Introduction in Molecular Modeling

Herma Cuppen

10th September 2020

Contents

1 Introduction **1**

2 Potential Energy Surface **3**
2.1 Stationary points . . . 4
2.2 Reaction coordinate . . . 5
2.3 How to obtain the Potential Energy Surface? . . . 6

3 Molecular Mechanics and Force fields **7**
3.1 Mutual Potential Energy . . . 8
3.2 Electrostatic interactions . . . 9
3.2.1 Polarizability . . . 12
3.2.2 Electrostatic interactions in force fields . . . 13
3.3 Intermolecular interactions . . . 14
3.4 Intramolecular interactions . . . 15
3.4.1 Stretching . . . 15
3.4.2 Bending . . . 16
3.4.3 Four-body terms . . . 17
3.5 How to parameterize a force field . . . 17

4 Geometry Optimization **19**
4.1 Derivative-free methods . . . 20
4.1.1 Multivariate Grid Search . . . 20
4.1.2 Simulated Annealing . . . 20
4.2 First-order methods . . . 20
4.2.1 Steepest descent . . . 21
4.2.2 Conjugate gradients . . . 21
4.3 Second-order methods . . . 21
4.4 Newton-Raphson . . . 22
4.5 Numerical Hessian . . . 22

5 Reaction path finding methods **23**
5.1 Double-ended methods . . . 23
5.2 Nudged Elastic Band method . . . 23
5.3 Transition Path Sampling . . . 24
5.4 Single-ended methods . . . 26

6 Statistical Thermodynamics and Transition State Theory **27**
6.1 The average mutual potential energy of two dipoles . . . 28
6.2 Transition State Theory . . . 29
6.2.1 Harmonic Transition State Theory . . . 32

7 Properties **33**
7.1 Radial Distribution Function 33
7.2 Pair Correlation Functions 34
7.3 Diffusion 34

8 Molecular Dynamics **37**
8.1 Initialization 38
8.2 Force Calculation 39
8.3 Periodic Boundary Conditions 40
8.4 Ewald summation 40
8.5 Verlet list 41
8.6 Integration 42
8.7 Simulations in different ensembles 44

9 Metropolis Monte Carlo **47**
9.1 Integration by Monte Carlo 48
9.2 Importance sampling 48
9.3 Umbrella sampling 50
9.4 Microscopic reversibility 51

10 Kinetic Monte Carlo **53**
10.1 Derivation of the master equation 55
10.2 Derivation of time evolution 56
10.3 Microscopic reversibility 56
10.4 KMC algorithm 57
10.5 Differences between Kinetic Monte Carlo and Metropolis Monte Carlo 59
10.6 Off-lattice methods 60

11 Free Energy Calculations **63**
11.1 Einstein Integration 64
11.2 Widom test particle insertion method 65

Chapter 1

Introduction

Computational Chemistry is widely used as a tool in chemistry and molecular sciences. It allows one to compute properties of molecular systems like geometry, spectra, reaction paths and mechanisms, band structure, (free) energy, diffusion constants, etc.. This syllabus aims to discuss different methods, *e.g.*, optimization methods, molecular dynamics, reaction path search algorithms, that can be applied to obtain these properties. This syllabus evolved from a collection of lecture notes and it should still be considered as such. Generally, in computational chemistry one starts from a systems (left starting point in Figure 1.1) and one would ultimately like to know some properties of this system (right end point in the flowchart). To achieve this one needs information about the energy, electron distribution or molecular orbitals as input. I have called this "energy function". This syllabus will only briefly discuss the use of force fields in this context and not go into detail on other methods to obtain energy. Next, we need to preform a geometry optimization and using this structure one can perform molecular simulations during which one can calculate the properties of interest.

This syllabus is organized to follow this flowchart. First, the basic concepts behind potential energy surfaces are described (Chapter 2). Then the use of force fields is discussed (Chapter 3). We are then set to start the introduction of different modeling methods. I will start with geometry optimization (Chapter 4) and continue with reaction path search algorithms (Chapter 5) which heavily lean on geometry optimization techniques. These are all however techniques that apply at 0 K. Chapter 6 will then briefly refresh the background of statistical thermodynamics which provides us with the tools to go to finite temperatures and transition state theory will be introduced. Chapters 8, 9, and 10 discuss three different modeling techniques: Molecular Dynamics, Metropolis Monte Carlo, and Kinetic Monte Carlo, respectively. Chapters 7 and 11 explain how these techniques can be applied to determine several properties of a system including the free energy.

