **1997**, 217, pp. 19–27.

(69) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials. *Theor. Chem. Acc.*, **1997**, 97, pp. 119–124.

(70) Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.*, **2006**, 8, pp. 1057–1065.

(71) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A*, **1998**, 102, pp. 1995–2001.

(72) York, D. M.; Karplus, M. A Smooth Solvation Potential Based on the Conductor-Like Screening Model. *J. Phys. Chem. A*, **1999**, 103, pp. 11060–11079.

(73) Garcia-Ratés; M., and Neese, F. Effect of the Solute Cavity on the Solvation Energy and its Derivatives within the Framework of the Gaussian Charge Scheme. *J. Comput. Chem.*, **2019**, 41, pp. 922–939.

(74) Fukui, K. The Path of Chemical Reactions - The IRC Approach. *Acc. Chem. Res.*, **1981**, 14, pp. 363–368.

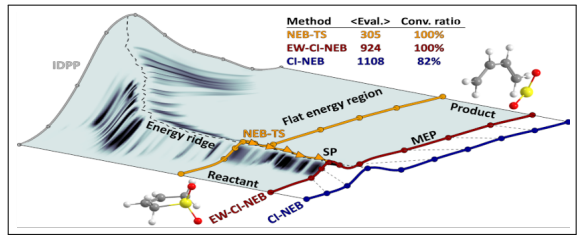
(75) Forrie, G.M, Valleau, J.P. Nonphysical sampling distributions in Monte Carlo free-energy estimations: Umbrella sampling. *J. Comput. Phys.*, **1977**, 2, pp. 187-199.

(76) Bolhuis, P.G., Chandler, D., Dellago, C. and Geissler, P.L. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.*, **2002**, 53(1), pp.291-318.

(77) Lessio, M. (University of Sydney Nano Institute), private communication, **2019**.

(78) Ye, S. (Max-Planck-Institut für Kohlenforschung), private communication, **2019**.

Graphical TOC Entry



Supporting Information: Nudged elastic band method for molecular reactions using energy-weighted springs combined with eigenvector following

Vilhjálmur Ásgeirsson,[†] Benedikt Orri Birgisson,[†] Ragnar Bjornsson,[‡] Ute Becker,[¶] Frank Neese,[¶] Christoph Riplinger,[§] and Hannes Jónsson[*,†]

[†]*Science Institute and Faculty of Physical Sciences, University of Iceland VR-III, 107 Reykjavík, Iceland*

[‡]*Max-Planck-Institute for Chemical Energy Conversion, Mülheim an der Ruhr 45470, Germany.*

[¶]*Max-Planck-Institute for Kohlenforschung, Mülheim an der Ruhr 45470, Germany*

[§]*FaccTs GmbH, Rolandstrasse 67, 50677 Cologne, Germany*

E-mail: hj@hi.is

SI-1 A large and diverse benchmark set of molecular reactions

In this section, the benchmark set of 121 main-group molecular reactions used to evaluate the (EW-)CI-NEB and NEB-TS methods is introduced.[1,2] The reactant, product and saddle point configurations are visualized. Modifications made to the reactant and product configurations are addressed. The potential energy barrier height, absolute imaginary frequency at the first order saddle points and reaction energy are given. As discussed in the manuscript, all calculations use B3LYP+D3(BJ)/def2-SVP level of theory.

SI-1.1 Visualization of reactant and product configurations

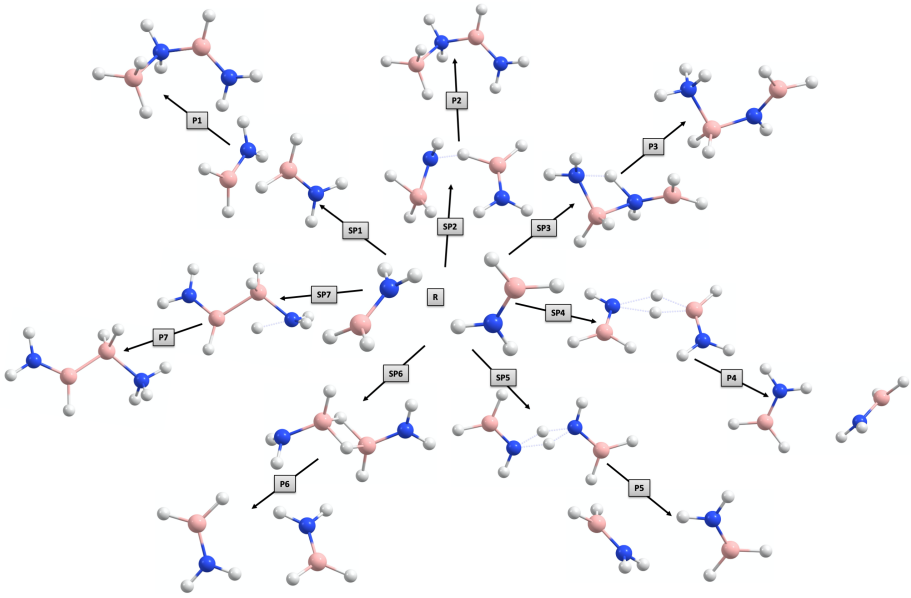


Figure 1: Reactant, saddle point and product configurations for reactions 1–7.

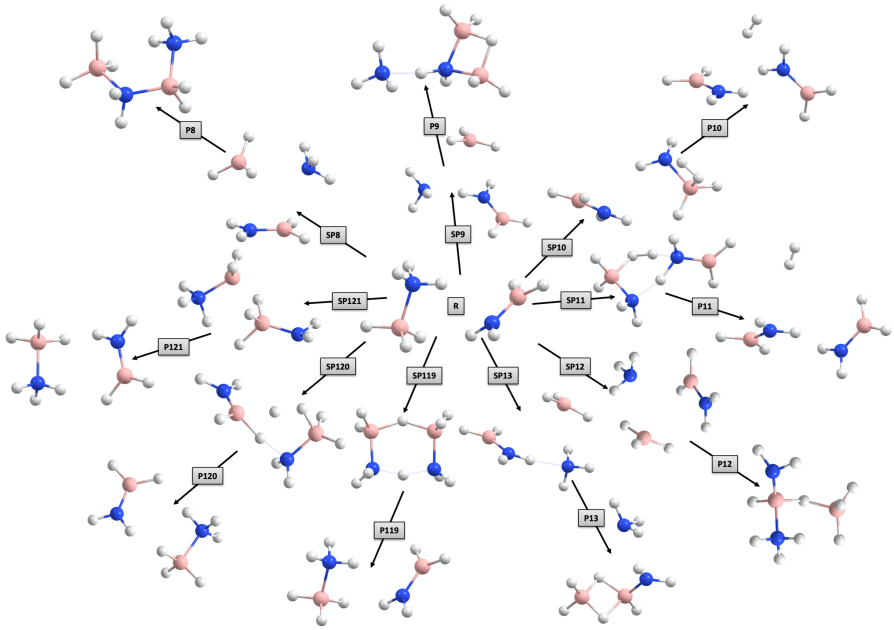


Figure 2: Reactant, saddle point and product configurations for reactions 8–13 and 119–121.

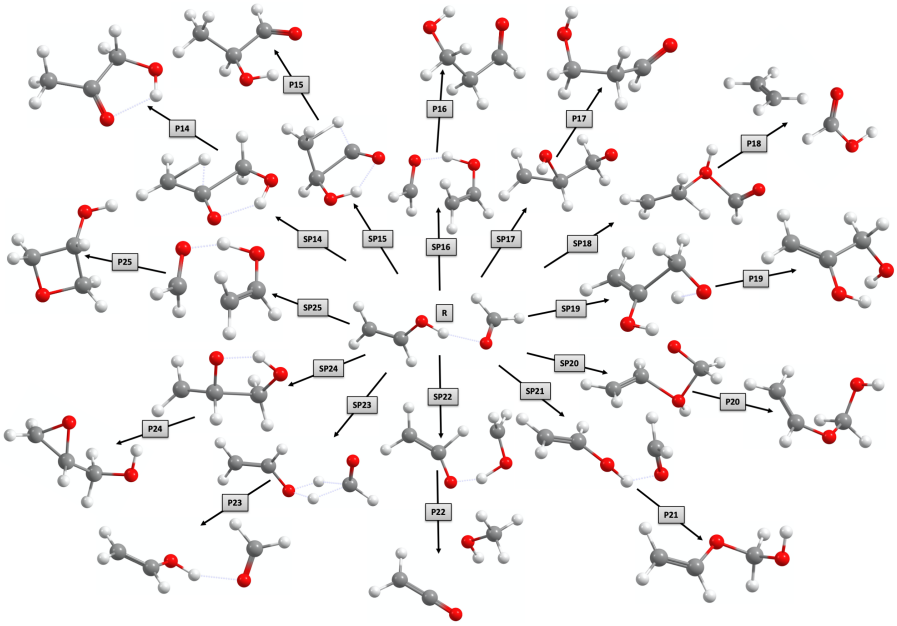


Figure 3: Reactant, saddle point and product configurations for reactions 14–25.

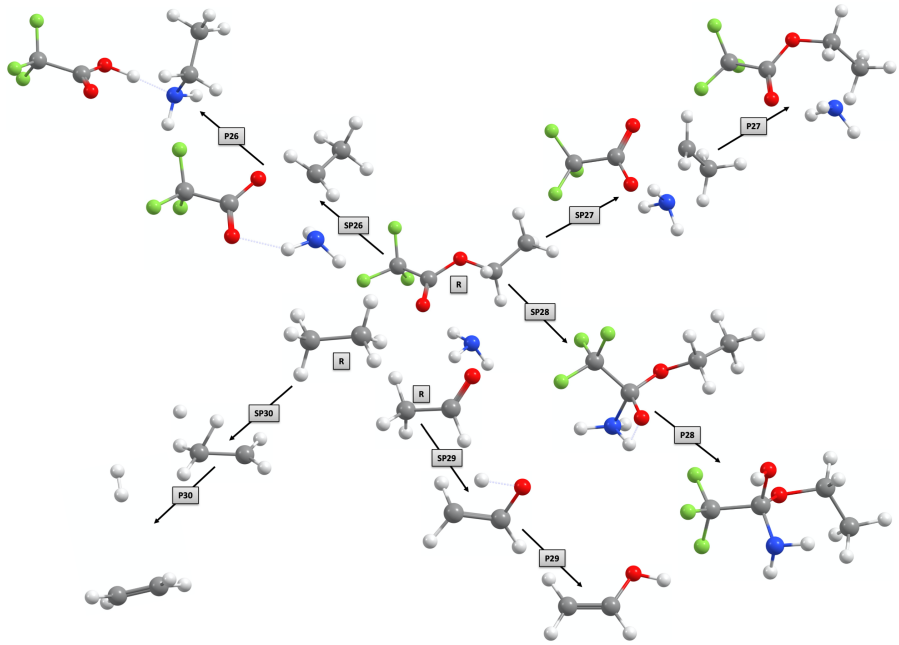


Figure 4: Reactant, saddle point and product configurations for reactions 26–30.

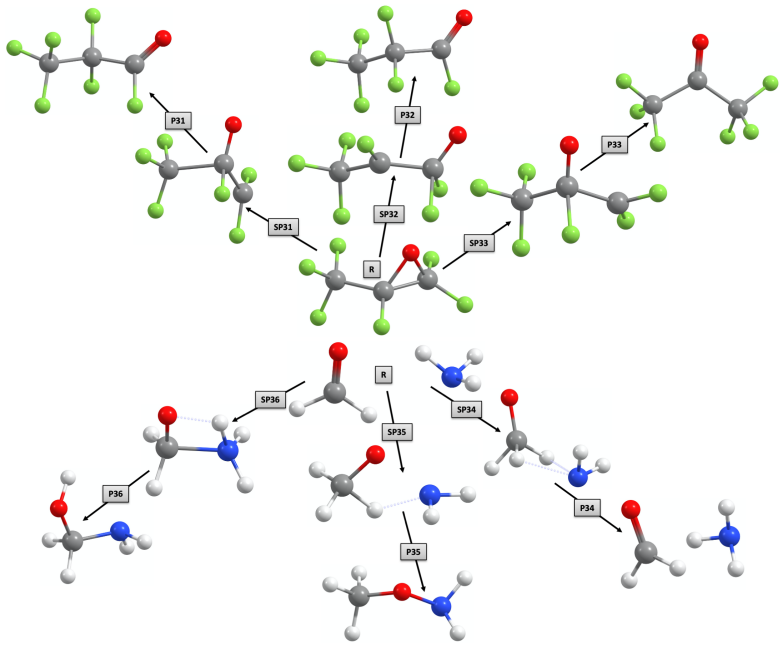


Figure 5: Reactant, saddle point and product configurations for reactions 31–36.

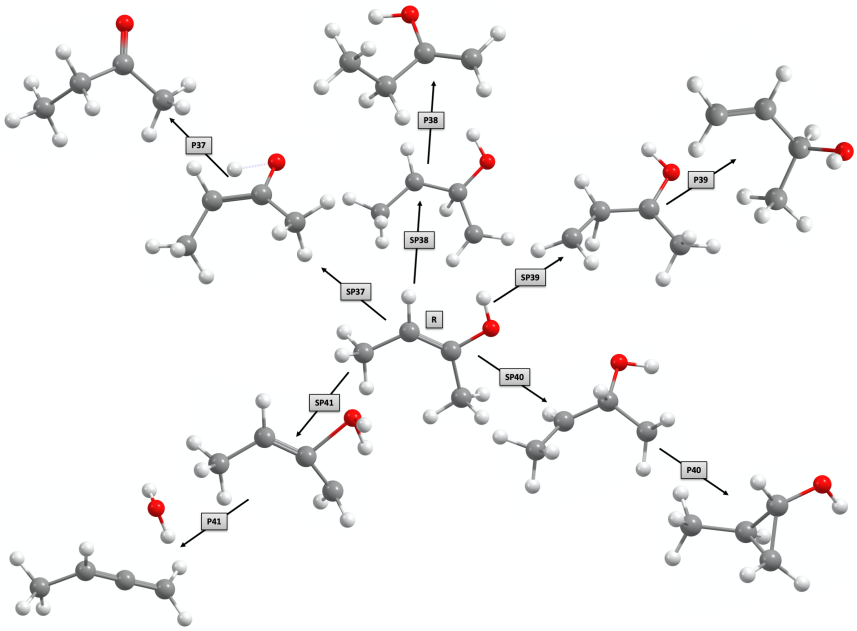


Figure 6: Reactant, saddle point and product configurations for reactions 37–41.

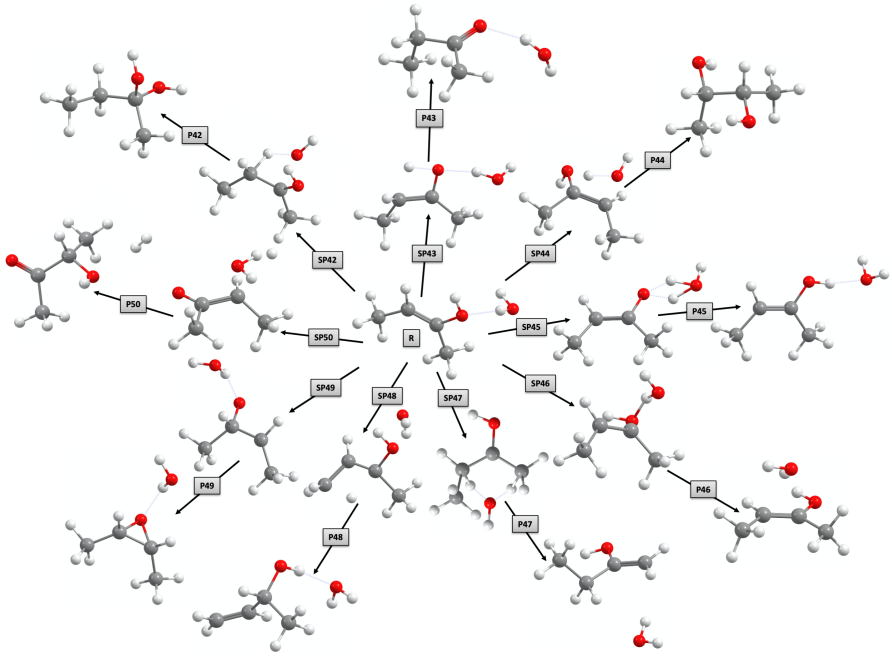


Figure 7: Reactant, saddle point and product configurations for reactions 42–50.

