22 Path Optimisation Methods

Many interesting chemical and physical processes involve transitions from one state to another. Typical examples are migration paths for defects in a solid or on a surface and chemical reactions. As a start to fully describing the kinetics of a transition the minimum-energy path is desired. As its name suggests the minimum-energy path is the lowest-energy path connecting two minima on a potential energy surface.

When the minimum energy path is known, the rate of occurrence for the related events can be easily related to the energy by the Arrhenius equation:

$$k = A \exp\left[-E_a/k_B T\right]$$

(22.1)

where $A$ is the pre-exponential coefficient, also often denoted as attempt frequency (even though it does not necessarily have the unit of frequency). It can be seen as a measure for how often a process potentially leading to a transition, e.g. a typical frequency for lattice vibrations in the case of defect migration, takes place. $E_a$ is the activation energy.

In a two-dimensional case, the most simple example is given by two energy minima with a single saddle point in the middle.

![Fig. 1. Equipotential contours in two dimensions, but intended here to suggest the general n-directional situation. A and B are minima, z=0 is a symmetry hyperplane and \( \beta \) the saddle point.](image)

In principle, such processes can be studied with the MD method. However, if the barrier is too high, the processes are so slow that they will not be observed during MD time scales at reasonable temperatures. This is similar to the case of free energy calculations in cases where there are significant differences in free energy. Therefore, we need a method which allows calculating these energy barriers without carrying out a dynamical simulation.
22.1 Transition State Theory

According to transition state theory (also known as activated-complex theory), in between reactant, and product state of a process, there is a state known as the transition state. It is defined by the maximum of the lowest-energy path between two minima on the energy surface. During the transition state, the reactants are combined to form a species called the activated complex.

This activated complex is assumed to be in chemical equilibrium with the reactant state. The overall rate for the reaction can then be calculated from the concentration of the activated complex and its rate of decomposition. This can be approximated in terms of the vibrational frequencies of initial state and activated complex to give

\[
k = \left( \prod_{j=1}^{N} \nu_j \right) \left( \prod_{j=1}^{N-1} \nu'_j \right)^{-1} e^{-\Delta^*G/k_B T} \tag{22.2}
\]

with \(\Delta^*G\) being the Gibbs free energy of forming the activated complex from the reactant state, and \(\nu_j\) and \(\nu'_j\) being the vibrational frequencies of the initial potential minimum and the activated complex, respectively. The transition state is associated with one less frequency, because there is one direction in which the system is driven away from the saddle point rather than towards it. In solid state physics calculations it is typical for defect migration to simply assume that the frequencies can be described by just
one material-dependent vibrational frequency $\nu_0$. Values for this are often in the range of $\nu_0 \approx 10^{12} \text{ s}^{-1}$.

This way, if we can follow the system over a period of time and observe the reaction rate, we can in principle calculate the reaction path energy barrier $\Delta^*G$. On the other hand, if we know the barrier, we can calculate the reaction rate without carrying out the (usually impossibly long) dynamical simulation.

### 22.2 Locating the Minimum Energy Reaction Path (MEP)

There are several methods for either directly locating the transition state, or finding the minimum energy path and thereby calculating the energy barrier:

- **Grid Search** The most primitive method for localising the transition state is the grid search method. In this method, the potential energy is evaluated over a grid of points which span the saddle point. In order to locate the saddle point accurately, however, a very fine grid is required, making this method essentially infeasible for all but the very smallest systems.

- **Reaction Coordinate** Another simplistic approach is to choose a certain internal coordinate of the system (e.g. a bond length) as the reaction coordinate. At various intervals along this coordinate, between its value for the reactant and product state, all other coordinates are then optimised. However, the reaction pathway found in this way is not necessarily physically meaningful; other choices of reaction coordinates can give different pathways. In fact, even one reaction coordinate can give different pathways, depending whether one starts from the reactant or product state.

- **Quasi Newton Methods** The Quasi Newton Methods discussed in the Energy minimisation chapter can in principle also be applied on the optimisation of a transition state. There is, however, no method to prevent the Hessian matrix from becoming positive definite, which would cause the method to find a minimum instead of the saddle point.

- **Conjugate Gradient Methods** Also the conjugate gradient method discussed in the energy minimisation chapter can be used in order to localise transition states. The methods starts with the midpoint
between reactant and product state as a starting point. From this point
the energy is maximised along the straight line connecting the two mi-
nima $p^1$. Now a conjugate gradient optimisation is performed starting
from this maximum, ensuring that the gradient in direction $p^1$ is kept
close to zero. This method has the advantage that it does not converge
to a minimum, but only to a transition state. Convergence, however
can be very slow.