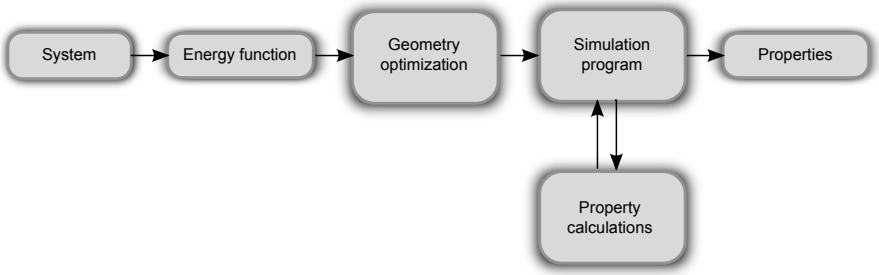


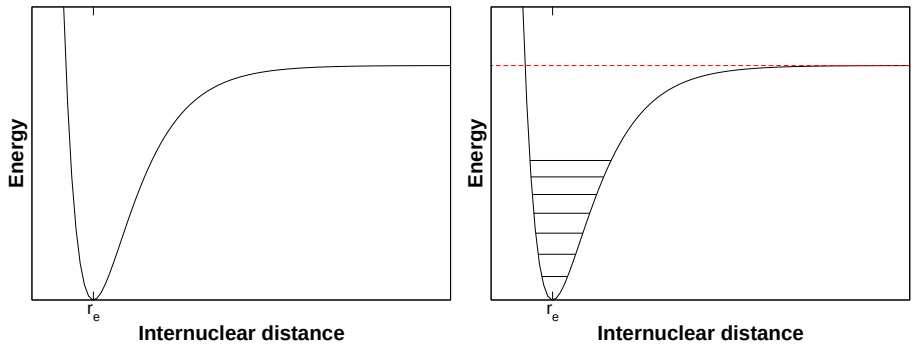
Figure 1.1: *Flowchart for a general modeling scheme to determine properties of a certain system.*

Chapter 2

Potential Energy Surface, Stationary points and reaction coordinate

Central to all subjects in this course is the potential energy surface (PES). Let us consider a diatom AB. We can plot the energy of this molecule as a function of the bond length (see Figure 2.1a). Since the model is motionless while we hold it at the new geometry, this energy is not kinetic and so is by default *potential* ("depending on position"). A PES is a plot of the energy of a collection of nuclei and electrons against the geometric coordinates of the nuclei. The reason why this can be done and makes sense lies in the Born–Oppenheimer approximation: the nuclei in a molecule are stationary with respect to the electrons.

Real molecules vibrate about the equilibrium bond length, so that they always possess kinetic energy (T) and/or potential energy (V). The fact that a molecule is never actually stationary with zero kinetic energy (*zero-point energy* (ZPE)) is usually shown on potential energy/bond length diagrams by drawing a series of lines above the bottom of the curve (see Figure 2.1b). Figure 2.1 represents an one-dimensional (1D) PES. This could be the potential energy surface of a diatomic system in the gas phase. We are usually interested in more complex systems. Generally isolated non-linear triatomic molecules have a 3 internal coordinates ($3N - 6$), linear molecules have $3N - 5$ internal coordinates. To take H_2O as an example, a single molecule as $3 \times 3 - 6 = 3$ coordinates: the two O-H distances and the angle between the two O-H bonds. One would need a 3D potential to describe the full system. We could assume that the two O-H bond lengths are equal and we could make an approximate 2D PES, which can be plotted in a 3D graph. 1D and 2D potentials can still be easily visualized; higher order potentials not and usually cuts are given with several parameters kept constant.



(a) (b)

Figure 2.1: *Example of an one-dimensional (1D) potential energy surface, without (a) and with (b) distinct vibrational levels.*

If one would describe a system consisting of two H_2O molecules, one has $3N - 6 = 3 \times 6 - 6 = 12$ independent coordinates describing the system: three coordinates for each molecule, the distance between the two molecules and five angles to describe the placement of the molecules with respect to each other. We see here already the concept of intramolecular and intermolecular PESs; we will return to this when discussing force fields. For simple species, it is easier and more intuitive to use a coordinate system that is based on internal coordinates, *i.e.*, bond lengths, angles, torsions, distances etc. These coordinates are usually labeled q_1, q_2, etc. For more complex systems often Cartesian coordinates are used where each atom gets three coordinates: (x_n, y_n, z_n) or $\boldsymbol{r_n}$. In this coordinate system no information about the connectivity is included and the system is described by $3N$ coordinates whereas there are only $3N - 6$ degrees of freedom. Translation and rotation of the whole system leads to degenerate results.

2.1 Stationary points

Let us consider the 2D PES shown in Figure 2.2. This PES has three stationary points. Stationary points are defined by

$$\frac{\partial E}{\partial q_1} = \frac{\partial E}{\partial q_2} = \cdots = \frac{\partial E}{\partial q_n} = 0. \tag{2.1}$$

These are minima, maxima and saddle points. Minima and maxima can be distinguished by their second derivatives. The easiest way to do this is to construct a Hessian second derivative matrix

$$H = \begin{bmatrix} \dfrac{\partial^2 E}{\partial q_1^2} & \dfrac{\partial^2 E}{\partial q_1 \, \partial q_2} & \cdots & \dfrac{\partial^2 E}{\partial q_1 \, \partial q_n} \\ \dfrac{\partial^2 E}{\partial q_2 \, \partial q_1} & \dfrac{\partial^2 E}{\partial q_2^2} & \cdots & \dfrac{\partial^2 E}{\partial q_2 \, \partial q_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 E}{\partial q_n \, \partial q_1} & \dfrac{\partial^2 E}{\partial q_n \, \partial q_2} & \cdots & \dfrac{\partial^2 E}{\partial q_n^2} \end{bmatrix} \tag{2.2}$$

and solve the following eigenvalue problem

$$\mathbf{H} = \mathbf{P}\mathbf{k}\mathbf{P}^{-1}. \tag{2.3}$$

The $\mathbf{P}$ matrix is the eigenvector matrix whose columns are "direction vectors" for the vibrations. The corresponding force constants are given by the $\mathbf{k}$ eigenvalue matrix. For a minimum in the PES,