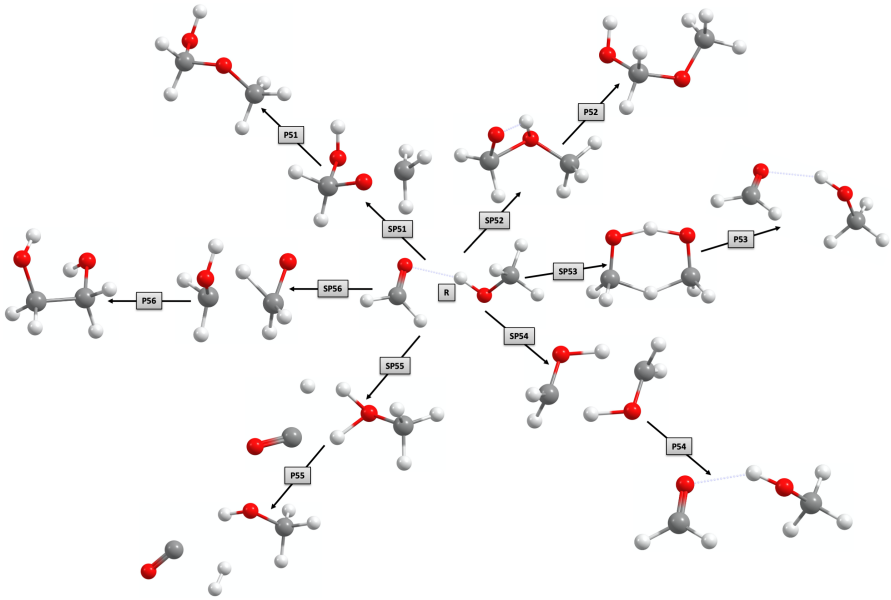


Figure 8: Reactant, saddle point and product configurations for reactions 51–56.

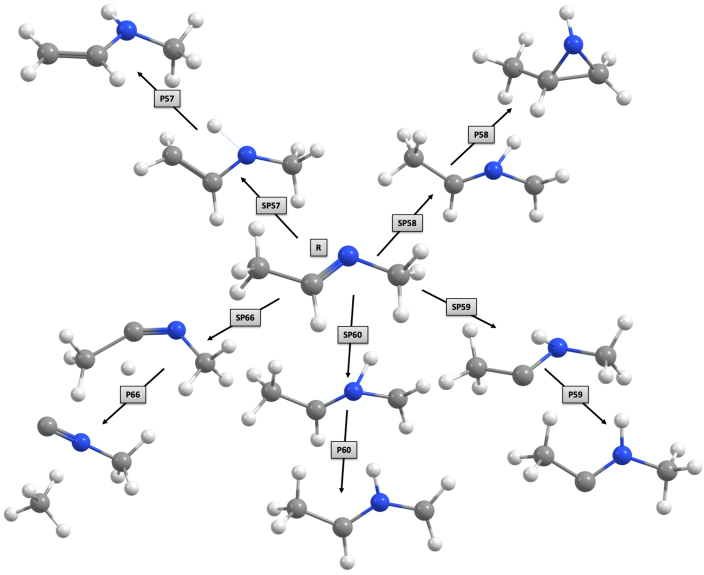


Figure 9: Reactant, saddle point and product configurations for reactions 57–59, 60 and 66.

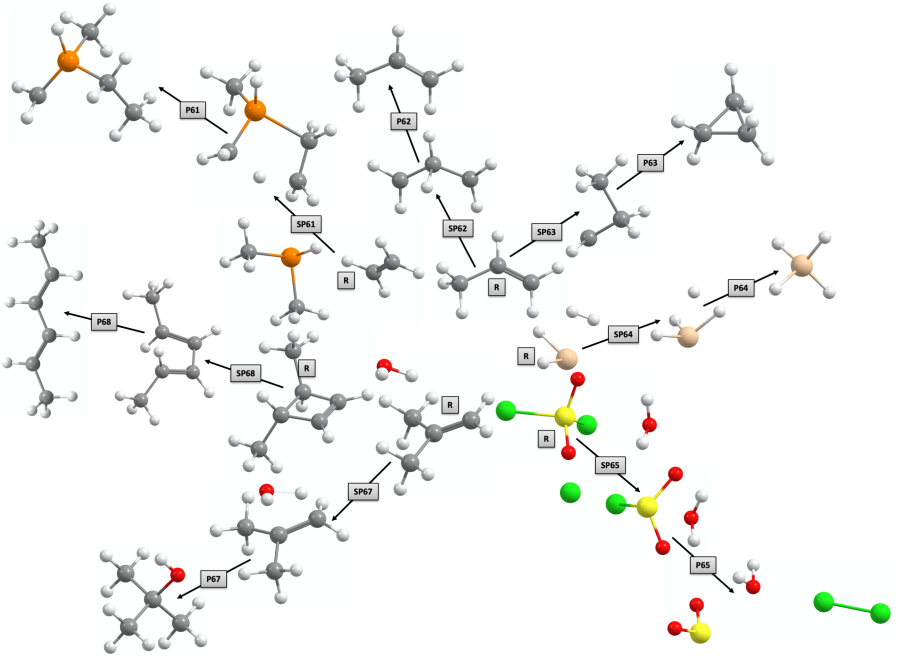


Figure 10: Reactant, saddle point and product configurations for reactions 61–65, 67 and 68.

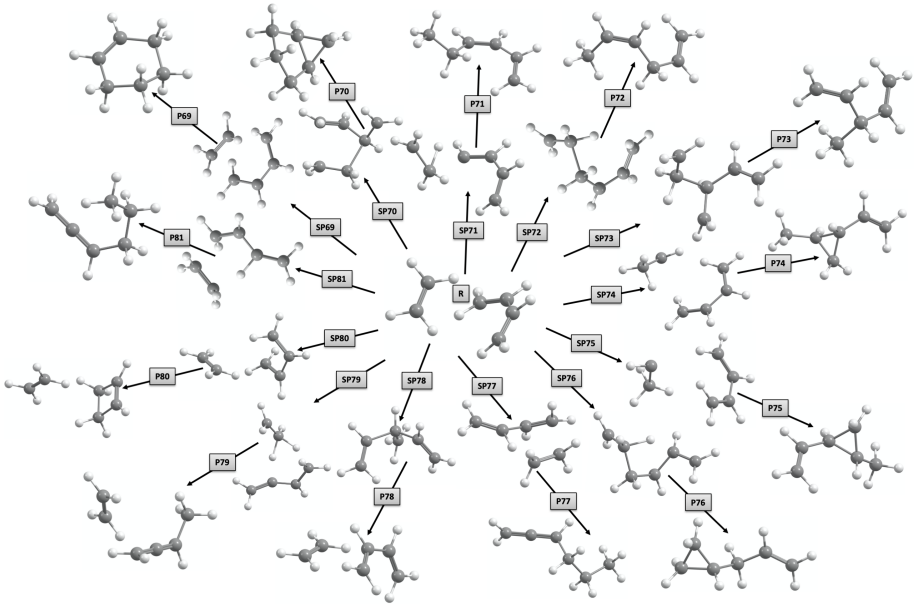


Figure 11: Reactant, saddle point and product configurations for reactions 69–81.

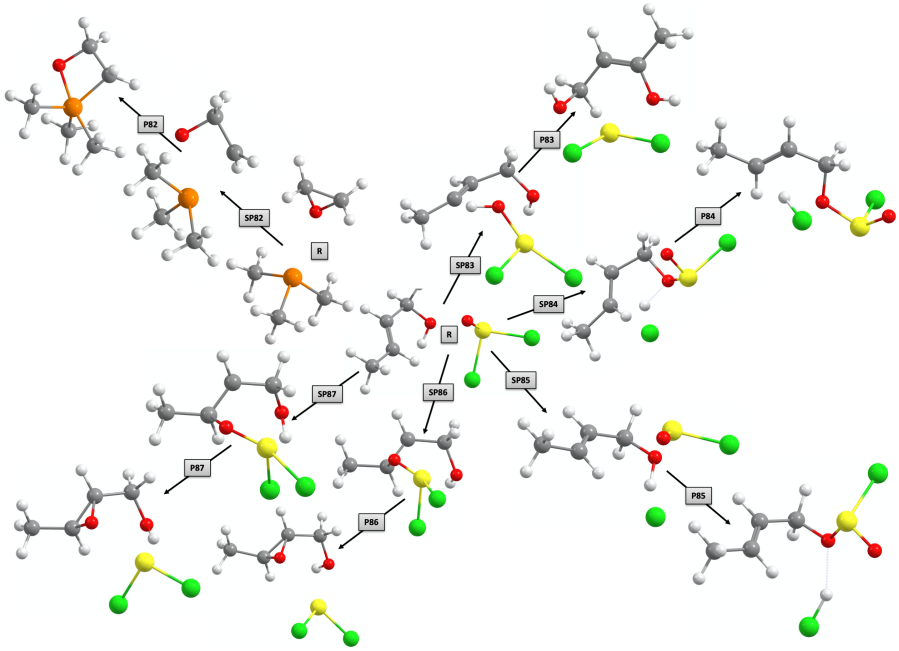


Figure 12: Reactant, saddle point and product configurations for reactions 82–87.

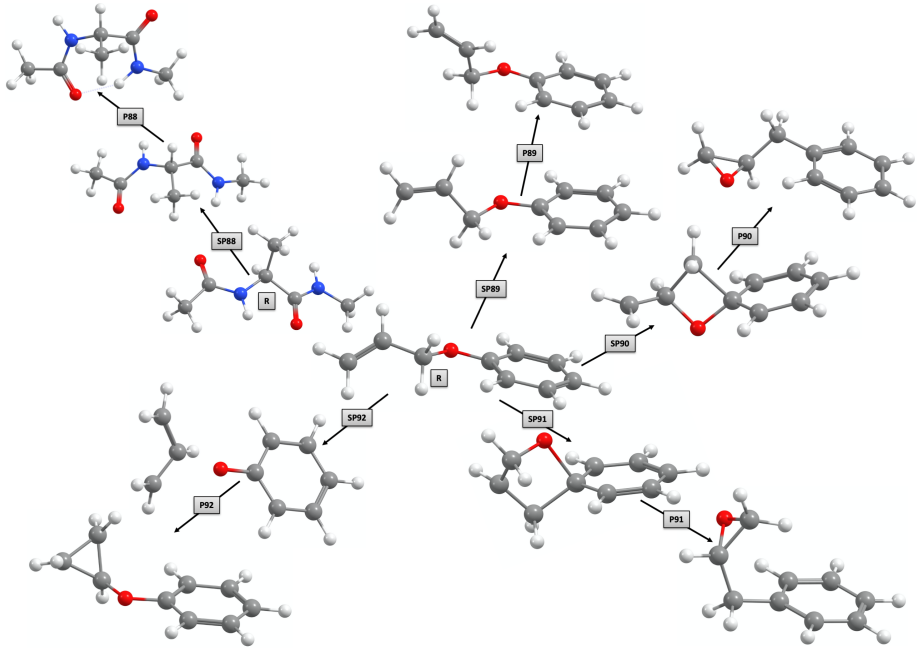


Figure 13: Reactant, saddle point and product configurations for reactions 88–92.

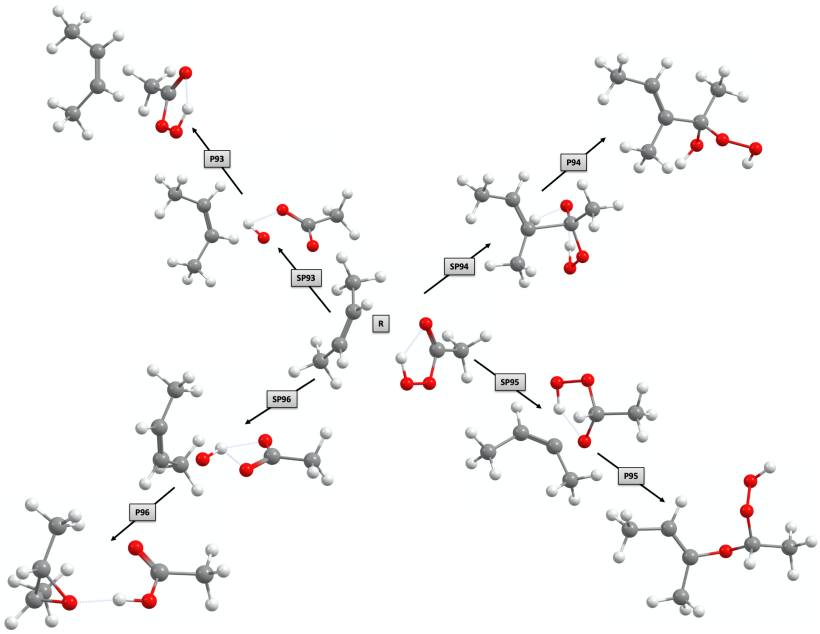


Figure 14: Reactant, saddle point and product configurations for reactions 89–92.

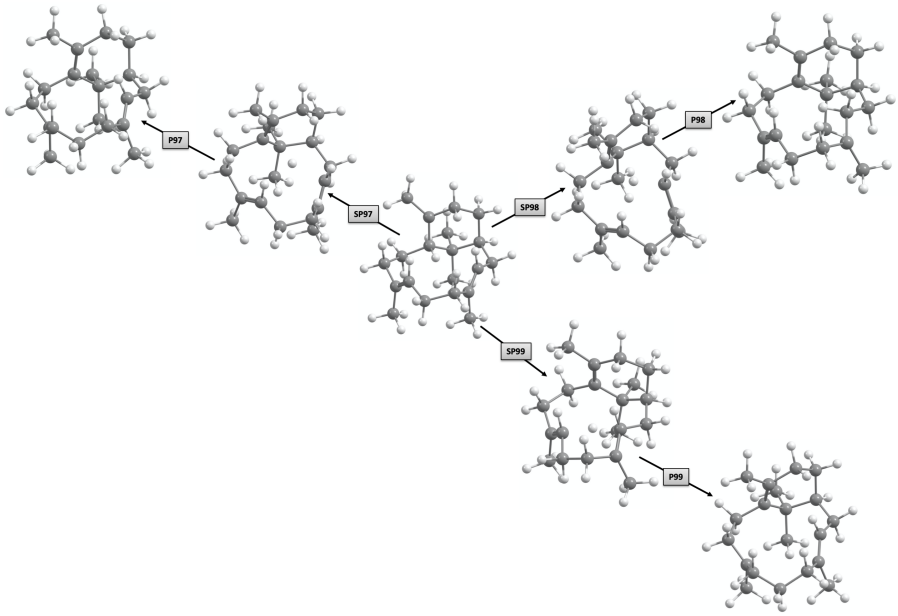


Figure 15: Reactant, saddle point and product configurations for reactions 93–96.

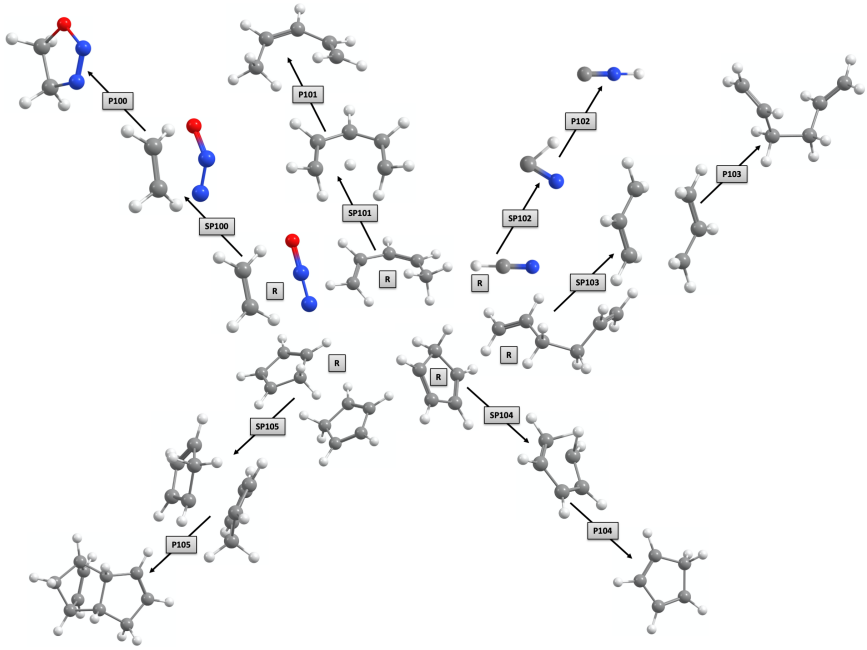


Figure 16: Reactant, saddle point and product configurations for reactions 97–99.

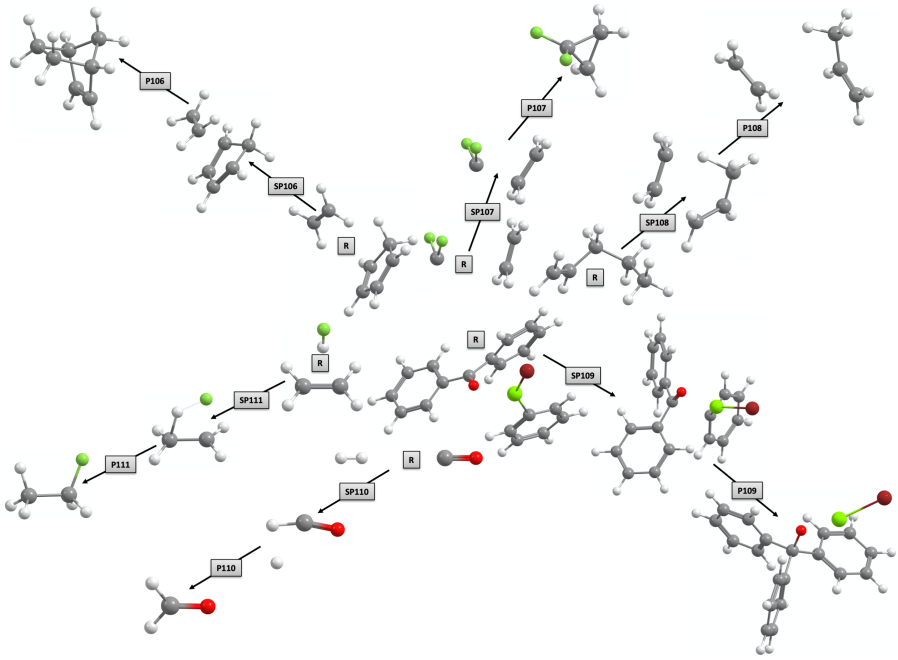


Figure 17: Reactant, saddle point and product configurations for reactions 100–105.