- **Synchronous Transit** The linear synchronous transit (LST) method
generates an estimate of the transition state by finding the highest point
along shortest line connecting two minima. A quadratic synchronous
transit method (QST) extends LST by subsequently searching for a
minimum along a line perpendicular to the previous one. The parabolic
path connecting the minima and the found point may then be searched
for a a maximum. This process of minimum and maximum searches
can be repeated until convergence is obtained.

- **Nudged Elastic Band** NEB works by guessing the Minimum energy
path which connects the two stable structures. A discrete number of
structures (images) are placed along the guessed path. These images
are moved according to: (A) the force acting on them perpendicular to
the path and (B) an artificial spring force keeping the images spaced
along the path. The highest energy image gives a good estimate of the
transition state.

- **String Method** The string method is similar to the NEB. It also in-
volves a series of images along a guess of the MEP, but in this case the
images are moved in two steps. Firstly, the images are moved according
to the force acting on them perpendicular to the path. Using an inter-
polated path, the images are moved short distances along the MEP to
make sure they are evenly spaced.

- **Dimer Method** The dimer method can be used to find possible tran-
sition states without knowledge of the final structure or to refine a good
guess of a transition structure. Thus, it is fundamentally different from
the above methods. The „dimer“ is formed by two images very close
to each other on the potential energy surface. The method works by
moving the dimer uphill from the starting position whilst rotating the
dimer to find the direction of lowest curvature (ultimately negative).
22.3 Nudged Elastic Band Methods


As mentioned, NEB starts by a guessed MEP which is generated as a set of images along the path.

Simplest way of generating the initial guess for the minimum energy path is, obviously, direct interpolation between the two minimum configurations.

- For the true MEP, all the forces acting on the atoms are required to act along the path.
- The relative distance along the MEP defines the reaction coordinate.

The minimum energy path often has one or more local minima (and maxima) in addition to the maximum representing the transition state and the minima of the end points. These correspond to stable intermediate structures.
(and transition states connecting them). For estimating the actual reaction rate $k$, it is critical that the real maximum along the MEP is found instead of one or more possible lower maxima.

In nudged elastic band methods, the spring interaction between the images along the MEP ensures continuity of the path. Hence, the system resembles an elastic rubber band. Optimizing this band, i.e. minimizing the force, brings the band to represent the MEP.

What distinguishes NEB from string method and other elastic band methods is a force projection, which ensures that the spring forces do not interfere with the convergence of the elastic band to the MEP. This also ensures that the true force does not affect the distribution of images along the MEP. By estimating the tangent of the path at each image, the force can be separated into two components: perpendicular and parallel to the MEP. Of these force components only the following are included in the calculation:

- Perpendicular component of the true force.
- Parallel component of the spring force.

This separation of the forces is called „nudging”. As a result of this nudging, the spring components only control the spacing of the images along the band.

The only adjustable parameter in the method is the spring constant. In the original work, the authors mention that it can be varied by several orders of magnitude without affecting the obtained minimum energy path.

### 22.4 NEB algorithm

- The elastic band with $N + 1$ images can be denoted by
  \[ [R_0, R_1, R_2, \ldots, R_N] \] (22.3)

- End points $R_0$ and $R_N$ are fixed and given by the energy minima corresponding to the initial and final states.

- The $N - 1$ middle images are adjusted by the optimization algorithm.

- As mentioned before, the force acting on the images is
  \[ F_i = F_i^S|_{||} - \nabla u(R_i)|_\perp \] (22.4)
The true force is given by
\[ \nabla U(R_i)\perp = \nabla U(R_i) - \nabla U(R_i) \cdot \hat{\tau}_i \]
(22.5)

Here \( \hat{\tau}_i \) is the normalized local tangent at image \( i \).

The spring force is given as
\[ F^S_i \| = k(|R_{i+1} - R_{i}| - |R_{i} - R_{i-1}|) \cdot \hat{\tau}_i \]
(22.6)

where \( k \) is the spring constant.

An optimization algorithm is used to move the images according to the force \( F_i \).
In principle, any optimization algorithm can be used, e.g. conjugate gradient method. In the original work, the authors used velocity Verlet algorithm.

The spacing between the images is determined by \( k \), and is thus a constant along the band if \( k \) is not varied. Therefore, it should come as no surprise that typically none of the images occurs exactly at the saddle point. Thus, the saddle point energy needs to be interpolated from the results. Even if carried out with care, the interpolation can result in large errors in the results. Luckily this problem can be solved by a rather small modification to the NEB algorithm.

### 22.5 Climbing-image NEB (CI-NEB)

Within the climbing-image NEB, the minimum energy path is not changed, but a rigorous convergence to the saddle point is also obtained. Also, it doesn’t add a significant computational effort.