$$k_i > 0 \tag{2.4}$$

for all k. For a maximum in the PES

$$k_i < 0 \tag{2.5}$$

for all k_i. All other stationary points are saddle points, where the number of negative eigenvalues or frequencies indicate the order of the saddle point. In chemistry we are generally interested in first order saddle points. A first order saddle point is often referred to as a *transition state* or an *activated complex*. Sometimes the distinction between transition state and transition structure is made. The transition state is the saddle point on the FREE energy surface, which represents the true activated complex, while the transition structure is a saddle point on a calculated enthalpy surface. The two differ by an entropy (TS) term. As we will see in Chapter 11, the entropy term is hard to calculate routinely and is therefore often ignored. Still the term transition state is used. Also ZPE is often ignored, but since the ZPE at the minima and the saddle point roughly cancel, the error in the relative enthalpy difference, which determines the activation energy or the barrier height, is small.

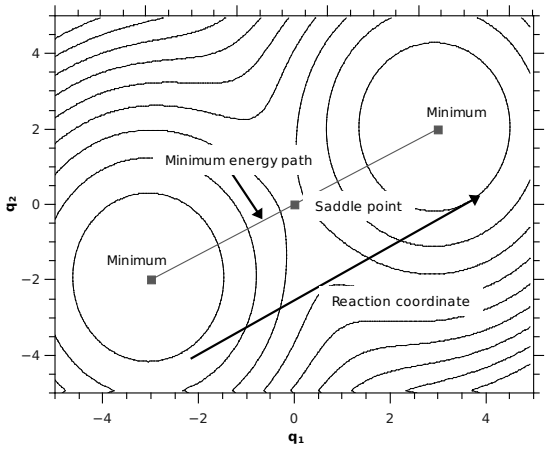


Figure 2.2: *Example 2D PES with two minima connected by a saddle point.*

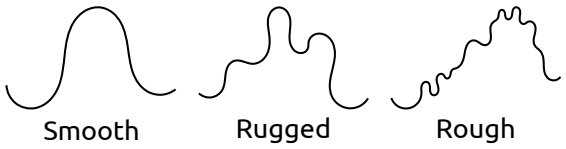


Figure 2.3: *Schematic representation of smooth, rugged, and rough potential energy surfaces.*

The PES in Figure 2.2 is very "smooth" with clear minima and one saddle point connecting them. Potential energy surfaces can however be much complicated. One speaks about a surface being "rugged" when the surface is still relatively smooth but the interesting deep wells are separated by several local minima or the surface can be "rough" if the interesting deep wells are separated many "minima" which are difficult to define or hard to count. These different type of potential energy surfaces are depicted in Figure 2.3.

2.2 Reaction coordinate

Let us return to Figure 2.2. Here the two minima represent two stable conformations. The system can go from one conformation to the other through the transition state at the saddle point. The line connecting the two local minima along the *minimum energy path* gives the *reaction coordinate*. In Figure 2.2 the reaction coordinate is a linear combination of two parameters (bond lengths, angles and dihedrals). When applying an internal coordinate system, the reaction coordinate is more likely to coincide with coordinates of the coordinate system than for Cartesian coordinates. Mathematically the reaction coordinate is given by the eigenvector that belongs to negative frequency of the saddle point.

On a high-dimensional PES it is however computationally expensive to obtain the transition state and reaction coordinate in this way. Then often the reaction coordinate is guessed based on chemical intuition and the (free) energy along this coordinate is maximized to find the transition state. Figure 2.4 indicates what can happen if a wrong choice of the reaction coordinate is made. The positions of A and B suggest that the reaction coordinate can be purely q independent of q'. The reaction coordinate would then follow the horizontal black line. Calculating the (free) energy landscape leads to a maximum which is not the true transition state (when checking the gradient in all other directions (costly) one would indeed find that is not a saddle point). The wrong initial reaction coordinate leads to hysteresis and can identify the wrong mechanism and kinetics. Furthermore, if one uses the wrong reaction coordinate to determine in which state the system is (A or B), one could be completely off. The red lines in Figure 2.4 indicate

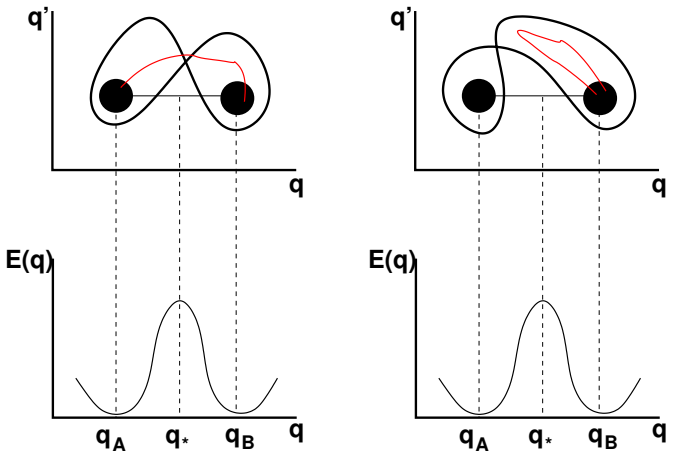


Figure 2.4: *Two schematic 2D PESs and a cut through this landscape along q as indicated by the line. We see here that the exact choice in the reaction coordinate is crucial to determine the correct behaviour.*

possible dynamics trajectories. Based on the chosen reaction coordinate, one would think that the trajectory right panel brought the system from B to A and back, since q_* was crossed. This would be counted as two barrier crossings. But in the full-dimensional picture one sees that the system stayed in state B the entire trajectory.