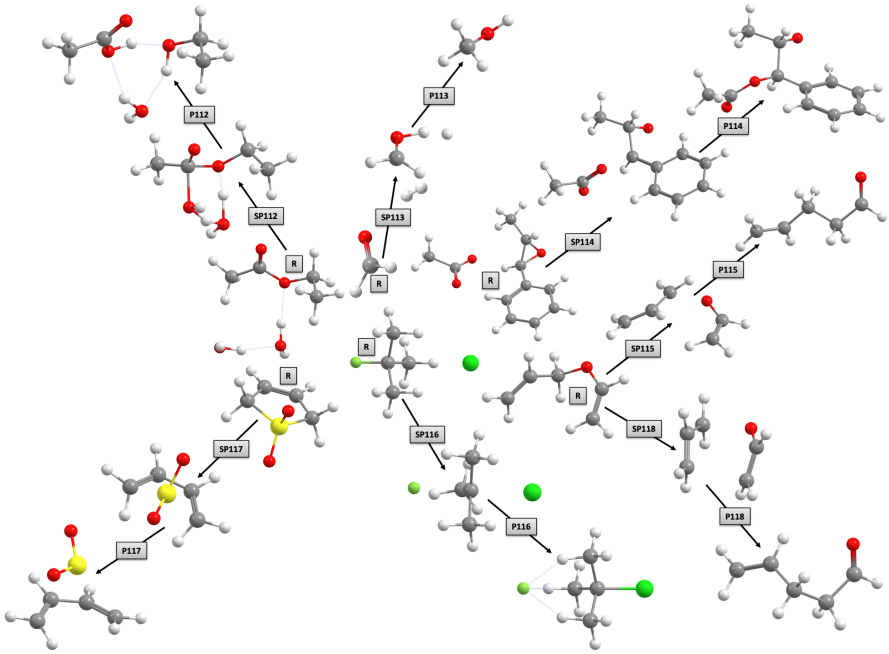


Figure 18: Reactant, saddle point and product configurations for reactions 106–111.

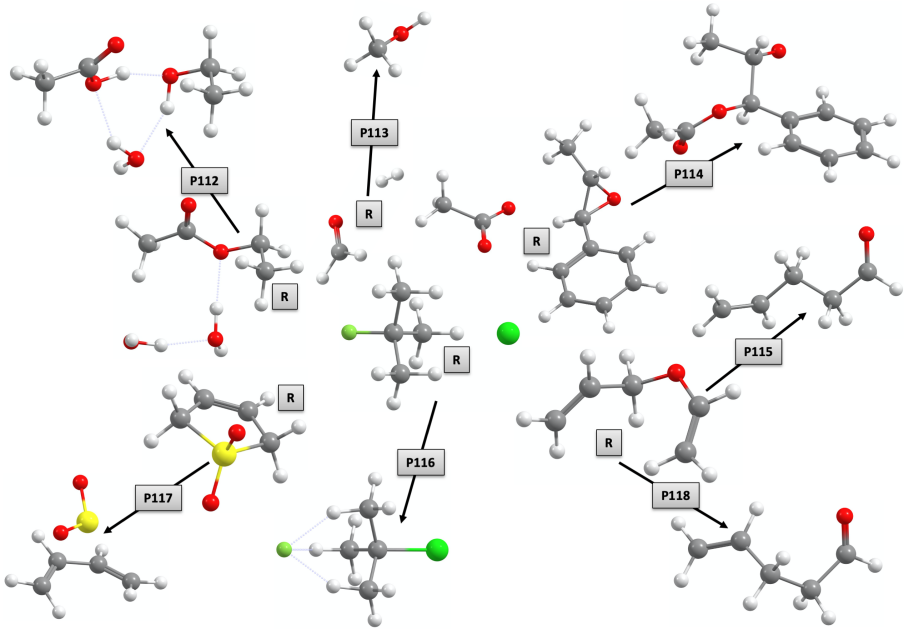
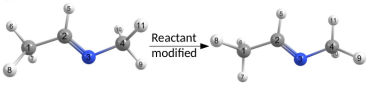
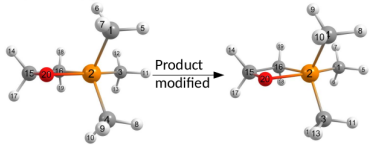


Figure 19: Reactant, saddle point and product configurations for reactions 112–118.

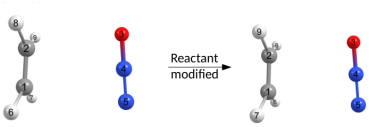
SI-1.2 Modifications of reactant and/or product configurations



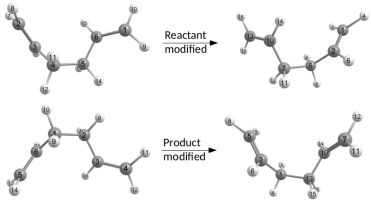
(a) Reaction 58



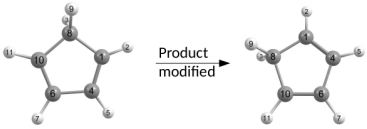
(b) Reaction 82



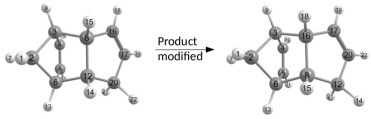
(c) Reaction 100



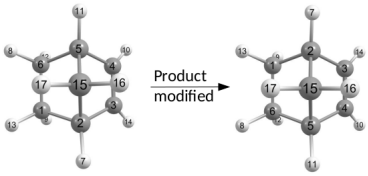
(d) Reaction 103



(e) Reaction 104



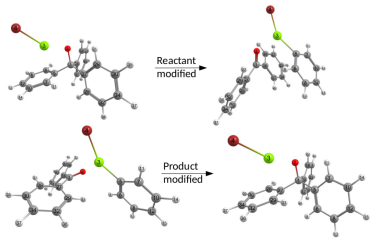
(f) Reaction 105



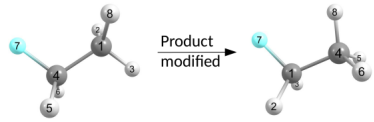
(g) Reaction 106

(h) Reaction 108

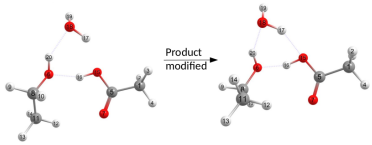
Figure 20: Modifications made to the reactant or product configurations of the original benchmark set of main-group molecular reactions.[1,2]



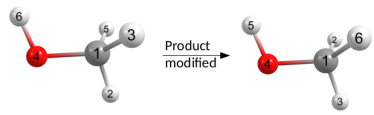
(a) Reaction 109



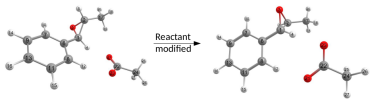
(b) Reaction 111



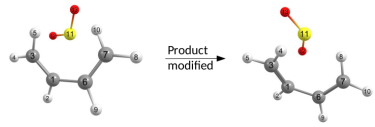
(c) Reaction 112



(d) Reaction 113



(e) Reaction 114



(f) Reaction 117

Figure 21: Modifications made to the reactant or product configurations of the original benchmark set of main-group molecular reactions. [1,2]

SI-1.3 Energy barrier, reaction energy and imaginary frequency

In this section, the potential energy barrier, reaction energy and imaginary frequency obtained at the first order saddle points are given for the large benchmark set of main-group molecular reactions.

Table 1: Potential energy barrier height, reaction energy and the absolute imaginary frequency at the saddle point for reactions 1–30

Rxn No.	$E^{\ddagger}$ [kcal/mol]	E_{rxn} [kcal/mol]	$\lvert\omega_0\rvert$ [cm^{-1}]
1	9.2	-6.3	343.5
2	88.2	-6.3	936.9
3	28.6	-4.5	1246.3
4	91.2	0.0	2450.5
5	67.6	0.0	2576.7
6	19.7	0.2	432.5
7	53.8	16.4	1486.6
8	35.9	-11.3	219.3
9	32.4	-2.7	169.1
10	42.1	3.3	1382.9
11	37.0	3.6	1337.6
12	36.6	11.7	217.1
13	32.5	17.9	158.7
14	56.4	-29.1	1217.2
15	50.6	-24.6	1217.5
16	12.2	-23.2	839.3
17	98.3	-19.4	592.6
18	79.4	-18.2	1081.4
19	78.8	-14.5	2144.9
20	67.1	-9.3	723.5
21	38.7	-8.6	1734.7
22	31.1	-1.6	1339.0
23	82.7	0.0	2854.8
24	62.6	3.6	774.6
25	72.1	2.7	1271.0
26	55.5	-12.9	535.5
27	48.5	0.0	506.0
28	37.1	3.6	1557.1
29	68.3	13.0	2139.1
30	120.2	41.4	2023.1

Table 2: Potential energy barrier height, reaction energy and the absolute imaginary frequency at the saddle point for reactions 31–60

Rxn No.	$E^{\ddagger}$ [kcal/mol]	E_{rxn} [kcal/mol]	$\lvert\omega_0\rvert$ [cm^{-1}]
31	54.6	-25.3	244.9
32	54.1	-25.3	402.1
33	46.8	-20.4	287.7
34	111.0	3.7	1306.7
35	99.4	28.2	1729.4
36	32.2	-9.6	1507.7
37	53.9	-11.4	2110.1
38	83.7	2.8	1541.1
39	82.5	9.3	601.9
40	87.6	17.1	397.8
41	81.6	31.9	1604.1
42	49.7	-16.6	1754.5
43	52.1	-13.3	2097.4
44	66.8	-5.4	2114.1
45	41.7	0.1	2108.7
46	68.4	2.8	1268.7
47	48.7	5.3	1958.0
48	83.9	9.1	1676.4
49	71.1	13.0	543.6
50	100.0	12.1	1471.9
51	67.9	-12.2	643.3
52	34.2	-9.8	1634.6
53	28.5	0.0	1489.4
54	48.1	0.0	1854.2
55	65.7	13.4	1950.7
56	81.1	-13.5	2020.1
57	65.6	5.0	1974.1
58	71.5	15.4	1666.0
59	91.7	33.6	1455.1
60	71.5	33.4	1666.2

Table 3: Potential energy barrier height, reaction energy and the absolute imaginary frequency at the saddle point for reactions 61-90

Rxn No.	$E^{\ddagger}$ [kcal/mol]	E_{rxn} [kcal/mol]	$\lvert\omega_0\rvert$ [cm^{-1}]
61	46.8	11.8	1337.4
62	82.0	0.0	1290.0
63	82.4	7.7	112.3
64	5.4	-52.6	1133.1
65	54.1	0.5	433.2
66	123.2	20.0	1371.0
67	50.5	-15.4	1997.4
68	31.2	-15.2	519.8
69	15.8	-48.3	516.7
70	56.5	-37.1	727.0
71	91.1	-27.4	1878.0
72	38.2	-25.8	434.4
73	67.5	-23.9	709.4
74	75.3	-21.1	92.0
75	75.0	-20.8	47.7
76	81.5	-18.6	1408.4
77	36.5	-16.5	925.2
78	124.3	0.2	1735.7
79	64.2	8.0	1549.4
80	44.0	9.1	719.1
81	36.5	-16.9	924.3
82	48.4	-11.8	418.8
83	80.7	-11.6	889.5
84	52.5	-0.6	1596.4
85	35.2	0.2	392.1
86	52.5	4.4	505.7
87	48.9	0.8	525.0
88	6.3	0.2	44.8
89	2.7	0.6	68.6
90	87.5	4.3	442.9

Table 4: Potential energy barrier height, reaction energy and the absolute imaginary frequency at the saddle point for reactions 91-121

Rxn No.	$E^{\ddagger}$ [kcal/mol]	E_{rxn} [kcal/mol]	$\lvert\omega_0\rvert$ [cm^{-1}]
91	84.6	3.9	778.8
92	75.1	4.0	571.5
93	53.1	0.1	558.7
94	73.5	8.2	1758.9
95	98.5	14.5	1387.3
96	17.1	-59.1	424.5
97	48.3	1.7	667.1
98	29.0	2.4	2132.7
99	28.1	-0.6	2127.8
100	25.6	-5.3	466.8
101	32.9	0.0	1549.5
102	47.8	13.6	1116.6
103	32.6	0.2	526.5
104	26.6	0.0	1194.2
105	15.2	-19.1	440.8
106	17.1	-25.6	488.4
107	8.8	-53.9	405.7
108	55.5	25.9	1156.6
109	18.3	-12.8	205.0
110	74.0	-10.4	1880.6
111	44.1	-16.1	1845.9
112	35.5	-5.3	1264.4
113	75.9	-25.1	1385.4
114	21.6	11.1	304.7
115	62.8	-18.0	401.3
116	45.4	29.5	391.1
117	18.9	2.4	348.2
118	31.6	-18.3	434.1
119	2.4	-4.1	861.0
120	52.6	0.0	1492.0
121	39.0	-4.1	916.4

SI-2 Data relevant to the (EW-)CI-NEB method

SI-2.1 Comparison of linear interpolation in Cartesian coordinates and the IDPP method

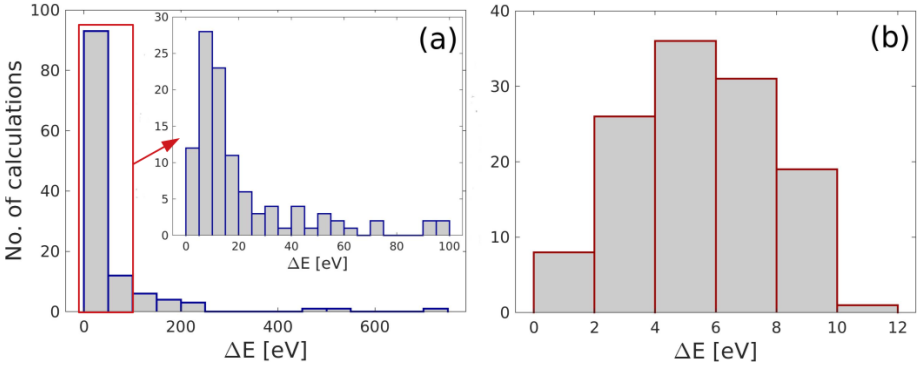


Figure 22: The maximum rise in energy ($\Delta E = E^{\mathrm{HEI}} - E^{\mathrm{reactant}}$) along initial paths generated by linear interpolation in Cartesian coordinates (shown in (a)) and by the IDPP method (shown in (b)) for the large benchmark set of main-group molecular reactions. For clarity the distribution of $\Delta E \in (0, 100]$ kcal/mol is shown as an inset in (a).

SI-2.2 Path lengths obtained by (EW-)CI-NEB

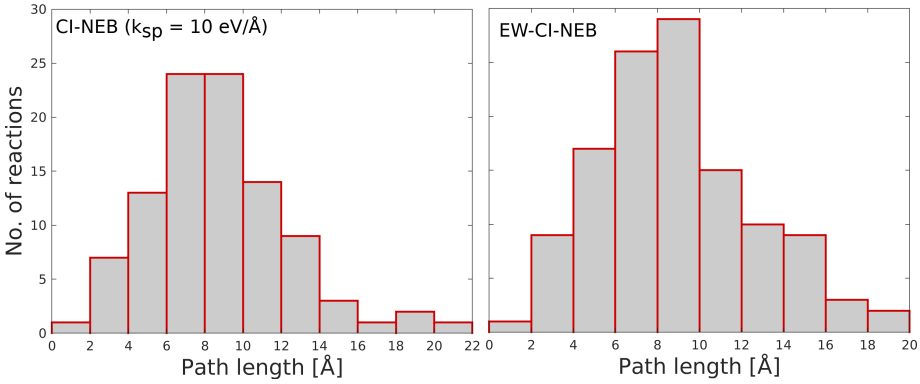


Figure 23: Distribution of path lengths of converged CI-NEB and EW-CI-NEB calculations, using $N_{\mathrm{im}} = 10$.