- In CI-NEB, after a few iterations of NEB, the image with the maximum energy is determined; this image is then driven to the saddle point.
- The force on this image is given by
\[ F_{i_{\text{max}}} = -\nabla U(R_{i_{\text{max}}}) + 2\nabla U(R_{i_{\text{max}}}) \| \]
\[ = -\nabla U(R_{i_{\text{max}}}) + 2\nabla U(R_{i_{\text{max}}}) \cdot \hat{\tau}_{i_{\text{max}}} \hat{\tau}_{i_{\text{max}}} \]
(22.7)
• This represents the full force due to the potential with the component along the elastic band inverted.

• So, the climbing image moves up the potential energy surface along the elastic band and down the potential surface perpendicular to the band.

• The maximum energy image is not affected by the spring forces at all.

• Other images define the one degree of freedom for which the energy maximization is carried out.

• Obviously, this method is much better than the standard NEB for localising the actual saddle point.

As the climbing image is not affected by the spring forces, the spacing between the images becomes non-uniform. However, since the saddle point is the most important point along the MEP, it would be desirable if the area around this point would be especially well sampled. Also, if the images get closer, the approximation of the tangent becomes better which leads to more accurate estimate for the saddle point.

Such an uneven sampling (tighter around the saddle point) can be achieved by using a variable spring constant. One approach to do this, is to make the spring constant linearly dependent on the potential energy. This will automatically lead to a better sampling of the important part of the MEP.

• One possible choice for the variable spring constant is

\[ k'_i = \begin{cases} 
  k_{\text{max}} - \Delta k \left( \frac{U_i - U_{\text{ref}}}{U_{\text{max}} - U_{\text{ref}}} \right) & \text{if } U_i > U_{\text{ref}} \\
  k_{\text{max}} - \Delta k & \text{else}
\end{cases} \]  

(22.8)
22.6 Combining the methods

\( U_i = \max \{U_i, U_{i-1}\} \) is the higher energy of the two images connected by spring \( i \), \( U_{\text{max}} \) is the maximum value of \( U_i \) for the band, and \( U_{\text{ref}} \) is a reference value for the energy, e.g., the energy of the higher energy endpoint of the MEP.

This way, \( k \) is scaled linearly from a maximum value for highest energy images to a minimum value for images with energy of \( U_{\text{ref}} \) or lower.

22.6 Combining the methods

The presented methods have different strengths and weaknesses. Climbing-image NEB, for example, is very efficient in finding the saddle point, but at the expense of poor resolution for the rest of the path, if the number of images (computational cost) is small.

Therefore, it can be useful to combine the methods, for example:

1. Use a minimal (5-image) climbing-image NEB to find the saddle point.

2. Then use a cheap optimizer (e.g. stepwise descent, steepest descent) to calculate the MEP down towards the minima.

Test have shown this to be superior in efficiency compared to NEB, at least in a simple test system:
How well the method works, depends of course on the complexity of the potential surface.

22.7 Implementing NEB

Integrating the NEB method into an existing MD code is straightforward. The following steps are involved in the process:

1. Input the coordinates for all atoms in all images of the band.
2. Construct a neighbour list excluding atom pairs which belong to different images.
3. Calculate the tangent $\hat{\tau}_i$.
4. Add the calculation of the spring force.
5. Modify the force routine to calculate only the perpendicular component of the force.
6. Choose a force constant $k$.
   • Run MD until the minimum force is reached.

Summary

• Many processes require an activation energy for overcoming a barrier, e.g., defect migration and chemical processes.
According to transition state theory, rate of occurrence for such rare events can be calculated if the characteristic vibrational frequencies and the energy barrier are known.

MD can be used to evaluate the energy barriers if the simulation time is long enough for obtaining the rate, but this is usually not possible.

For such cases, methods for finding the minimum energy path (MEP) have been developed, e.g., the climbing image nudged elastic band method.

In this method, a series of images created along a guessed MEP, joined by springs, and allowed to relax into the actual MEP by including only forces along direction perpendicular to the MEP tangent.
Exam

- Wed 11.12.2013 at 10.00-14.00 (note the extended time and different place compared to the lectures): D106, Physicum.

- Exam will include questions that can be
  1. Explanation of concepts.
  2. Broader essay questions, e.g. for a given problem, what kind of tools could be used to solve it.
  3. Simple calculations.

- You do not need to remember the exact mathematical forms of different potentials or models. However, the basic forms and the theoretical basis of different models should be known

- No 'real' code writing will be asked in the exam.

- Preparation material: Lecture notes.