2.3 How to obtain the Potential Energy Surface?

The energy of a system can be calculated in different ways and generally four classes are distinguished: *ab initio* calculations, semiempirical calculations, Denisty Functional Theory, and Molecular mechanics.

Ab initio calculations start from the Schrödinger equation. As you will be aware this equation cannot be exactly solved for more than one electron. This has lead to a range of different methods with different approximations. *Ab initio* calculations give the energy and the wave function as an output. One of the basic methods is the Hartree-Fock method and most other *ab initio* methods have build on this. For this reason, there often referred to as post-Hartree-Fock methods. Hartree-Fock will be discussed in detail in the Quantum Chemistry course. The output of these methods can be very accurate, but the calculations are slow and scale quite dramatically with system size. They are therefore usually only applied to small gas phase molecules.

Semiempirical calculations also start from the Schrödinger equation but use empirical parameterization for some integrals in the calculation. This makes them faster and allows for a bit larger molecules. At the same time the accuracy goes down.

Denisty Functional Theory uses an electron density instead of wave functions. It is often referred to as an *ab initio* method as well, although strictly speaking it is not, since it requires an unknown functional to describe the electron correlation. These functionals are often parameterized to reproduce experimental results. The method is still rather slow (depending on the functional), but scales a bit better with system size and it can be used for periodic systems such as crystals. This method will also be discussed in the Quantum Chemistry course.

Finally, Molecular Mechanics uses a force field, *i.e.*, a set of equations that describe "balls on springs". The input parameters are the force constants and the analytical description of the different components. The calculations are extremely fast as compared to the other methods. It is therefore well-suited to use on systems with very large (bio)molecules and solids in combination with methods that require many force calls. This is typically the case for the methods that will be discussed in this course. We will therefore limit ourselves to the use of force fields.

Chapter 3

Molecular Mechanics and Force fields

Molecular Mechanics uses a mechanical model to describe a molecule and to find its minimum-energy geometry. The energy of a molecule is expressed as a function of its resistance toward bond stretching, bond bending, and atom crowding. This model is conceptually very close to the intuitive feel for molecular energetics that one obtains when manipulating molecular models of plastic or metal; it completely ignores electrons. The mathematical expressions that are used to calculate the energy constitute a force field which is parameterized by force constants. A force field is generally divided in an intramolecular part and an intermolecular part which can be further divided in Van der Waals terms and Coulomb interactions

$$U_{\mathrm{MM}} = U_{\mathrm{intra}} + \sum_{\mathrm{pairs}} U_{\mathrm{vdw}} + \sum_{\mathrm{pairs}} U_{\mathrm{coul}}. \tag{3.1}$$

Sometimes additional terms that account for hydrogen bonding are added as well. The intramolecular term can be further divided in

$$U_{\mathrm{intra}} = \sum_{\mathrm{bonds}} U_{\mathrm{stretch}} + \sum_{\mathrm{angles}} U_{\mathrm{bend}} + \sum_{\mathrm{dihedrals}} U_{\mathrm{torsion}} + \cdots \tag{3.2}$$

The input for this method is not only the position of each atom but also which atoms are connected and with what kind of bond (single, double, resonant etc). In electronic structure calculations this bonding information is not an input but rather an output of the calculation (given by the amount of overlap and electron density).

Some force fields only deal with the intramolecular terms, others only with the van der Waals component. The Coulomb interactions often have to be calculated separately, but there are also force fields that have parameterized these interactions as well and cover the full package: intramolecular, Van der Waals and Coulomb interactions.

Force fields cannot be applied to calculate electronic properties like charge distributions or nucleophilic and electrophilic behaviour, since electrons are ignored in the description. An advantage of the application of force fields is that the results are conceptually easy to interpret. It is computationally cheap and can therefore be used for large systems (biomolecules etc) or in combination with modeling methods that require many ($>$ millions) force calls, like Molecular Dynamics and Monte Carlo simulations. Since the energy is calculated using mathematical expressions, the first derivative of the energy

$$\nabla U = \left(\frac{\partial U}{\partial x}\mathbf{e}_x + \frac{\partial U}{\partial y}\mathbf{e}_y + \frac{\partial U}{\partial z}\mathbf{e}_z \right) \tag{3.3}$$

which is often needed for different methods can be easily calculated as well. When the parameterization is done well, very accurate results can be obtained, but one should check against other methods (*ab initio*, DFT).

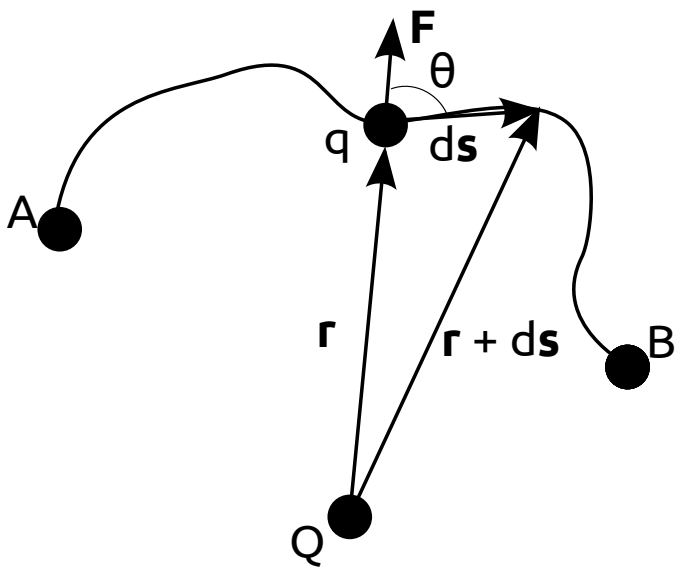


Figure 3.1: *Force that works on point charge q due to point charge Q when traveling from A to B.*