SI-2.3 Evolution of inter-image distance in (EW-)CI-NEB

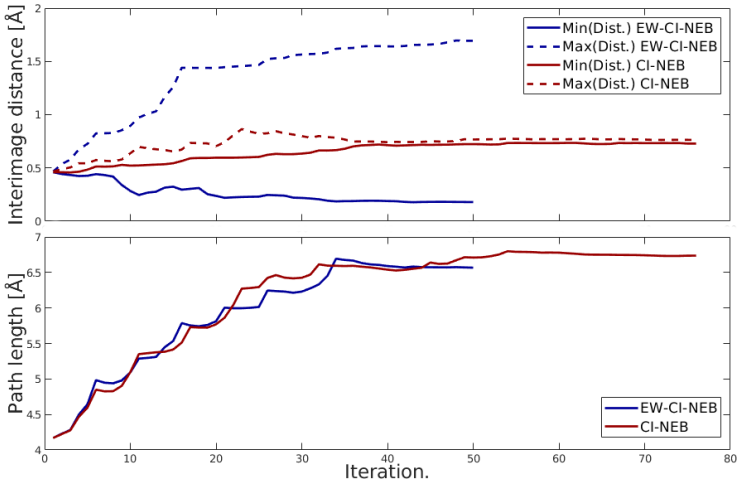


Figure 24: Features of the reaction path in CI-NEB and EW-CI-NEB calculations on the ene-reaction of 1-propylene and ethylene. In the upper panel, the evolution of the maximum and minimum inter-image distance along the reaction path during the CI-NEB (red) and EW-CI-NEB (blue) optimization. In the lower panel, the evolution of the path length. The calculations are started from an IDPP initial path.

SI-2.4 Inter-image distances and angle distribution in (EW-)CI-NEB

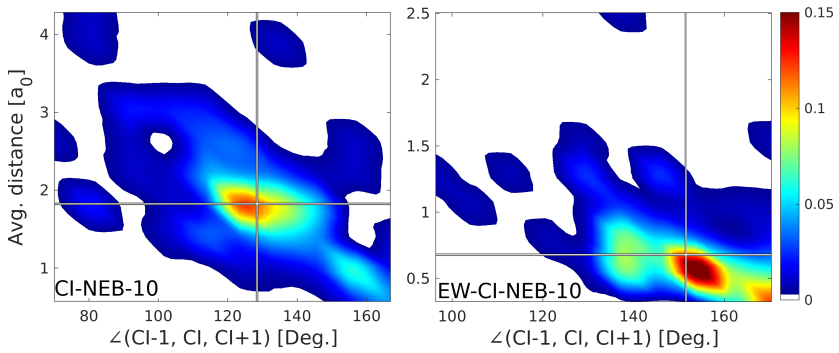


Figure 25: Bivariate distribution of average inter-image distances of CI to its two neighboring images and the angle formed by the three images along the converged path for CI-NEB and EW-CI-NEB calculations on the large benchmark set of main-group molecular reactions.

To further investigate the image distribution along the reaction paths obtained by CI-NEB and EW-CI-NEB calculations on the large benchmark set of main-group molecular reactions, the deviation from 'ideality' is computed. For this purpose, the reaction is partitioned into two segments, to the left and right of CI. The deviation from the ideal even distribution is then computed,

$$\gamma_{\text{left}} = \frac{|\mathbf{R}_{\text{CI}} - \mathbf{R}_{\text{CI}-1}|}{\sum_{i=0}^{\text{CI}-1} |\mathbf{R}_{i+1} - \mathbf{R}_i|}$$
$$\gamma_{\text{right}} = \frac{|\mathbf{R}_{\text{CI}+1} - \mathbf{R}_{\text{CI}}|}{\sum_{i=\text{CI}}^{N} |\mathbf{R}_{i+1} - \mathbf{R}_i|}$$

Then, γ is selected as the interval that exhibits a larger deviation from an even distribution (i.e., $\gamma = 1.0$). The results are shown in Fig. 26.

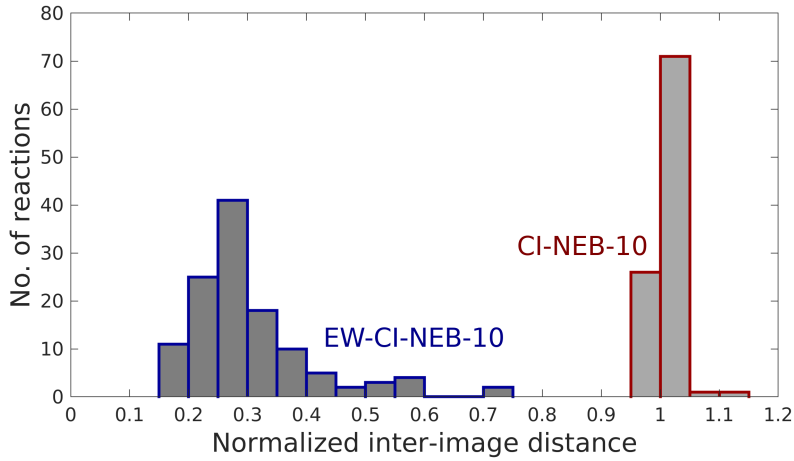


Figure 26: The deviation of the inter-image distance of CI to its neighboring images from an ideal even distribution along the reaction path. The results are shown for CI-NEB and EW-CI-NEB calculations on the large benchmark set of molecular reactions.

SI-2.5 Optimization profile to monitor (EW-)CI-NEB calculations

In this study, we present a new tool to visualize and monitor (EW-)CI-NEB calculations and call it an optimization profile. In this scheme, the path at every optimization step is interpolated using a piecewise-cubic polynomial[3] and plotted along with the position of the intermediate images. This allows us to visualize how the path/images 'slide down' on the energy surface towards the MEP. Optimization profiles may also reveal whether the path may be kinked, calculation may become non-convergent or if the optimization has become unstable. Furthermore, an optimization profile will also reveal whether an intermediate energy minima is to be found along the path. In such cases, it may be the best choice of action to halt the calculation, locate the intermediate energy minimum and carry out CI-NEB calculation for the two path fragments, i.e. from reactant to intermediate state and from intermediate to product state. The tool to generate an optimization profile from a NEB calculation in ORCA is available in Ref.[4]

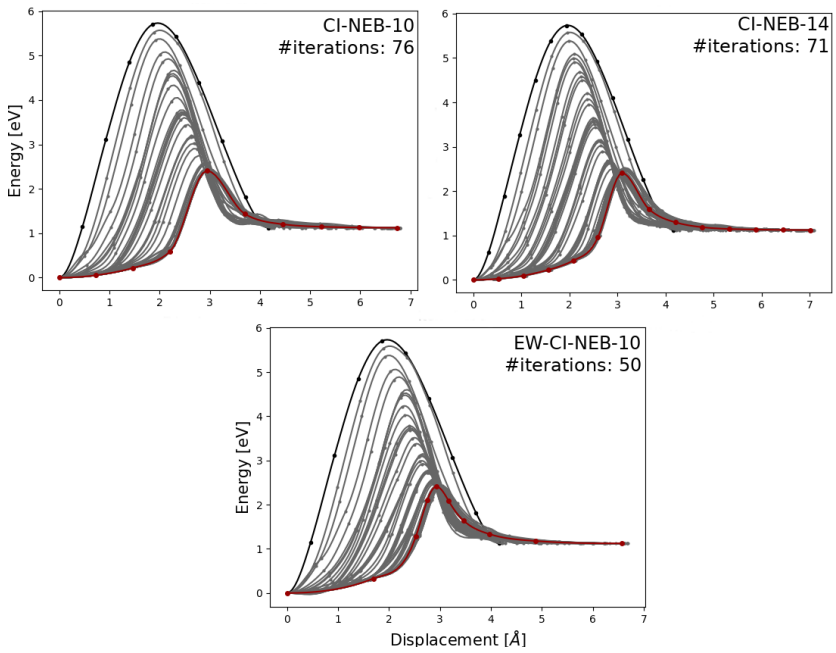


Figure 27: Optimization profiles for CI-NEB and EW-CI-NEB calculations using $N_{\text{im}} = 10$ and 14 of the ene-reaction of 1-propylene and ethylene.

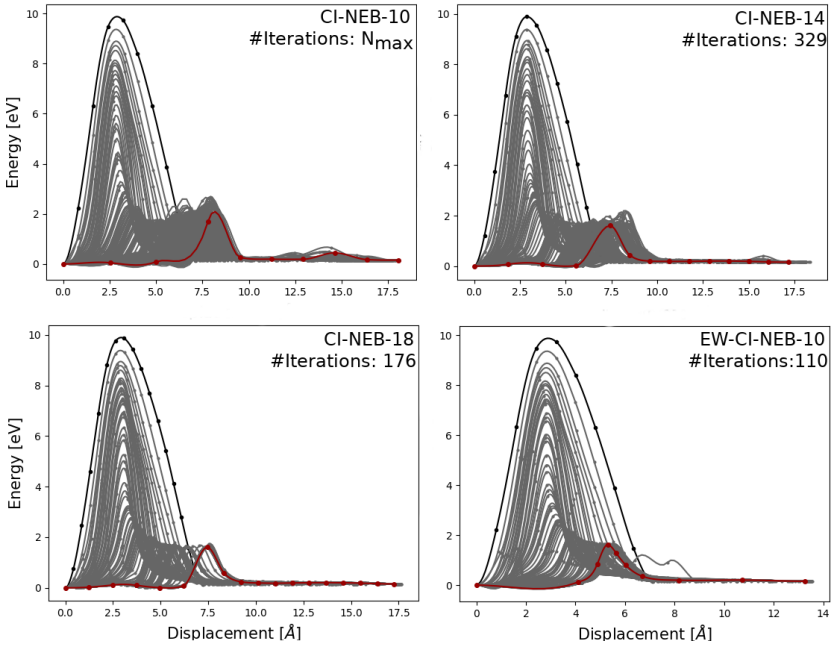


Figure 28: CI-NEB and EW-CI-NEB optimization profiles for the reaction $BH_3NH_3 + BH_2NH_2 \rightarrow 2BH_2NH_2 + H_2$ using $N_{\text{im}} = 10, 14$ and 18.

SI-2.6 Convergence behavior of CI and TS

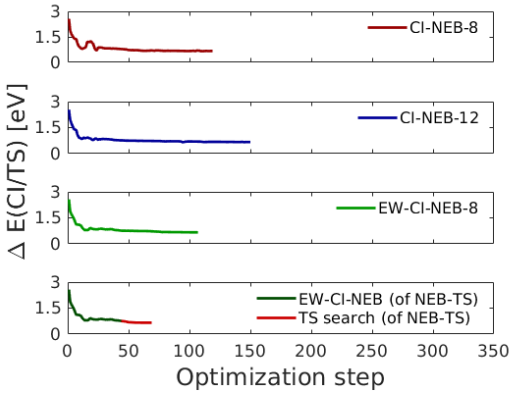


Figure 29: The relative energy of CI and TS as a function of optimization step for CI-NEB, EW-CI-NEB and NEB-TS calculations, using both $N_{\mathrm{im}} = 8$ and 12, of the Diels-Alder addition of two cyclopentadienes.

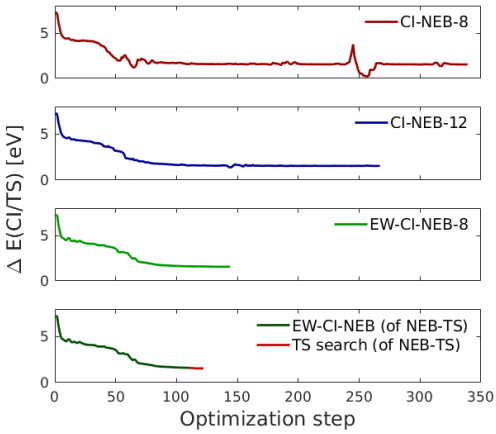


Figure 30: The relative energy of CI and TS as a function of optimization steps for CI-NEB, EW-CI-NEB and NEB-TS calculations, using both $N_{\mathrm{im}} = 8$ and 12, of the hydrolysis reaction of ethyl acetate.

SI-3 Additional data for NEB-TS and IDPP-TS methods

SI-3.1 Energy deviation of NEB-TS (and IDPP-TS) saddle points from reference set of saddle points

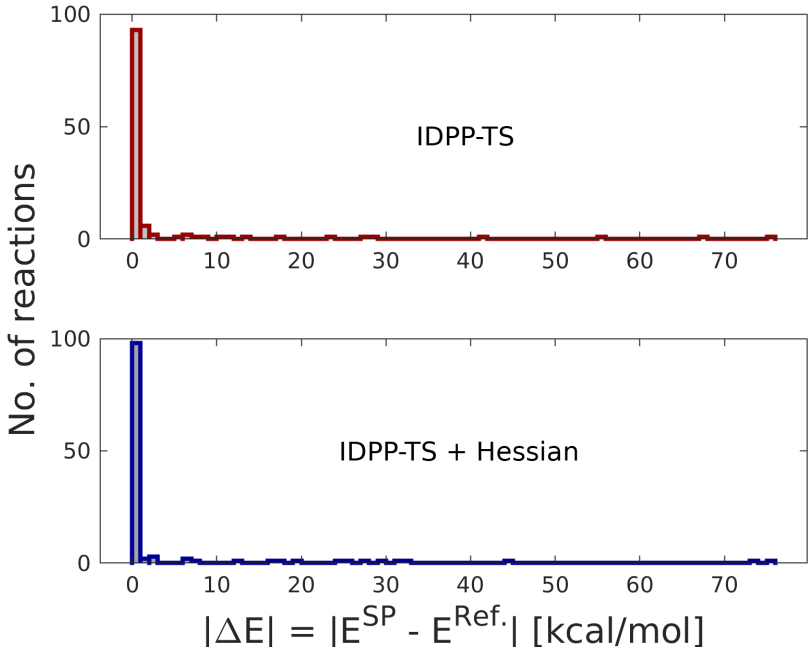


Figure 31: Energy deviation of saddle points obtained by IDPP-TS from the reference set of saddle points. In the upper panel, the TS search is started using a modified empirical Hessian matrix. In the lower panel, the TS search is started using the exact Hessian matrix. In both sets of calculations, the initial configuration of the TS search is taken as the highest energy image along an IDPP initial path.

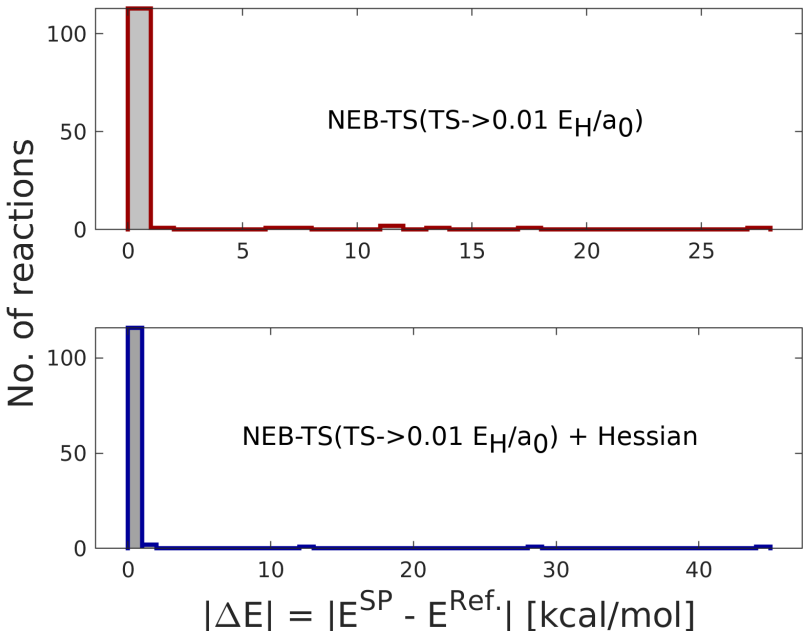


Figure 32: Energy deviation of saddle points obtained by NEB-TS from the reference set of saddle points. The NEB-TS calculations use a TS activation threshold of $\epsilon_{max}^{\mathrm{TS}} = 0.01\ \mathrm{E}_H/\mathrm{a}_0$. In the upper panel, the TS search is started using a modified empirical Hessian matrix. In the lower panel, the TS search is started using the exact Hessian matrix. In both sets of calculations, the initial configuration of the TS search is taken as the climbing image from a partially converged EW-CI-NEB calculation.

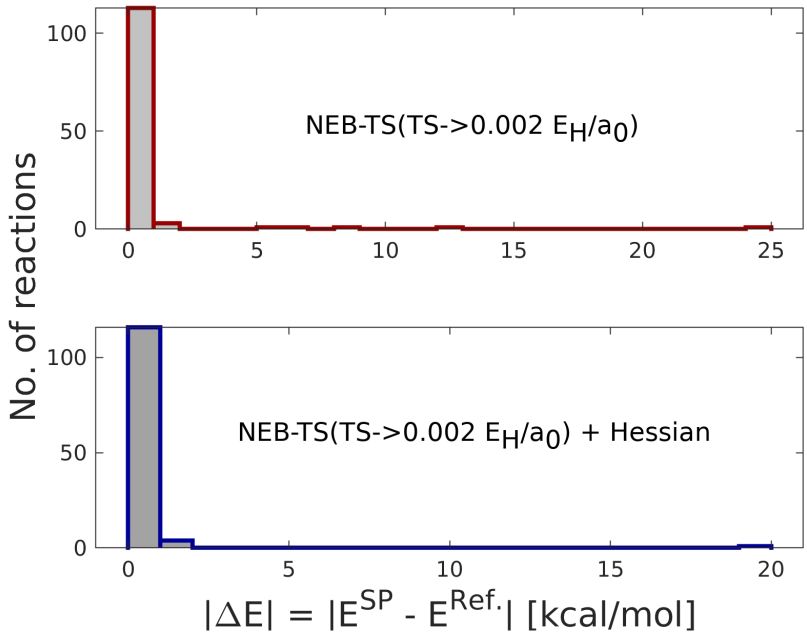


Figure 33: Energy deviation of saddle points obtained by NEB-TS from the reference set of saddle points. The NEB-TS calculations use a TS activation threshold of $\epsilon_{max}^{\mathrm{TS}} = 0.002$ $\mathrm{E}_H/\mathrm{a}_0$. In the upper panel, the TS search is started using a modified empirical Hessian matrix. In the lower panel, the TS search is started using the exact Hessian matrix. In both sets of calculations, the initial configuration of the TS search is taken as the climbing image from a partially converged EW-CI-NEB calculation.