3.1 Mutual Potential Energy

Let us calculate the work that is needed to bring a point charge q from point A to point B near a point charge Q (see Figure 3.1). The work on the system w.r.t. its surrounding is

$$w_{\mathrm{on}} = -\int \mathbf{F}_{\mathrm{on\,q}} \cdot \mathrm{d}\mathbf{s} \tag{3.4}$$

The force on q due to Q is

$$\mathbf{F}_{\mathrm{Q\,on\,q}} = \frac{1}{4\pi\epsilon_0\epsilon_r}\frac{Qq}{r^3}\mathbf{r} \tag{3.5}$$

with ϵ_r the relative permittivity and ϵ_0 the permittivity in free space. According to Newton's Third Law

$$\mathbf{F}_{\mathrm{Q\,on\,q}} = -\mathbf{F}_{\mathrm{q\,on\,Q}} \tag{3.6}$$

So the work becomes

$$w_{\mathrm{on}} = -\int \frac{1}{4\pi\epsilon_0\epsilon_r}\frac{Qq}{r^3}\mathbf{r} \cdot \mathrm{d}\mathbf{s}. \tag{3.7}$$

We can use

$$\mathbf{r} \cdot \mathrm{d}\mathbf{s} = r\mathrm{d}r \tag{3.8}$$

which gives

$$\begin{aligned} w_{\mathrm{on}} &= -\int \frac{1}{4\pi\epsilon_0}\frac{Qq}{r^2}\mathrm{d}r \\ &= -\int_{\mathrm{A}}^{\mathrm{B}} \frac{1}{4\pi\epsilon_0}\frac{Qq}{r^2}\mathrm{d}r \\ &= \frac{1}{4\pi\epsilon_0}Qq\left(\frac{1}{r_{\mathrm{B}}} - \frac{1}{r_{\mathrm{A}}}\right) \end{aligned} \tag{3.9}$$

So the work only depends on the initial A and final B position and is independent of route connecting them. This can be applied to determine the Mutual Potential Energy, U or Φ, which is the work to bring a charge q from infinity to point $\mathbf{r}$.

$$U = \frac{1}{4\pi\epsilon_0}\frac{Qq}{r} \tag{3.10}$$

The total of Mutual Potential Energy of a system is the work done to build up this system. The electric force is *pairwise additive* which means that the force on a point charge A due to multiple point charges can determined by simply adding the individual contributions

$$\mathbf{F}_{\text{on A}} = \frac{Q_{\text{A}}}{4\pi\epsilon_0\epsilon_r}\sum_{i\neq\text{A}}\frac{Q_i}{R_{i\text{A}}^3}\mathbf{R}_{i\text{A}}. \tag{3.11}$$

The work and hence the mutual potential energy is therefore also pairwise additive and one gets

$$U_{2\text{ species}} = U_{12} \tag{3.12}$$
$$U_{3\text{ species}} = U_{12} + U_{13} + U_{23} \tag{3.13}$$
$$U_{4\text{ species}} = U_{12} + U_{13} + U_{23} + U_{14} + U_{24} + U_{34}. \tag{3.14}$$

The total energy of the system is the sum of the kinetic energy and the potential energy

$$E_{\text{total}} = \frac{1}{2}mv^2 + U. \tag{3.15}$$

Because of conservation of energy

$$\frac{\mathrm{d}E_{\text{total}}}{\mathrm{d}t} = 0 \tag{3.16}$$

$$\frac{\mathrm{d}E_{\text{total}}}{\mathrm{d}t} = mv\frac{\mathrm{d}v}{\mathrm{d}t} + \frac{\mathrm{d}U}{\mathrm{d}t} = mv\frac{\mathrm{d}v}{\mathrm{d}t} + \frac{\mathrm{d}U}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}t} = mva + \frac{\mathrm{d}U}{\mathrm{d}x}v = 0 \tag{3.17}$$

$$v\left(ma + \frac{\mathrm{d}U}{\mathrm{d}x}\right) = v\left(F + \frac{\mathrm{d}U}{\mathrm{d}x}\right) = 0 \tag{3.18}$$

$$F = -\frac{\mathrm{d}U}{\mathrm{d}x}. \tag{3.19}$$

In three dimensions this becomes

$$\mathbf{F} = -\nabla U = -\left(\frac{\partial U}{\partial x}\mathbf{e}_x + \frac{\partial U}{\partial y}\mathbf{e}_y + \frac{\partial U}{\partial z}\mathbf{e}_z\right). \tag{3.20}$$

This expression is used in Molecular Dynamics simulations to calculate the forces on atoms. This does not only hold for the potential energy due to electrostatic interactions but also for other types of interactions, which we will discuss later, as long as they are pairwise additive. Since force fields use analytical descriptions to describe the potential energy, the force can be derived analytically as well, hence the term “force fields”.

3.2 Electrostatic interactions

Electrostatic or Coulomb interactions are usually considered to be between separate molecules, for instance between two water molecules, or between atoms that are at least three or four bonds apart. The latter is especially important for large (bio)molecules or polymorphs which can fold up such that atoms can interaction which are part of the same molecule but many bonds apart.

The dipole moment of a molecule consisting of several atoms at location $\mathbf{R}_i$ that can be described by point charges with charge Q_i is

$$\mathbf{p}_e = \sum_i Q_i\mathbf{R}_i \tag{3.21}$$

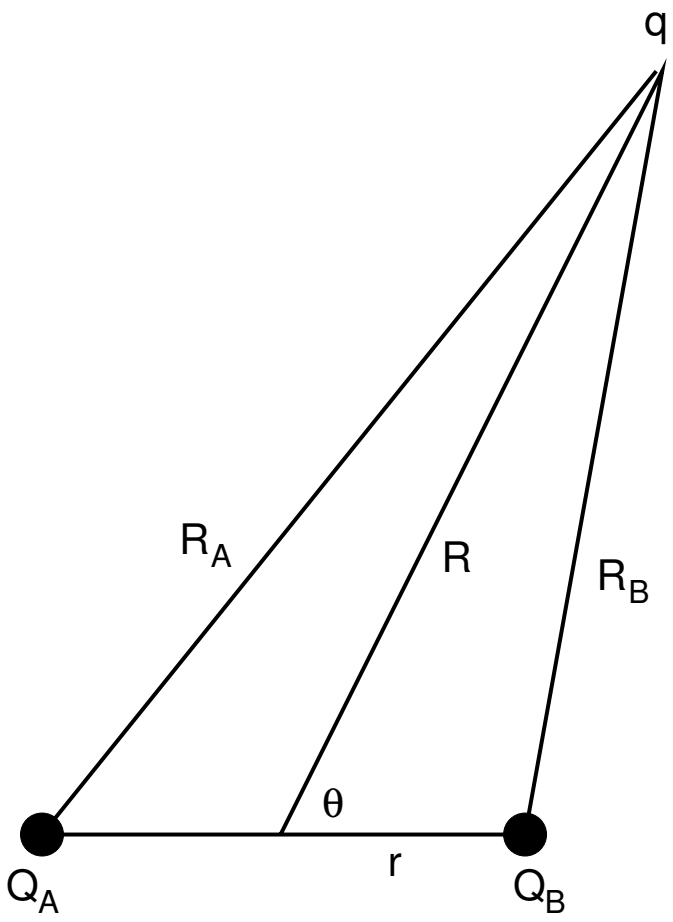


Figure 3.2: *Dipole A-B with respect to charge q.*

If we now move the origin such that $\mathbf{R}'_\mathbf{i} = \mathbf{R_i} + \mathbf{\Delta}$ the dipole moment becomes

$$\begin{aligned} \mathbf{p}'_e &= \sum_i Q_i \mathbf{R}'_i = \sum_i Q_i (\mathbf{R}_i + \mathbf{\Delta}) \\ \mathbf{p}'_e &= \mathbf{p}_e + \mathbf{\Delta} \sum_i Q_i. \end{aligned} \tag{3.22}$$

This leads to the same result as Eq. (3.21) for neutral systems $\sum_i Q_i = 0$. For charged systems, ions for instance, the origin needs to be defined when giving a dipole moment since for those systems the dipole moment is not gauge invariant (independent of the choice of the origin).

The energy between two point charges is:

$$U(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{Q_A Q_B}{r} \tag{3.23}$$

Between a point charge q and an electric dipole with charge Q_A and Q_B as depicted in Figure 3.2, the interaction is

$$U(\mathbf{R}) = \frac{q}{4\pi\epsilon_0} \left(\frac{Q_\mathrm{A}}{R_\mathrm{A}} + \frac{Q_\mathrm{B}}{R_\mathrm{B}} \right). \tag{3.24}$$

$$U(\mathbf{R}) = \frac{q}{4\pi\epsilon_0}\left(\frac{Q_\mathrm{A}}{\sqrt{(r+R\cos\theta)^2+(R\sin\theta)^2}} + \frac{Q_\mathrm{B}}{\sqrt{(r-R\cos\theta)^2+(R\sin\theta)^2}}\right)$$
$$U(\mathbf{R}) = \frac{q}{4\pi\epsilon_0}\left(\frac{Q_\mathrm{A}}{\sqrt{R^2+2Rr\cos\theta+r^2}} + \frac{Q_\mathrm{B}}{\sqrt{R^2-2Rr\cos\theta+r^2}}\right)$$
$$U(\mathbf{R}) = \frac{q}{4\pi\epsilon_0}\left(\frac{Q_\mathrm{A}}{\sqrt{R^2+2Rr\cos\theta+r^2}} + \frac{Q_\mathrm{B}}{\sqrt{R^2-2Rr\cos\theta+r^2}}\right)$$
$$U(\mathbf{R}) = \frac{q}{4\pi\epsilon_0 R}\left(\frac{Q_\mathrm{A}}{\sqrt{1+2r/R\cos\theta+(r/R)^2}} + \frac{Q_\mathrm{B}}{\sqrt{1-2r/R\cos\theta+(r/R)^2}}\right) \tag{3.25}$$