SI-3.2 Further analysis of selected reactions

SI-3.2.1 NEB-TS converges to incorrect saddle point

The NEB-TS method can converge to an incorrect saddle point, if not executed carefully, i.e. a saddle point that is not connected to the given reactant or product state. In the following, an example of one such reaction is taken. The reaction of 2-butanol and H_2O to form acetoin. The results of EW-CI-NEB and NEB-TS calculations for this reaction are summarized in Table 5 and Fig. 34.

Table 5: Investigation of the character of the saddle points obtained from NEB-TS calculations on the formation of acetoin from 2-butanol and H_2O. The table shows results from both EW-CI-NEB and NEB-TS calculations, i.e. an estimate of the activation energy $E^{\ddagger}$ and the absolute value of the imaginary frequency $\omega_0^{\ddagger}$ at the first order saddle point obtained by the calculations. All calculations use $N_{\mathrm{im}} = 10$. The two sets of NEB-TS calculations are carried out using $\epsilon_{\mathrm{max}}^{\mathrm{TS}} \approx 0.1$ and 0.5 eV/Å.

Method	$E^{\ddagger}$ [eV]	$\omega_0^{\ddagger}$ [cm^{-1}]
EW-CI-NEB	4.34	1464.5
NEB-TS($\approx$0.1 eV/Å)	4.34	1445.8
NEB-TS($\approx$0.5 eV/Å)	3.83	539.3

The saddle point obtained by EW-CI-NEB and NEB-TS($\epsilon_{\mathrm{max}}^{\mathrm{TS}} \approx 0.1$ eV/Å) are in agreement. This saddle point is denoted as (denoted by SP-I in Fig. 34. The saddle point obtained by NEB-TS($\epsilon_{\mathrm{max}}^{\mathrm{TS}} \approx 0.5$ eV/Å), labeled as SP-II, is approximately 0.51 eV lower in energy than SP-I. Internal reaction coordinate analysis reveals that SP-II is indeed not connected to the given reactant state. The correct saddle point, SP-I, can hence be obtained by lowering the TS activation activation threshold. Also, sometimes, it may suffice to use a more accurate initial Hessian, than the Almlöf model Hessian matrix used in these NEB-TS calculations. For this particular case, however, NEB-TS($\epsilon_{\mathrm{max}}^{\mathrm{TS}} \approx 0.5$ eV/Å) using an exact Hessian also converges to SP-II.

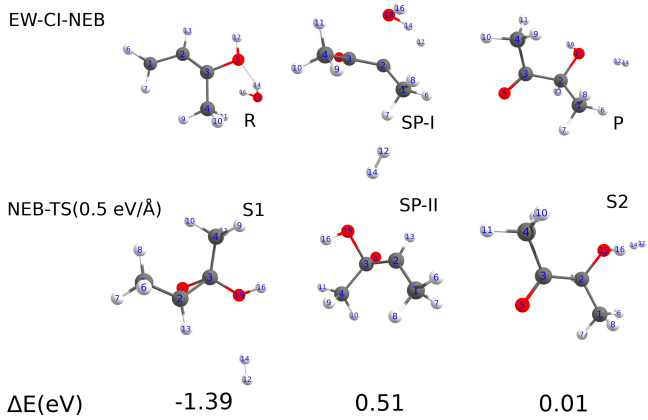


Figure 34: Internal reaction coordinate analysis of saddle points obtained by NEB-TS calculations carried out on the reaction of 2-butanol and H_2O to form acetoin. The reactant, product and correct saddle point (SP-I) configurations are shown in the upper panel. The configuration of the incorrect saddle point (SP-II), obtained by a NEB-TS(≈ 0.5 eV/Å) calculation, is shown in the center of the lower panel. From an internal reaction coordinate analysis started on SP-II, two energy minima are obtained, labeled as S1 and S2. The configuration of these energy minima is shown in the lower panel. ΔE is the energy difference between the configurations of the upper and lower panel.

SI-3.2.2 An alternative saddle point obtained by NEB-TS

The NEB-TS method can converge to saddle points that belong to a different reaction path connecting the same given reactant and product states. In the following, an example of one such reaction is taken. The cyclization of propene to form cyclopropane. The results of EW-CI-NEB and NEB-TS calculations for this reaction are summarized in Table. 6 and Figs. 35, 36 37. As is evident, the cyclization reaction is a very challenging reaction path that is characterized by an extremely flat and high energy barrier, see Fig. 35.

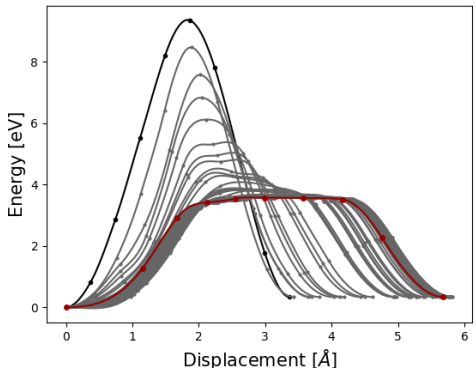


Figure 35: Optimization profile for a EW-CI-NEB-10 calculation the cyclization reaction of propene.

The EW-CI-NEB method converges to a saddle point (SP-I) that gives an activation energy of $E^{\ddagger} = 4.34$ in agreement to the saddle point obtained by NEB-TS($\epsilon^{\mathrm{TS}}_{\mathrm{max}} \approx 0.5$). Note the geometries are also nearly the same. However, the magnitude of the vibrational frequency calculated from the two saddle points is 74.0 and 330.9 cm^{-1}. While, the saddle point (SP-II) obtained by NEB-TS($\epsilon^{\mathrm{TS}}_{\mathrm{max}} \approx 0.1$), is roughly 0.25 eV higher in energy than that of EW-CI-NEB.

Table 6: Investigation of the character of the saddle points obtained from NEB-TS calculations on the cyclization reaction of propene. The table shows results from both EW-CI-NEB and NEB-TS calculations, i.e. an estimate of the activation energy $E^{\ddagger}$ and the absolute value of the imaginary frequency $\omega_0^{\ddagger}$ at the first order saddle point obtained by the calculations. All calculations use $N_{\mathrm{im}} = 10$. The two sets of NEB-TS calculations are carried out using $\epsilon^{\mathrm{TS}}_{\mathrm{max}} \approx 0.1$ and 0.5 eV/Å.

Method	$E^{\ddagger}$ [eV]	$\lvert\omega_0^{\ddagger}\rvert$ [cm^{-1}]
EW-CI-NEB	3.57	74.0
NEB-TS($\approx$0.1 eV/Å)	3.82	574.1
NEB-TS($\approx$0.5 eV/Å)	3.59	330.9

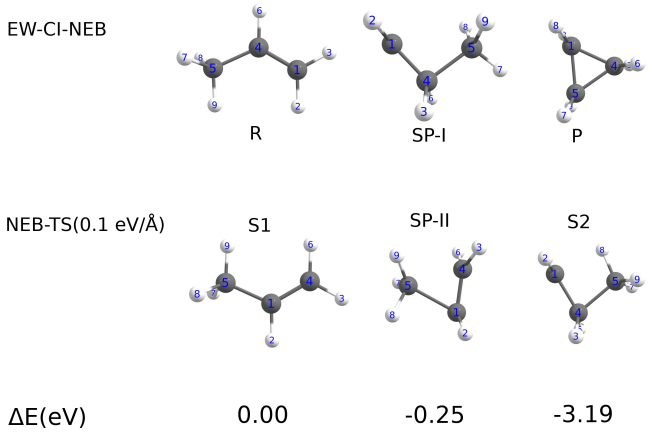


Figure 36: Internal reaction coordinate analysis of saddle points obtained by NEB-TS calculations carried out on the reaction of cyclization reaction of propene. The reactant, product and correct saddle point (SP-I) configurations are shown in the upper panel. The configuration of the other saddle point (SP-II), obtained by a NEB-TS(≈ 0.1 eV/Å) calculation, is shown in the center of the lower panel. From an internal reaction coordinate analysis started on SP-II, two energy minima are obtained, labeled as S1 and S2. The S1 corresponds to the given reactant energy minimum, while S2 corresponds to some very shallow intermediate configuration on the flat surface along the top of the energy barrier. The configuration of these energy minima is shown in the lower panel. ΔE is the energy difference between the configurations of the upper and lower panel.

IRC analysis from SP-II reveals that this saddle point is connected to the reactant state. However, the second state obtained by the IRC calculations is a very shallow intermediate energy minimum located on the top of the flat energy barrier. Therefore, to further investigate the connectivity of SP-II, two EW-CI-NEB calculations are carried out. Namely, from the reactant state to SP-II and from SP-II to the product state. The results are summarized in Fig. 37.

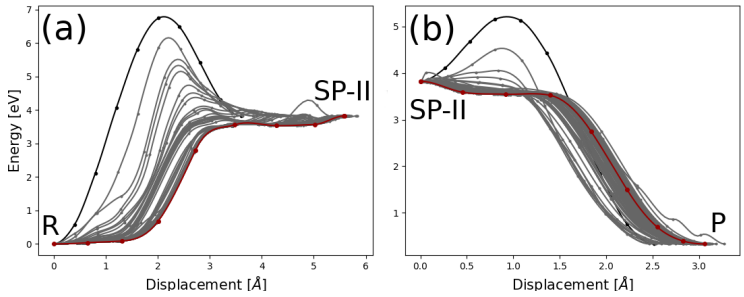


Figure 37: Optimization profile for two EW-CI-NEB-10 calculations on the cyclization reaction of propene. The calculations are started from an initial path constructed using the reactant state and SP-II (left) and from SP-II to the product state (right). SP-II is the saddle point obtained by the NEB-TS($\epsilon_{\max}^{\mathrm{TS}} \approx 0.1$ eV/Å) calculations. The EW-CI-NEB calculations indicate that SP-II is indeed connected to the same reactant and product states as SP-I, i.e. the SP obtained by the NEB-TS($\epsilon_{\max}^{\mathrm{TS}} \approx 0.1$ eV/Å) calculation.

The two partitioned EW-CI-NEB calculations carried out indicate that SP-II is likely to be connected to the given reactant and product energy minima, i.e. if the very shallow intermediate energy minimum along the path is excluded. Therefore, to conclude, it is highly likely that the NEB-TS($\epsilon_{\max}^{\mathrm{TS}} \approx 0.1$ eV/Å) calculation identifies a saddle point that characterizes an alternative reaction coordinate than the one obtained by EW-CI-NEB and NEB-TS($\epsilon_{\max}^{\mathrm{TS}} \approx 0.5$ eV/Å) calculations. Interestingly, if the exact Hessian is used as the initial Hessian matrix in NEB-TS($\epsilon_{\max}^{\mathrm{TS}} \approx 0.1$ eV/Å) the calculation converges to SP-I.

SI-3.2.3 NEB-TS identifies different saddle point on same path

The NEB-TS method can converge to a different saddle point (than the highest energy one) along a multiple extrema reaction path. In the following, an example of one such reaction is taken. The rearrangement of allyl-phenyl-ether to phenylpropylene oxide. The results of EW-CI-NEB and NEB-TS calculations for this reaction are summarized in Table 7 and Figs.38 and 39. This reaction is (at least) a two-step reaction pathway. Therefore, by locating the intermediate energy minimum and carrying out two EW-CI-NEB calculations, i.e., from the reactant to intermediate state and from the intermediate to the product state, the two saddle points along the reaction coordinate are obtained accurately, see Fig. 38.

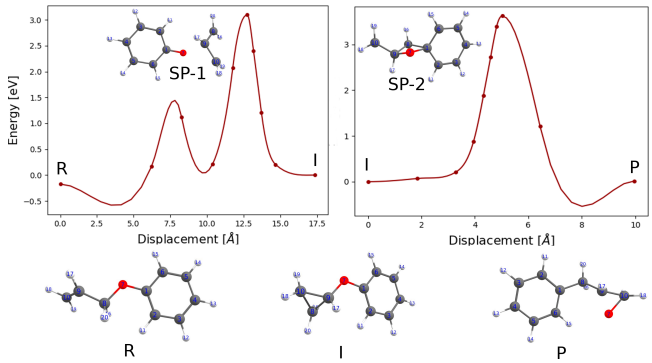


Figure 38: Energy profiles for the rearrangement of allyl-phenyl-ether to phenylpropyelene. The reaction coordinate includes an intermediate energy minimum (labeled as I) and hence two EW-CI-NEB calculations are carried out, from the reactant (R) to the intermediate state and from the intermediate to the product state (P). The two saddle points along the coordinate are labeled as SP-1 and SP-2. The reactant, product, and intermediate state configurations along with the two saddle point configurations are shown as insets. The EW-CI-NEB calculations are considered converged when the atom forces acting on CI drop below the prescribed convergence thresholds. Therefore, the initial energy maximum, at around $x = 7.5$ Å, is not necessarily a true energy maximum along the reaction path (and hence a first order saddle point on the energy surface). It could be only an artefact of the interpolation, as the atom forces acting tangential to the path are used by the interpolant.

The saddle point included in the reference set of saddle points is SP-2 and corresponds to the highest energy, first order, saddle point along the reaction path. While, the saddle points obtained by NEB-TS($\epsilon^{\mathrm{TS}}_{\mathrm{max}} \approx 0.5$) eV/Å and NEB-TS($\epsilon^{\mathrm{TS}}_{\mathrm{max}} \approx 0.1$) eV/Å correspond to SP-1, see Fig. 39. If the exact Hessian is used as the initial Hessian matrix, the two NEB-TS calculations converge to the higher energy saddle point, i.e. SP-2.

Table 7: Investigation of the character of the saddle points obtained from NEB-TS calculations on the rearrangement of allyl-phenyl-ether to phenylpropylene oxide. The table shows results from two intermediate EW-CI-NEB calculations and NEB-TS calculations, i.e. an estimate of the activation energy $E^{\ddagger}$ and the absolute value of the imaginary frequency $\omega_0^{\ddagger}$ at the first order saddle point obtained by the calculations. All calculations use $N_{\mathrm{im}} = 10$. The two sets of NEB-TS calculations are carried out using $\epsilon^{\mathrm{TS}}_{\mathrm{max}} \approx 0.1$ and 0.5 eV/Å.

Method	E‡ [eV]	$\lvert\omega_0^{\ddagger}\rvert$ [cm^{-1}]
EW-CI-NEB (SP-1)	3.79	442.0
EW-CI-NEB (SP-2)	3.27	561.2
NEB-TS(≈0.1 eV/Å)	3.26	568.2
NEB-TS(≈0.5 eV/Å)	3.28	557.6 (note that $\omega_1 = -34.0$)

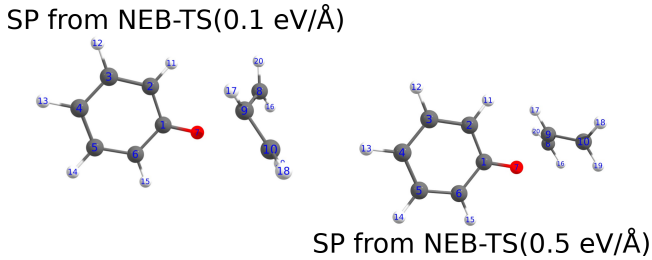


Figure 39: Saddle point configurations obtained from NEB-TS using $\epsilon^{\mathrm{TS}}_{\mathrm{max}} \approx 0.1$ and 0.5 eV/Å are shown and are in agreement to SP-1 shown in Fig. 38.

SI-3.3 Vibrational analysis of allyl-vinyl-ether reaction

Table 8: Imaginary frequencies $(\omega_0, \omega_1, \omega_2)$ for the NEB-TS calculations of the rearrangement of allyl-vinyl-ether to 1-pentene-5-one, where the TS activation threshold ($\epsilon^{\mathrm{TS}}_{\mathrm{max}}$) is varied. Note that $\epsilon^{\mathrm{TS}}_{\mathrm{RMS}} = \frac{1}{2}\epsilon^{\mathrm{TS}}_{\mathrm{max}}$ in all calculations.

$\epsilon^{\mathrm{TS}}_{\mathrm{max}}$ [$\mathrm{E_H/a_0}$]	ω_0 [$\mathrm{cm^{-1}}$]	ω_1 [$\mathrm{cm^{-1}}$]	ω_2 [$\mathrm{cm^{-1}}$]
0.02	-488.5	-158.4	-23.0
0.018	-478.9	-151.7	-25.5
0.016	-426.9	-139.0	-38.2
0.014	-426.9	-139.0	-38.2
0.012	-291.4	-165.0	–
0.010	-295.6	-158.6	–
0.008	-295.6	-158.6	–
0.006	-462.6	–	–
0.004	-473.9	–	–
0.002	-466.4	–	–

SI-4 CI-NEB and NEB-TS method parameters

To investigate selected CI-NEB and NEB-TS method parameters, a smaller benchmark set of 5 relatively simple reactions is used. The set includes HCN isomerization (**system 1**), rearrangement reaction of 1,5 hexadiene (**system 2**), Diels-alder cycloaddition (**system 3**), Ene-reaction of 1-propylene and ethylene (**system 4**) and addition of H_2 to formaldehyde (**system 5**).