We assume that the distance between the charge and the dipole is larger than the distance within the dipole ($t = r/R < 1$). This is usually the case when calculating the relevant electrostatic potential for a molecular dipole. We can therefore use $\frac{1}{\sqrt{1-2xt+t^2}} = \sum_{l=0}^{\infty} t^l P_l(x)$ with P_l the Legendre polynomial

$$U(\mathbf{R}) = q\frac{Q_\mathrm{A}\left[1-\frac{r}{R}\cos\theta+\frac{r^2}{2R^2}(3\cos^2\theta-1)\right] + Q_\mathrm{B}\left[1+\frac{r}{R}\cos\theta+\frac{r^2}{2R^2}(3\cos^2\theta-1)\right]}{4\pi\epsilon_0 R}$$
$$U(\mathbf{R}) = \frac{q}{4\pi\epsilon_0}\left(\frac{Q_\mathrm{A}+Q_\mathrm{B}}{R} + \frac{Rp\cos\theta}{R^3} + (Q_\mathrm{A}+Q_\mathrm{B})\frac{R^2r^2}{2R^5}(3\cos^2\theta-1)\right) \tag{3.26}$$

Use $\mathbf{a}\cdot\mathbf{b} = ab\cos\theta$

$$U(\mathbf{R}) = \frac{q}{4\pi\epsilon_0}\left(\frac{Q_\mathrm{A}+Q_\mathrm{B}}{R} + \frac{\mathbf{pR}}{R^3} + \frac{(Q_\mathrm{A}+Q_\mathrm{A})(3(\mathbf{r}\cdot\mathbf{R})^2-r^2R^2)}{2R^5}\right) \tag{3.27}$$

Or more general

$$U(\mathbf{r}) = \frac{q}{4\pi\epsilon_0}\left(\frac{\sum Q_i}{r} + \frac{\mathbf{p}\cdot\mathbf{r}}{r^3} + \frac{\sum Q_i(3(\mathbf{r}_i\cdot\mathbf{r})^2-r_i^2r^2)}{2r^5}\right) \tag{3.28}$$

The first term represents the contribution of the first electric moment (charge), the second of the second electric moment (dipole moment), and the third the second electric moment (quadrupole moment). Higher order terms that are cut off from the expansion represent higher order terms.

The quadrupole moment can be defined similarly to the dipole moment

$$\mathbf{\Theta}_e = \frac{1}{2}\begin{pmatrix} \sum Q_i(3X_i^2-R_i^2) & 3\sum Q_iX_iY_i & 3\sum Q_iX_iZ_i \\ 3\sum Q_iY_iX_i & \sum Q_i(3Y_i^2-R_i^2) & 3\sum Q_iY_iZ_i \\ 3\sum Q_iZ_iX_i & 3\sum Q_iZ_iY_i & \sum Q_i(3Z_i^2-R_i^2) \end{pmatrix}. \tag{3.29}$$

This has zero trace and is gauge invariant when the system is charge neutral and has no dipole moment. Using this definition the charge-dipole potential energy becomes

$$U(\mathbf{r}) = \frac{q}{4\pi\epsilon_0}\left(\frac{\sum Q_i}{r} + \frac{\mathbf{p}\cdot\mathbf{r}}{r^3} + \frac{\mathbf{\Theta}_e\cdot\mathbf{r}^2}{2r^5}\right). \tag{3.30}$$

For neutral dipole and ignoring the quadrupole interaction the mutual potential energy from Eq. (3.27) becomes

$$U(\mathbf{r}) = \frac{q}{4\pi\epsilon_0}\frac{\mathbf{p}\cdot\mathbf{r}}{r^3} \tag{3.31}$$

which falls off as $1/r^2$ which is faster than $1/r$ for the point charges. For molecular systems however dipole-dipole interactions are more common than charge-dipole interactions. Figure 3.3 shows such a system with the following relationships

$$Q_{A1} + Q_{A2} = Q_A, \tag{3.32}$$

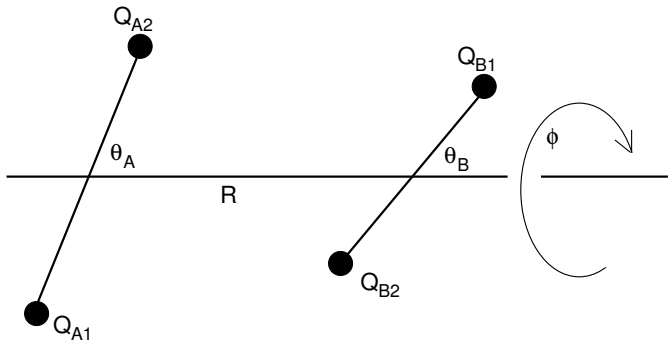


Figure 3.3: *Coordinate definition for a two dipole system.*

Table 3.1: Dipole-dipole mutual potential energies for different configurations

θ_A	θ_B	U (dipole-dipole)
0	0	$-\frac{2p_Ap_B}{4\pi\epsilon_0 R^3}$
0	π	$+\frac{2p_Ap_B}{4\pi\epsilon_0 R^3}$
0	$\pi/2$	0
$\pi/2$	$\pi/2$	$\frac{p_Ap_B}{4\pi\epsilon_0 R^3}\cos\phi$

$$Q_{B1} + Q_{B2} = Q_B, \tag{3.33}$$

p_A and p_B are the magnitudes of the dipole moment, and Θ_A, Θ_B the magnitudes of the quadrupole moment. Then the mutual potential energy between dipoles A and B is

$$\begin{aligned}(4\pi\epsilon_0)U_{\rm AB} &= \frac{Q_AQ_B}{R} + \frac{1}{R^2}\left(Q_Bp_A\cos\theta_A - Q_Ap_B\cos\theta_B\right)\\ &\quad -\frac{p_Ap_B}{R^3}\left(2\cos\theta_A\cos\theta_B - \sin\theta_A\sin\theta_B\cos\phi\right)\\ &\quad +\frac{1}{2R^3}\left(Q_A\Theta_B\left(3\cos^2\theta_B - 1\right) + Q_B\Theta_A\left(3\cos^2\theta_A - 1\right)\right) + \cdots\end{aligned} \tag{3.34}$$