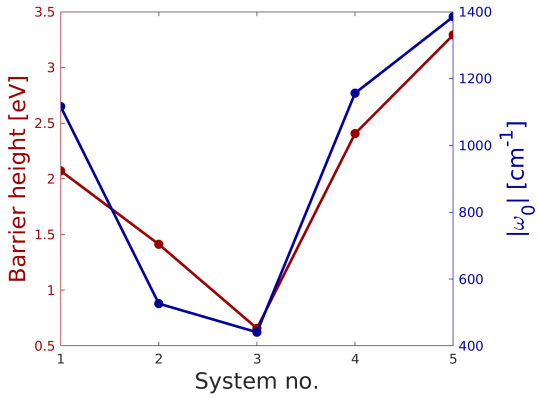


Figure 40: Barrier heights for the five reactions of the small benchmark set are shown by red (left vertical axis). Absolute imaginary frequencies (obtained from the analytical Hessian matrix computed at the saddle points) are shown by blue (right vertical axis).

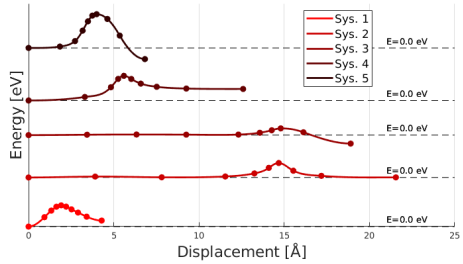


Figure 41: Minimum energy paths for the five reactions of the small benchmark set.

SI-4.1 Effect of the spring constant value

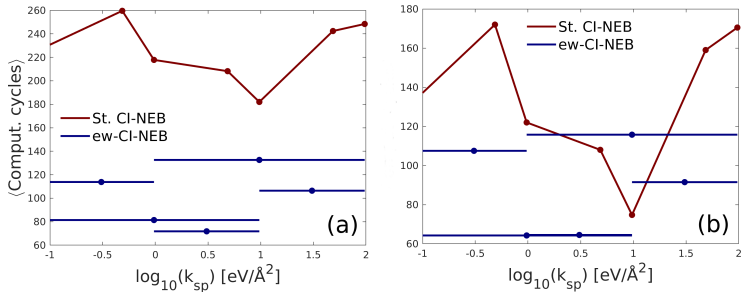


Figure 42: The average number of computational cycles required to complete the small benchmark set using both CI-NEB and EW-CI-NEB as a function of the logarithm of the spring constant, k^{SP}. Note that CI-NEB is unable to converge for system 2. In (a) system 2 is included in the average number of computational cycles, while in (b) it is excluded.

SI-4.2 Effect of the number of images

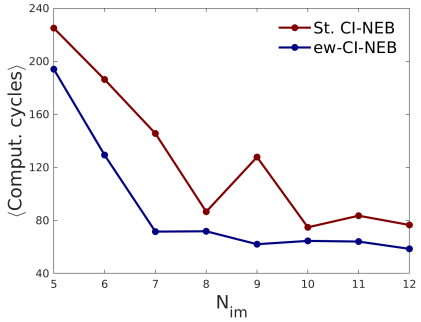


Figure 43: Computational efficiency of CI-NEB (red) and EW-CI-NEB (blue) on the small benchmark set, using variable number of images. Since, CI-NEB calculation of system 2 is non-convergent. The data from CI-NEB and EW-CI-NEB calculations on system 2 are omitted.

SI-4.3 Comparison: L-BFGS and VPO

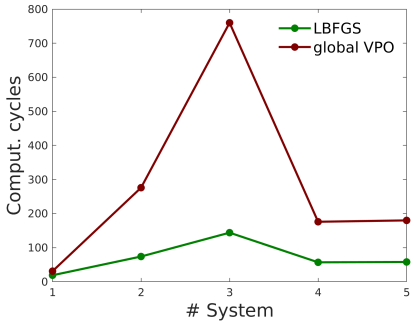


Figure 44: Computational efficiency of EW-CI-NEB calculations measured on the small benchmark set, using both L-BFGS and velocity projection optimization (VPO) method. Note that the total number of optimization steps allowed was increased from 500 to 1000 for VPO.

SI-4.4 Minimization of root-mean-square deviation (RMSD)

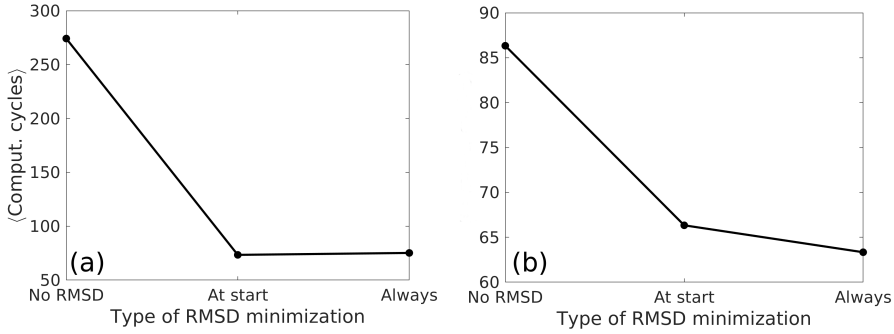


Figure 45: Computational efficiency of EW-CI-NEB calculations measured on a slightly modified benchmark set, where the product configurations have been uniformly displaced by 0.5 Å in x-direction and globally rotated by 30° ($\phi, \theta, \psi = \pi/6$). Three different sets of EW-CI-NEB calculations are then carried out; (i) without any RMSD minimization of the reactant and product configurations (ii) RMSD minimization a priori to the initial path generation of NEB and (iii) RMSD minimization is carried out, both a priori to the initial path generation and in each optimization step of EW-CI-NEB. The EW-CI-NEB calculations for systems 3 and 5 are unable to converge when no RMSD minimization is employed. In (a) the data for systems 3 and 5 is included, while in (b) it is omitted.

SI-4.5 Effect of activating climbing image

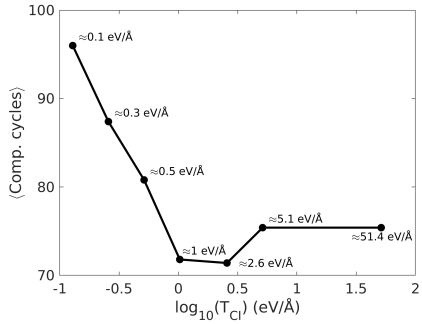


Figure 46: Computational efficiency of EW-CI-NEB/L-BFGS calculations measured on the small benchmark set, where the threshold to activate CI is varied from $\approx 0.1 - 51$ eV/Å

SI-4.6 Effect of the TS activation threshold

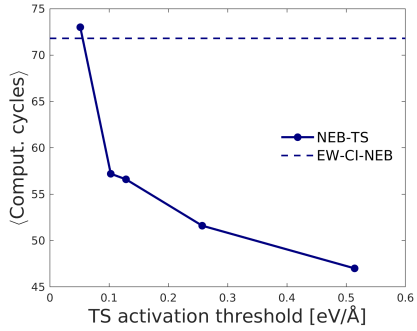


Figure 47: Computational efficiency of NEB-TS calculations using a variable TS activation threshold from $\epsilon_{\max}^{\mathrm{TS}} \approx 0.05 - 0.5$ eV/Å on the small benchmark set. For these relatively simple systems it is beneficial to activate TS in the very early stages of the EW-CI-NEB optimization.

SI-4.7 Constant trust-radius and resetting L-BFGS memory

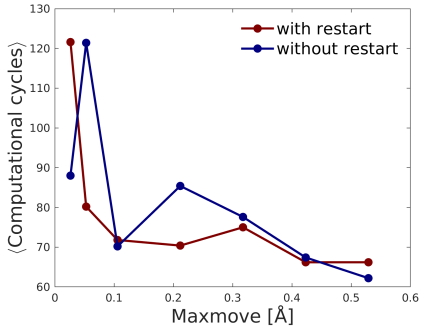


Figure 48: Computational efficiency of EW-CI-NEB/L-BFGS calculations on the small benchmark set where the allowed maximum step-size is varied, both with and without resetting the L-BFGS memory when the maximum step-size condition is invoked.

Table 9: Computational efficiency and success rate of EW-CI-NEB/L-BFGS calculations on the large benchmark set of molecular reactions. The allowed maximum step-size used is 0.2 and 0.4 a_0, with and without resetting the L-BFGS memory when the maximum step-size condition is invoked.

$\Delta_{\text{step}}[a_0]$	Reset memory	$\langle$Eval.$\rangle$	Success rate [%]
0.2	True	924±603	100
0.2	False	1001±843	96
0.4	True	897±675	98

References

(1) Birkholz, A. B. and Schlegel, H. B. Using bonding to guide transition state optimization. *J. Comput. Chem.*, **2015**, 36(15), pp. 1157–1166.

(2) Zimmerman, P. Reliable Transition State Searches Integrated with the Growing String Method. *J. Chem. theory and Comput.*, **2013**, 9, pp. 3043–3050.

(3) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. Climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys*, **2000**, 113, pp. 9901–9904

(4) Ásgeirsson, V. (2020) https://github.com/via9a/neb_visualize.git (last visited: 10.05.2021)

Article IV

Localized and Delocalized States of a Diamine Cation: Resolution of a Controversy

Galyńska, M., Ásgeirsson, V., Jónsson, H. and Bjornsson, R.
J. Pys. Chem. Lett., **12**, 1250–1255, 2021

THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCL

Letter

Localized and Delocalized States of a Diamine Cation: Resolution of a Controversy

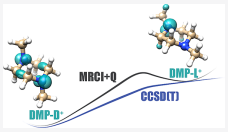
Marta Gałyńska, Vilhjálmur Ásgeirsson, Hannes Jónsson,* and Ragnar Bjornsson*

Cite This: *J. Phys. Chem. Lett.* 2021, 12, 1250–1255

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Recent Rydberg spectroscopy measurements of a diamine molecule, *N*,*N*′-dimethylpiperazine (DMP), indicate the existence of a localized electronic state as well as a delocalized electronic state. This implies that the cation, DMP^+, can similarly have its positive charge either localized on one of the N atoms or delocalized over both. This interpretation of the experiments has, however, been questioned based on coupled cluster calculations. In this article, results of high-level multireference configuration interaction calculations are presented where a localized state of DMP^+ is indeed found to be present with an energy barrier separating it from the delocalized state. The energy difference between the two states is in excellent agreement with the experimental estimate. The results presented here, therefore, support the original interpretation of the experiments and illustrate a rare shortcoming of CCSD(T), the "gold standard" of quantum chemistry. These results have implications for the development of density functionals, as most functionals fail to produce the localized state.



A mixed-valence molecule, defined as a molecule containing two or more redox sites in different oxidation states, can exhibit a localized or a delocalized electronic structure. The occurrence and energy difference between the two types of states depend on the detailed molecular structure, such as the distance between redox sites or the chemical bonds connecting them, with either through-space and through-bond mechanisms playing a role.

The radical cation of *N*,*N*′-dimethylpiperazine (DMP), see Figure 1, has been identified as an interesting organic mixed-valence molecule for studying lone-pair interactions and delocalization/localization phenomena.[1−7] Experimental studies (EPR/Raman) on DMP^+ were originally performed in solution,[1−3] where resonance Raman spectra together with calculations indicated the electronic structure of DMP^+ to be most consistent with a C_{2h}-symmetric delocalized state (here called DMP-D$^+$).[1,2] Until recently, a localized electronic structure of DMP^+ had not been observed experimentally. Ultrafast time-resolved Rydberg spectroscopy was carried out in the gas phase, where an excitation from the ground state of the DMP molecule to the 3p Rydberg state was used to monitor the picosecond time-scale dynamics from a localized to a delocalized state. By varying the energy of the photon, the energy difference between the two states could be determined as 0.33 eV, in favor of the delocalized state.[4,5] This is a rare case where the energy difference between localized and delocalized electronic states in a molecule has been determined experimentally, and it provides an important test case for theoretical methods where the balance in the electronic structure description of the two types of states can be problematic.

Since the Rydberg state electron is distributed over a large region, the Rydberg excited molecule can be assumed to resemble closely the cation. Density functional theory (DFT) calculations with commonly used density functionals, however, fail to give a localized state of DMP^+ (the BHLYP functional[8] being the exception), while calculations with a functional where self-interaction error is explicitly removed give results consistent with the Rydberg state experiments.[5] These DFT results have subsequently led to a debate in the literature, and the existence of a localized state of DMP^+ on the potential energy surface (PES) been questioned based on the fact that coupled cluster theory calculations at the CCSD(T) level do not produce a localized state.[6,7] The interpretation of the experimental measurements has thus also been questioned based on the assumption that CCSD(T) calculations, the "gold standard" of quantum chemistry, produce a reliable description of the PES. An important question, therefore, arises as to whether DFT and CCSD(T) calculations are sufficiently accurate to describe DMP^+ or whether the experimental observations need to be reinterpreted.

Organic mixed-valence cations have actually been found to present considerable challenges to both state-of-the-art wave function theory and DFT approaches.[9,10] As discussed by Kaupp and co-workers, wave function theory approaches based

Received: December 11, 2020
Accepted: January 20, 2021

https://dx.doi.org/10.1021/acs.jpclett.0c03651
J. Phys. Chem. Lett. 2021, 12, 1250−1255

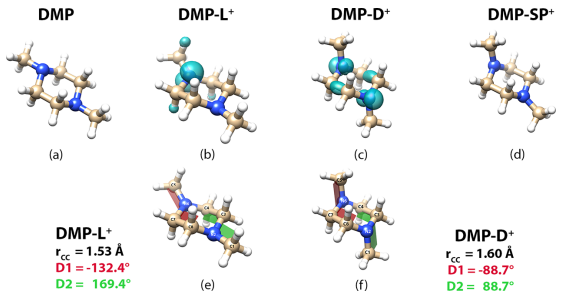


Figure 1. Structure of (a) neutral *N,N′*-dimethylpiperazine, DMP (the primary *eq-eq* conformer), (b) and (e) the localized cation, DMP-L$^+$, (c) and (f) the delocalized cation, DMP-D$^+$, and (d) the saddle point for the transition between DMP-L$^+$ and DMP-D$^+$, denoted DMP-SP$^+$. The spin densities shown in (b) and (c) correspond to an isosurface level of 0.01 electron/Å^3. The definition of the dihedral angles D1 and D2 that are used to span a cut through the energy surface are shown in (e) and (f), and their values in DMP-L$^+$ and DMP-D$^+$ are given, as well as the C−C bond length in the two structures (BHLYP level of theory).

on unrestricted Hartree−Fock (UHF) wave functions have an initial bias toward symmetry breaking and hence charge localization, in addition to spin contamination. Electron correlation is imperative for describing the possible delocalization present in such systems, but unrestricted MP2 calculations have been found to suffer from exaggerated spin contamination. Therefore, a robust dynamic correlation treatment appears necessary, and as these systems by nature have near-degeneracies, it may also be important to describe the static correlation reliably from the start (i.e., via multireference approaches) in addition to the dynamic correlation (though electron correlation is not always clearly separable).

In this study, multireference wave function calculations of the PES of the DMP cation are carried out with the aim of resolving the controversy and answering definitively the question of whether both localized and delocalized states exist on the ground-state DMP$^+$ energy surface. Using a 78-point cut of the energy surface calculated with multireference configuration interaction (FIC-MRCI+Q), we firmly establish the PES of the DMP cation as containing both a localized state and a delocalized state.

The lowest energy structure of the neutral DMP molecule is a chair conformer of C_{2h} symmetry with both methyl groups in equatorial positions (Figure 1a), which clearly reveals the presence of lone pairs on each sp^3-hybridized nitrogen (atoms N2 and N6 in Figure 1e,f). After removing one electron from a nitrogen lone-pair orbital, a positive hole remains, which results in Jahn−Teller-type distortion and the formation of a distorted structure of C_s symmetry. The geometry of this localized structure, DMP-L$^+$, is shown in Figure 1b (a symmetrically equivalent conformer also exists where the hole is localized on the other nitrogen atom). The structure of DMP-L$^+$ clearly reflects the different electronic nature of the nitrogen atoms, where the ionized nitrogen site exhibits more sp^2-like character and the nonionized site remains sp^3-like. This is also seen in the spin density in Figure 1b, showing the unpaired electron localized on only one nitrogen atom. Alternatively, the charge can be delocalized between both nitrogen atoms, resulting in a DMP-D$^+$ structure of C_{2h} symmetry. The delocalized DMP$^+$ structure shown in Figure 1c reveals the spin density as delocalized over both nitrogen atoms, and interestingly a contribution from the bridging C−C atoms can be seen, suggesting the involvement of the C−C bonds in lone-pair interactions that result in the delocalized state. The transition from the localized to the delocalized state of DMP$^+$ involves geometrical changes such as the bending and rotation of the methyl groups, as well as elongation of C−C bonds. Taking this into account, we find that two dihedral angles, D1 and D2, serve the purpose of being suitable descriptors for characterizing a cut of the PES that connects both delocalized and localized minima. The D1 (D2) angle is defined via the two planes created by C5−N6−C7 (C1−N2−C3) and N6−C7−C8 (N2−C3−C4) atoms (Figure 1e,f).

Previous theoretical studies of DMP$^+$ have involved geometry optimizations to find the lowest energy atomic configuration and minimum energy paths at the lower levels of theory as well as single-point calculations at higher levels of theory. However, the absence of the DMP-D$^+$ and DMP-L$^+$ minima for some electronic structure methods complicates comparisons, introduces an unfortunate dependence on the minimization algorithm employed as well as the initial structure, and ideally requires analytical gradients. In order to conveniently compare different electronic structure methods (with or without available analytical gradients) and to quantify the differences between methods, we instead utilize a 78-point PES cut where the dihedral angles D1 and D2 vary from 70 to 175°. The surfaces are interpolated using a biharmonic spline interpolation provided in Matlab.[11] For each surface point, constrained geometry optimizations at the BHLYP/aug-cc-pVDZ level are performed where the D1 and D2 angles are fixed while the energy is minimized with respect to all other atom coordinates. Single-point energy evaluations for all methods are then carried out on the constraint-optimized BHLYP structures (an alternative choice of configurations is discussed in the Supporting Information (SI)).

Figure 2a shows the energy surfaces calculated with the BLYP and B3LYP density functionals. Only a single minimum is found, corresponding to the delocalized state at ∼±90°. These results represent the behavior of most common density functionals. Figure 2b (bottom) shows the energy surface obtained with the BHLYP functional, where instead two well-resolved minima are obtained, one corresponding to a localized state, DMP-L$^+$, at D1 = −132.4° and D2 = 169.4°, and the other corresponding to a delocalized state, DMP-D$^+$, at

https://dx.doi.org/10.1021/acs.jpclett.0c03651
J. Phys. Chem. Lett. 2021, 12, 1250−1255

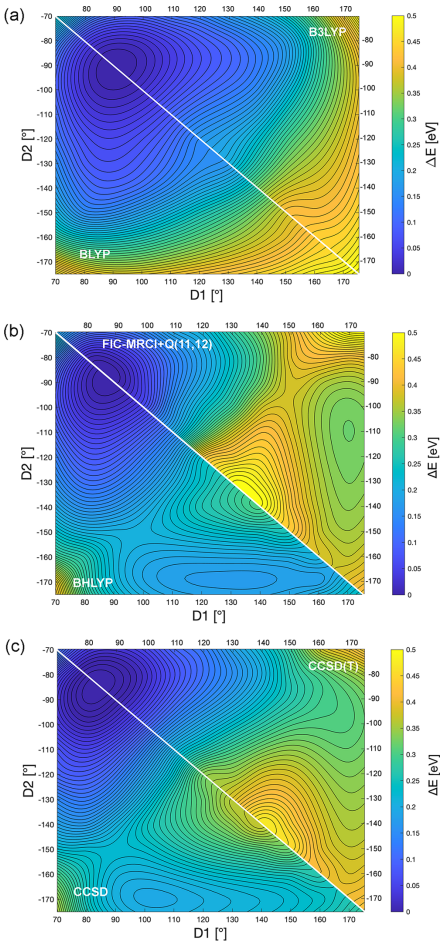


Figure 2. Potential energy surfaces calculated at the (a) B3LYP (upper) and BLYP (lower), (b) FIC-MRCI+Q(11,12) (upper) and BHLYP (lower), and (c) CCSD(T) (upper) and CCSD (lower) levels of theory. The aug-cc-pVDZ basis set is used, except in the MRCI calculation where cc-pVDZ is used. The BHLYP functional with 50% exact exchange produces a minimum on the energy surface (unlike BLYP and B3LYP) corresponding to a localized state, 0.18 eV higher in energy than the delocalized state. The structures are relaxed at the BHLYP level, subject to the constraints on the two dihedral angles defining the energy surface, and the other calculations are carried out for those structures. The high-level FIC-MRCI+Q calculations give a localized state that is 0.34 eV higher in energy than the delocalized state, in close agreement with the experimental estimate of 0.33 eV. The CCSD(T) calculation fails to give a localized state even though it is present in the CCSD calculation.

±88.7°. These values are obtained from full optimization. A first-order saddle point representing a transition state between the two states (DMP-SP$^+$ in Figure 1d) is obtained at D1 = −97.3° and D2 = 151.8°. The energy difference between DMP-D$^+$ and DMP-L$^+$ is found to be +0.178 eV using fully relaxed structures, with DMP-D$^+$ being more stable. An energy barrier of +0.033 eV is found for the transition from DMP-L$^+$ to DMP-D$^+$. Figure 2b (top) shows the energy surface calculated using the highest level of theory considered in this work, the multireference wave function method, MRCI+Q. A cc-pVDZ basis set was used, and the structures were those obtained from the BHLYP surface. The reference in the MRCI+Q calculation is a CASSCF wave function with a large active space of 11 electrons in 12 orbitals (CAS(11,12)). The fully internally contracted version of MRCI (FIC-MRCI, implemented in ORCA[12]) was used, and the Davidson size-consistency correction[13] for unlinked quadruples (Q) was applied. The internal contraction avoids bottlenecks associated with the traditional uncontracted MRCI approaches by applying an excitation operator to the whole reference wave function. The CAS(11,12) active space, including σ H_3C−N and C−C orbitals, as well as the corresponding virtual orbitals and the natural orbitals are shown in Figure 3. This active space was deemed large enough to capture the essential orbital interactions in the system.

The PESs of BHLYP and FIC-MRCI+Q are overall remarkably similar, exhibiting resolved minima for both DMP-D$^+$ and DMP-L$^+$ for similar coordinate values as well as a clearly defined energy barrier region. The magnitude of the energy difference between the delocalized and localized regions is the main difference, with BHLYP overstabilizing the localized state.

Energy surfaces were also calculated at the coupled cluster level of theory using either a singles−doubles expansion (CCSD) or the singles−doubles and perturbative triples expansion (CCSD(T)) as shown in Figure 2c. Importantly, for each surface point, a stability analysis (calculation of the electronic Hessian) was performed for the UHF SCF solution, and in many cases instabilities were found; new stable solutions were subsequently generated. In fact, it was found that all stable UHF SCF solutions had a localized electronic structure, and no delocalized minimum could be found at the HF level. The CCSD surface (shown in Figure 2c, lower) is overall comparable to BHLYP surface in terms of relative energy, while the position of the DMP-L$^+$ minimum is closer to the MRCI+Q minimum. The two minima are directly visible on the surface and are consistent with the results of previous CCSD geometry optimizations.[5,6] The energy surface shows, furthermore, the presence of a low-energy energy barrier. Spin population analysis of the unrelaxed CCSD density for the D1 = −90°, D2 = 90° point (Mulliken nitrogen spin populations of 0.38 and 0.30, respectively) confirms the electronic structure as mostly delocalized with CCSD, despite being expanded from a localized UHF reference wave function. Despite the minor symmetry breaking present, the CCSD wave function appears to describe the electronic structure and energy surface of DMP$^+$ qualitatively correctly. The energy difference between DMP-D$^+$ and DMP-L$^+$ is calculated to be +0.22 eV, in agreement with previous results (0.23 eV).[5] However, remarkably, when perturbative triples excitations are added (i.e., the CCSD(T) method), the PES changes substantially. The CCSD(T) surface (Figure 2c, upper) has only one

https://dx.doi.org/10.1021/acs.jpclett.0c03651
J. Phys. Chem. Lett. 2021, 12, 1250−1255

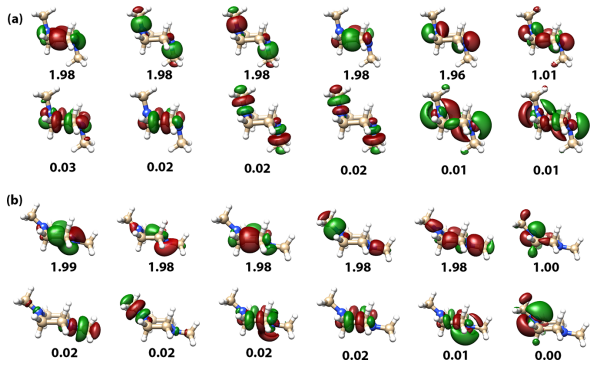


Figure 3. CAS(11,12) active space for (a) DMP-D$^+$ and (b) DMP-L$^+$. Natural orbitals calculated at the CASSCF/aug-cc-pVDZ level of theory on the BHLYP/aug-cc-pVDZ structures are shown with their natural occupation numbers. The active space contains two N lone pairs, two σ H_3C-N and two σ C−C bonding orbitals, and the corresponding virtual orbitals. As seen, the orbitals associated with the N lone pairs have acquired considerable C−C bond character in both DMP-D$^+$ and DMP-L$^+$ states, and the C−C σ^* orbitals have significant fractional occupation (0.02−0.03), suggesting the importance of the C−C bonds in a through-bond delocalization mechanism. It was found necessary to include virtual N 3p atom-like orbitals for a balanced active space.

minimum, analogous to the results of most DFT methods. The localized state is missing.

The results presented here (see Table 1) obtained at the MRCI+Q level of theory with a large CASSCF(11,12) reference wave function clearly establish the presence of both a localized and a delocalized minimum on the PES of the DMP cation. Furthermore, the energy difference calculated using the aug-cc-pVDZ basis set is 0.336 eV, in excellent agreement with the measured energy difference between delocalized and localized Rydberg energy states.[5] While the MRCI+Q calculations depend on the theory level used for structural optimization, we note that the use of alternative structures at the CCSD or DMRG-CASSCF(19,20) level (to be reported on later) result in small changes to the energy difference (less than 0.02 eV). The barrier from DMP-L$^+$ to DMP-D$^+$ (using the BHLYP-located minima and saddle point geometries) is calculated to be 0.05 eV. The multireference wave function energy surface hence reinforces the original interpretation of the experimental measurements that a localized state exists, separated by an energy barrier from the delocalized state. This still hinges on the assumption that the Rydberg states resemble the states of the molecular cation.

Table 1. Single-Point BHLYP, MP2, CCSD, and FIC-MRCI +Q(11,12) Energies (eV) of DMP-D$^+$ and DMP-SP$^+$ Relative to DMP-L$^+$, Calculated Using the Relaxed BHLYP/aug-cc-pVDZ Structures

method	ΔE	$E_{SP} - E_{loc}$
BHLYP/aug-cc-pVDZ	0.178	0.033
MP2/aug-cc-pVDZ	0.153	0.081
CCSD/aug-cc-pVDZ	0.221	0.012
FIC-MRCI+Q(11,12)/aug-cc-pVDZ	0.336	0.050
experiment[5]	0.33	

The cc-pVDZ basis set was used for the calculations of the energy surface, while a larger aug-cc-pVDZs basis set was used for the calculations of the local minima. We estimate that increasing the basis set further would change the relative energy by less than ∼0.02 eV, based on explicitly correlated CCSD-F12 calculations (CCSD-F12/cc-pVDZ-F12 results give an energy difference of 0.238 eV) as well as large-basis local-correlation CCSD calculations (see the SI).

In a recent comment, Ali et al.[6] presented CCSD and CCSD(T) calculations of DMP$^+$ and compared them with DFT results. They argued that CCSD(T) gives accurate results and that the absence of a localized state in most DFT calculations does not represent a shortcoming of those functionals. From the high-level MRCI+Q results presented here, it is clear that the CCSD(T) calculations are in fact in error, possibly in part due to its single-reference nature. The lack of a stable delocalized solution at the reference UHF level (as shown by stability analysis) and the contrasting behavior of CCSD and CCSD(T) suggest this. In order to understand whether the behavior of CCSD(T) stems from the flawed UHF reference function or alternatively the perturbative triples correction, we performed CCSD(T) calculations using alternative reference wave functions, as detailed in the SI. CCSD(T) calculations using Brueckner orbitals,[14,15] quasi-restricted orbitals,[16] and UKS-DFT orbitals did not, however, lead to an improved CCSD(T) energy surface; see SI for a discussion. We also performed orbital-optimized coupled cluster theory (OO-CCD(T)) calculations, where the orbitals are variationally optimized at the CCSD level (instead of the HF level), thus effectively removing any effect of the HF reference. While this encouragingly gave a relaxed OO-CCD density that showed complete delocalization (without symmetry-breaking) according to spin population analysis of the D1 = −90°, D2 = 90° point (see Table S2 in the SI), the OO−CCD(T) surface is still missing the DMP-L$^+$ minimum, as shown in Figure S1 in the SI. These results suggest that the HF reference wave function is not the culprit in the CCSD(T) calculation, but rather the wave function expansion itself, perhaps due to an imbalance in the static and dynamic correlation of the wave function when the perturbative triples correction is included. It would be interesting to see whether

https://dx.doi.org/10.1021/acs.jpclett.0c03651
J. Phys. Chem. Lett. 2021, 12, 1250−1255

these problems are resolved at the (very expensive) CCSDT (full triples)[17] or CCSDT(Q)[18] levels (full triples and perturbative quadruples) or via alternative triples approximations.[19−24]

In summary, we have presented a combined single-reference and multireference wave function theory investigation of the DMP cation and have shown that the molecule represents an unusual challenge to standard quantum chemistry methods, whether at the DFT level or the wave function theory level. Surprisingly, the CCSD(T) level of theory gives a qualitatively incorrect energy surface, failing to describe the localized state of the DMP cation that is unquestionably present at the MRCI +Q level of theory and is inferred from Rydberg spectroscopy measurements. The DMP cation, despite its apparent simplicity, is a truly challenging case for correlated wave function theory, and the presented results should make the system useful as a benchmark system for the study of electronic state localization and guide the development of more robust and affordable correlated wave function methods as well as density functionals.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.0c03651.

Computational details, basis set convergence, Mulliken spin populations of wave functions, and Cartesian coordinates of optimized structures (PDF)

A compressed directory with XYZ files for the entire 78-point BHLYP-constraint-optimized surface (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

Hannes Jónsson − *Science Institute and Faculty of Physical Sciences, VR-III, University of Iceland, 107 Reykjavík, Iceland;* orcid.org/0000-0001-8285-5421; Email: hj@hi.is

Ragnar Bjornsson − *Max-Planck Institute for Chemical Energy Conversion, 45470 Mülheim an der Ruhr, Germany;* orcid.org/0000-0003-2167-8374; Email: ragnar.bjornsson@cec.mpg.de

Authors

Marta Galyńska − *Science Institute and Faculty of Physical Sciences, VR-III, University of Iceland, 107 Reykjavík, Iceland; Institute of Physics, Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University, 87-100 Toruń, Poland*

Vilhjálmur Ásgeirsson − *Science Institute and Faculty of Physical Sciences, VR-III, University of Iceland, 107 Reykjavík, Iceland*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.0c03651

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

M.G. acknowledges a postdoctoral fellowship from the University of Iceland Research Fund. This work was supported by the Icelandic Research Fund. R.B. acknowledges the Max Planck Society for funding. The computations were performed on resources provided by the Icelandic Research High Performance Computing Centre (IRHPC) at the University of Iceland.

■ REFERENCES

(1) Brouwer, A. M.; Zwier, J. M.; Svendsen, C.; Mortensen, O. S.; Langkilde, F. W.; Wilbrandt, R. Radical Cation of N,N-Dimethylpiperazine: Dramatic Structural Effects of Orbital Interactions through Bonds. *J. Am. Chem. Soc.* **1998**, *120*, 3748−3757.

(2) Brouwer, A. M.; Langkilde, F. W.; Bajdor, K.; Wilbrandt, R. Through-Bond Interaction in the Radical Cation of N,N-Dimethylpiperazine. Resonance Raman Spectroscopy and Quantum Chemical Calculations. *Chem. Phys. Lett.* **1994**, *225*, 386.

(3) Eastland, G. W.; Rao, D. N. R.; Symons, M. C. R. Amine Radical Cations: An Electron Spin Resonance Study of Cations Generated by Radiolysis. *J. Chem. Soc., Perkin Trans. 2* **1984**, *9*, 1551−1557.

(4) Deb, S.; Cheng, X.; Weber, P. M. Structural Dynamics and Charge Transfer in Electronically Excited N,N-Dimethylpiperazine. *J. Phys. Chem. Lett.* **2013**, *4*, 2780−2784.

(5) Cheng, X.; Zhang, Y.; Jónsson, E.; Jónsson, H.; Weber, P. M. Charge localization in a diamine cation provides a test of energy functionals and self-interaction correction. *Nat. Commun.* **2016**, *7*, 11013.

(6) Ali, Z.; Aquino, F. W.; Wong, B. The diamine cation is not a chemical example where density functional theory fails. *Nat. Commun.* **2018**, *9*, 4733.

(7) Cheng, X.; Jónsson, E.; Jónsson, H.; Weber, P. M. Reply to: "The diamine cation is not a chemical example where density functional theory fails. *Nat. Commun.* **2018**, *9*, 5348.

(8) Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372−1377.

(9) Parthey, M.; Kaupp, M. Quantum-Chemical Insights into Mixed-Valence Systems: Within and beyond the Robin-Day Scheme. *Chem. Soc. Rev.* **2014**, *43*, 5067−5088.

(10) Kaupp, M.; Renz, M.; Parthey, M.; Stolte, M.; Würthner, F.; Lambert, C. Computational and Spectroscopic Studies of Organic Mixed-Valence Compounds: Where Is the Charge? *Phys. Chem. Chem. Phys.* **2011**, *13*, 16973−16986.

(11) *MATLAB and Statistics Toolbox*, Release 2019b, The MathWorks Inc., Natick, MA, 2019.

(12) Sivalingam, K.; Krupicka, M.; Auer, A. A.; Neese, F. Comparison of Fully Internally and Strongly Contracted Multi-reference Configuration Interaction Procedures. *J. Chem. Phys.* **2016**, *145*, 054104.

(13) Langhoff, S. R.; Davidson, E. R. Configuration Interaction Calculations on the Nitrogen Molecule. *Int. J. Quantum Chem.* **1974**, *8*, 61−72.

(14) Brueckner, K. A. Nuclear Saturation and Two-Body Forces. II. Tensor Forces. *Phys. Rev.* **1954**, *96*, 508−516.

(15) Handy, N. C.; Pople, J. A.; Head-Gordon, M.; Raghavachari, K.; Trucks, G. W. Size-consistent Brueckner theory limited to double substitutions. *Chem. Phys. Lett.* **1989**, *164*, 185−192.

(16) Neese, F. Importance of Direct SpinSpin Coupling and Spin-Flip Excitations for the Zero-Field Splittings of Transition Metal Complexes: A Case Study. *J. Am. Chem. Soc.* **2006**, *128*, 10213−10222.

(17) Noga, J.; Bartlett, R. J. he Full CCSDT Model for Molecular Electronic Structure. *J. Chem. Phys.* **1987**, *86*, 7041−7050.

(18) Kucharski, S. A.; Bartlett, R. J. An Efficient Way to Include Connected Quadruple Contributions into the Coupled Cluster Method. *J. Chem. Phys.* **1998**, *108*, 9221−9226.

(19) Urban, M.; Noga, J.; Cole, S. J.; Bartlett, R. J. Towards a Full CCSDT Model for Electron Correlation. *J. Chem. Phys.* **1985**, *83*, 4041−4046.

(20) Noga, J.; Bartlett, R. J.; Urban, M. Towards a Full CCSDT Model for Electron Correlation. CCSDT-n Models. *Chem. Phys. Lett.* **1987**, *134*, 126−132.

(21) Taube, A. G.; Bartlett, R. J. Improving upon CCSD(T): CCSD(T). I. Potential Energy Surfaces. *J. Chem. Phys.* **2008**, *128*, 044110.
(22) Eriksen, J. J.; Matthews, D. A.; Jørgensen, P.; Gauss, J. Assessment of the Accuracy of Coupled Cluster Perturbation Theory for Open-Shell Systems. I. Triples Expansions. *J. Chem. Phys.* **2016**, *144*, 194102.
(23) Piecuch, P.; Wloch, M. Renormalized Coupled-Cluster Methods Exploiting Left Eigenstates of the Similarity-Transformed Hamiltonian. *J. Chem. Phys.* **2005**, *123*, 224105.
(24) Piecuch, P. Active-Space Coupled-Cluster Methods. *Mol. Phys.* **2010**, *108*, 2987−3015.

https://dx.doi.org/10.1021/acs.jpclett.0c03651
J. Phys. Chem. Lett. 2021, 12, 1250−1255

Supporting Information for: Localized and Delocalized States of a Diamine Cation: Resolution of a Controversy

Marta Gałyńska,[†,‡] Vilhjálmur Ásgeirsson,[†] Hannes Jónsson,[*,†] and Ragnar Bjornsson[*,¶]

†*Faculty of Physical Science, VR-III, University of Iceland, 107 Reykjavik, Iceland*

‡*Institute of Physics, Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University, Grudziadzka 5, 87-100 Toruń, Poland*

¶*Max-Planck Institute for Chemical Energy Conversion, Stiftstrasse 34-36, 45470 Mülheim an der Ruhr, Germany*

E-mail: hj@hi.is; ragnar.bjornsson@cec.mpg.de

1. Computational details

All calculations were performed using the ORCA 4.0 quantum chemistry code.[1] Calculations were performed using the correlation-consistent basis sets.,[2,3] Coupled cluster calculations were performed using multiple reference wavefunctions as described below. Stability analysis was performed for all HF calculations. Saddle points connecting the localized and delocalized minima were located at the BHLYP level of theory using an eigenvector-following method using the full analytical Hessian as implemented in ORCA.

2. Choice of optimization method for constrained optimizations

The choice to use BHLYP to generate delocalized and localized geometries for all surfaces is justified as BHLYP is one of very few functionals (as discussed in Cheng et al.[4]) that gave converged minima for both DMP-L+ and DMP-D+ states and analytical gradients are readily available. We also compared the quality of the BHLYP structures to CCSD structures for DMP-L^+ and DMP-D^+ via higher level single-point FIC-MRCI+Q/aug-cc-pVDZ single point calculations. The FIC-MRCI+Q total energies were lowest when using CCSD structures (hence being closer to the FIC-MRCI+Q minima) but the BHLYP structures were not far off. Finally, the energy difference between DMP-L^+ and DMP-D^+ is hardly affected (< 0.02 eV) by the geometric difference between CCSD and BHLYP.

3. Basis set convergence of correlated wavefunctions

The basis set convergence of the DMP-L^+ and DMP-D^+ energy difference (using BHLYP/aug-cc-pVDZ structures) was studied further at the CCSD level using both local correlation theory (domain-based local pair natural orbitals, DLPNO) and F12 explicit correlation. The DLPNO approximation enables the use of large basis sets while the explicitly correlated

method exhibits faster convergence to the basis set limit. Tight thresholds were used in the DLPNO approximation (TightPNO keyword in ORCA).

The results in Table 1 reveals that the basis set convergence of DMP-L$^+$ and DMP-D$^+$ is relatively mild with the small cc-pVDZ and aug-cc-pVDZ basis sets giving very small errors ($<$ 0.02 eV) compared to either DLPNO-CCSD/aug-cc-pVQZ or DLPNO-CCSD-F12/cc-pVQZ-F12. This justifies the use of these basis sets in our correlated wavefunction calculations for the energy surfaces, both at the coupled cluster level and the MRCI+Q level, as correlated wavefunction methods exhibit very similar basis set dependence in general.

Table 1: The basis set dependence of the energy difference between DMP-L$^+$ and DMP-D$^+$ at the DLPNO-CCSD level.

DLPNO-CCSD/cc-pVDZ	0.210
DLPNO-CCSD/cc-pVTZ	0.197
DLPNO-CCSD/cc-pVQZ	0.163
DLPNO-CCSD/aug-cc-pVDZ	0.179
DLPNO-CCSD/aug-cc-pVTZ	0.169
DLPNO-CCSD/aug-cc-pVQZ	0.170
DLPNO-CCSD-F12/cc-pVDZ-F12	0.198
DLPNO-CCSD-F12/cc-pVTZ-F12	0.191
DLPNO-CCSD-F12/cc-pVQZ-F12	0.194

4. Coupled-cluster calculations with different reference wavefunctions.

The CCSD and CCSD(T) PES surfaces were calculated using different reference wave functions: quasi-restricted orbitals (QRO),[5] PBE orbitals and Brueckner orbitals. All CC calculations started from a stable SCF solution. The data is presented in Figure S1. For CCSD, DMP-L$^+$ and DMP-D$^+$ minima were always present on the surfaces regardless of the reference wave function used, with only some changes in energies. The energy surface changed slightly compared to UHF-CCSD, especially regarding the position of the DMP-L$^+$ minimum and the height of the saddlepoint region between DMP-L$^+$ and of DMP-D$^+$. For CCSD(T),

the use of QRO and PBE references did not influence the surface much (Figure S1a and S1b) compared to UHF-CCSD(T). However, the CCSD(T) PES obtained using Brueckner orbitals (Figure S1c) was found to be strangely chaotic in comparison to other CCSD(T) results. The reasons for this are not presently clear.

Calculations were also performed using orbital-optimized coupled-cluster theory where the orbitals are simultaneously relaxed at the CCSD level, resulting in optimal orbitals for the CCSD wavefunction, which results in the absence of single excitations, hence the name OO-CCD. On top of the OO-CCD wavefunction, the perturbative (T) correction was also calculated. The results for OO-CCD give a rather similar surface as UHF-CCSD, with some changes in the depth of the energy wells. The OO-CCD(T) surface suffers from the same problem as the other CCSD(T) surfaces.

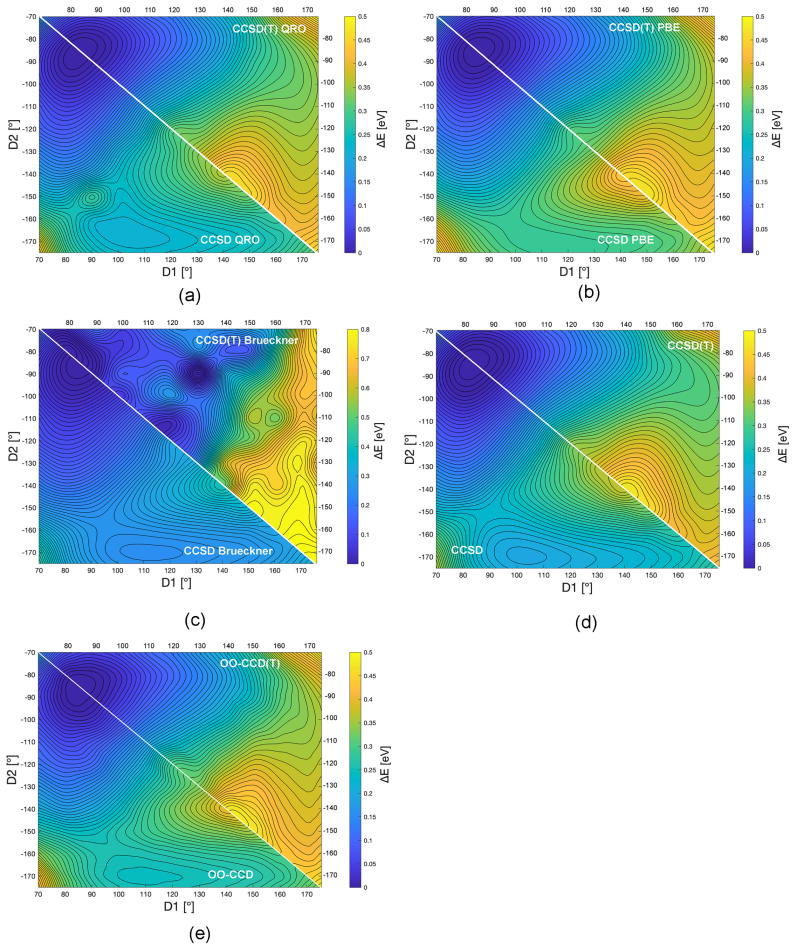


Figure 1: The potential energy surfaces calculated at the CCSD(T)/cc-pVDZ(upper) and CCSD/cc-pVDZ (lower) level of theory using different reference wave functions: (a) QRO, (b) PBE, (c) Brueckner and (d) UHF. Additionally, orbital optimized coupled cluster surfaces are shown in (e): OO-CCD(T) (upper) and OO-CCD (lower).

5. Symmetry-breaking in the electronic structure of DMP-D$^+$.

The DMP-D$^+$ state is symmetric with respect to nuclear coordinates but the wavefunction can artificially break symmetry as can be seen by inspecting the difference in nitrogen atom spin populations of a Mulliken decomposition of the wavefunction. Table 2 below compares Mulliken spin population on the nitrogen atoms from different electronic structure methods on selected points in the delocalized region.

Table 2: Mulliken spin populations on nitrogen atoms for the surface points which correspond to the DMP-D$^+$ minimum found on the potential energy surfaces calculated at various levels of theory. BHLYP constraint-optimized structures were used.

HF (D1=-80, D2=80), stable WF	0.98	0.09
HF (D1=-80, D2=80), unstable WF	0.51	0.51
B3LYP (D1=-90, D2=90)	0.40	0.40
BHLYP (D1=-90, D2=90)	0.43	0.43
CCSD (D1=-90, D2=90) linearized, stable HF WF	0.32	0.34
CCSD (D1=-90, D2=90) unrelaxed, stable HF WF	0.38	0.30
CCSD (D1=-90, D2=90) linearized, unstable HF WF	0.34	0.34
CCSD (D1=-90, D2=90) unrelaxed, unstable HF WF	0.34	0.34
OO-CCD (D1=-90, D2=90) relaxed	0.40	0.40
CASSCF(11,12) (D1=-90, D2=90)	0.37	0.37

6. Cartesian coordinates of optimized geometries

A compressed directory with XYZ files for the entire 78-point BHLYP-constraint-optimized surface is provided as additional supporting information.

Table 3: The Cartesian coordinates [Å] of the DMP-D+ minimum relaxed at the BHLYP/aug-cc-pVDZ level.

N	0.22030638483835	1.38335169439629	-0.00141428270678
C	-0.23307706173303	0.76534307809786	1.18740587238139
C	0.23894425174175	-0.76386646452091	1.18722001517577
N	-0.22030684419671	-1.38335113899910	0.00141431037829
C	0.23307659981281	-0.76534262310306	-1.18740587001465
C	-0.23894470846030	0.76386687853965	-1.18721999841372
C	1.53471993591523	1.99276339527823	-0.00517159621707
C	-1.53471981040781	-1.99276394447674	0.00517158067032
H	-1.32174960069968	0.77017282021861	1.23316844148976
H	0.17448322791798	1.25447472839007	2.06793203095279
H	-0.16433157233165	-1.25187750268697	2.07034182284149
H	1.32782459474641	-0.76866820858800	1.22766333849882
H	0.16433122635944	1.25187782151717	-2.07034182920387
H	-1.32782506498139	0.76866867233390	-1.22766348128770
H	1.32174917170525	-0.77017245222663	-1.23316826841778
H	-0.17448366508040	-1.25447438095232	-2.06793194920358
H	1.64798713622730	2.61399603089722	0.87961502815664
H	1.64232763481302	2.61519954488115	-0.88983629749963
H	2.33102223344338	1.23797485970780	-0.00826747341729
H	-1.64232681989818	-2.61520057736802	0.88983602393025
H	-1.64798657934418	-2.61399616615818	-0.87961535284821
H	-2.33102267038759	-1.23797606517800	0.00826793475475

Table 4: The Cartesian coordinates [Å] of the DMP-L+ minimum relaxed at the BHLYP/aug-cc-pVDZ level.

N	-1.378932	-0.003352	-0.226901
C	-0.705445	1.187327	0.222762
C	0.700124	1.248790	-0.380457
N	1.410727	0.013527	-0.128835
C	0.713221	-1.228795	-0.388546
C	-0.690703	-1.187437	0.218188
C	-2.794842	-0.012378	0.105594
C	2.757891	-0.005867	0.378014
H	-0.629913	1.247429	1.319155
H	-1.245791	2.068554	-0.118935
H	1.280830	2.076787	0.016026
H	0.618690	1.350018	-1.467284
H	-1.221707	-2.073941	-0.124470
H	-0.612504	-1.249663	1.314463
H	0.632766	-1.324689	-1.475509
H	1.306062	-2.051484	0.002621
H	-3.272092	0.866545	-0.323696
H	-2.973101	-0.015002	1.188882
H	-3.261381	-0.895772	-0.326273
H	3.335898	-0.751465	-0.168224
H	2.728250	-0.305096	1.431918
H	3.209944	0.976712	0.288157

Table 5: The Cartesian coordinates [Å] of the DMP+ saddlepoint located at the BHLYP/aug-cc-pVDZ level.

N	1.131597	0.828767	0.002660
C	0.422674	0.572147	1.208039
C	-0.103337	-0.901559	1.172916
N	-0.862151	-1.059246	-0.030418
C	-0.164942	-0.794520	-1.251810
C	0.363590	0.677288	-1.184182
C	2.197429	1.811730	0.019502
C	-2.298031	-0.935325	0.015275
H	-0.446649	1.226865	1.359272
H	1.085291	0.678073	2.064392
H	-0.726512	-1.109469	2.037769
H	0.754386	-1.570691	1.154351
H	0.984504	0.857744	-2.059074
H	-0.510080	1.341604	-1.235907
H	0.690230	-1.461981	-1.335208
H	-0.831287	-0.926148	-2.099190
H	1.819671	2.838707	0.074090
H	2.795376	1.708868	-0.884254
H	2.839693	1.630581	0.879473
H	-2.684936	-1.469849	0.878986
H	-2.731732	-1.332189	-0.898356
H	-2.570440	0.124572	0.106291

References

(1) Neese, F. Software Update: The ORCA Program System, Version 4.0. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1327.

(2) Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(3) Kendall, T. H. H. R. J., R. A.; Dunning Electron Affinities of the First-row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(4) Cheng, X.; Zhang, Y.; Jónsson, E.; Jónsson, H.; Weber, P. M. Charge localization in a diamine cation provides a test of energy functionals and self-interaction correction. *Nature Communications* **2016**, *7*, 11013.

(5) Neese, F. Importance of Direct SpinSpin Coupling and Spin-Flip Excitations for the Zero-Field Splittings of Transition Metal Complexes: A Case Study. *J. Am. Chem. Soc.* **2006**, *128*, 10213–10222.