Again this expression does not include higher order terms due to dipole-quadrupole, and charge-octupole interactions or even higher order terms. For neutral molecules all charge contributions cancel and we only have

$$U_{\rm AB} = -\frac{p_Ap_B}{4\pi\epsilon_0 R^3}\left(2\cos\theta_A\cos\theta_B - \sin\theta_A\sin\theta_B\cos\phi\right). \tag{3.35}$$

Table 3.1 gives the interaction for some special configurations.

3.2.1 Polarizability

The electric moments of a molecule can change depending on their environment. An electric field $\mathbf{E}$ caused by, for instance, a neighboring molecule can induce a dipole

$$\mathbf{p}_e(induced) = \mathbf{p}_e(permanent) + \alpha\mathbf{E} + \ldots. \tag{3.36}$$

The extend of the induced dipole depends on polarizability of the molecule in which the dipole is induced. This polarizability is a tensor α of rank 2 and is and represented by a 3×3 matrix. This matrix can always be written in diagonal form by a suitable rotation of the Cartesian axes. The quantities α_{aa}, α_{bb} and α_{cc} are called the principle values of the polarizability tensor. Let

us assume a molecule A with and a molecule B without permanent dipole. Molecule A can then induce a dipole in molecule B. The corresponding induction energy is

$$U_{\mathrm{AB\ induced}} = -\frac{1}{2}\alpha_{\mathrm{B}}|\mathbf{E}_{\mathrm{A}}(\mathbf{R})|^2. \tag{3.37}$$

$\mathbf{R}$, the vector between A and B. In analogy with the electric force

$$\mathbf{F} = -\nabla U, \tag{3.38}$$

the electric field depends on the electrostatic potential

$$\mathbf{E} = -\nabla\phi. \tag{3.39}$$

The electrostatic potential due to a neutral dipole at distance $\mathbf{R}$ is

$$\phi_{\mathrm{A}}(\mathbf{R}) = \frac{1}{4\pi\epsilon_0}\frac{\mathbf{p}_{\mathrm{A}}\mathbf{R}}{R^3} \tag{3.40}$$

which results in

$$\mathbf{E}_{\mathrm{A}}(\mathbf{R}) = -\frac{1}{4\pi\epsilon_0}\left\{\frac{\mathbf{p}_{\mathrm{A}}}{R^3} - 3\frac{\mathbf{p}_{\mathrm{A}}\mathbf{R}}{R^5}\mathbf{R}\right\}. \tag{3.41}$$

The modulus (absolute value) of this vector at the center of mass of molecule B is

$$E_{\mathrm{A}} = \frac{1}{4\pi\epsilon_0}\frac{p_A}{R^3}\sqrt{1+3\cos^2\theta} \tag{3.42}$$

with R the modulus of $\mathbf{R}$ and θ the angle between the dipole of A, p_{A}, and $\mathbf{R}$. The induction energy $U_{\mathrm{AB\ induced}}$ depends quadratically on the field and is therefore proportional to $-\frac{1}{R^6}$.

3.2.2 Electrostatic interactions in force fields

In force fields usually a set of effective point charges is used. Often the positions of these point charges coincide with the atom positions of the molecule. The electrostatic interactions are then calculated by applying Eq. (3.10). Different recipes for constructing such a set exist. When deciding which one to use, several things should be considered.

1. One should consider the environment in which the charges will be used, since this can induce a dipole and the charge distribution can be different. Also its ionization state might change, depending on the environment. Amino acids are for instance in their zwitterionic form in solution, but whereas they are in their "normal" form in the gas phase.

2. The charge distribution should reproduce the electric moments (dipole, quadrupole, etc.). A good test would be to compare the interaction energy at large distance by applying Eqs. (3.10) and 3.34.

3. If a modeling method with flexible molecules is applied, the charge set should not be too conformationally dependent.

4. Often the force field is fitted using a specific recipe for the point charges. Errors on the point charges can be partly canceled by other intermolecular interactions which will now have an error as well. Applying a different recipe of the point charges (even if this is a better one) can lead to worse overall results because of there will be no cancellation of errors.

Generally, the following recipes exist

Mulliken charges are one of the oldest and arise from Mulliken population analysis. Once an electronic structure calculation based on molecular orbitals is performed, it is a relatively cheap calculation. The charge is calculated by subtracting the sum of the electron density in all atomic and molecular orbitals in which the atom is involved from the charge of the nucleus. Here the overlap population between two atoms is divided equally between the two partners, which is a rather arbitrary choice. The result is that often bigger basis sets produce less believable Mulliken charges. The method is therefore not often used anymore, but the charges are still used as a reference.

Gasteiger, Qeq, EEM charges are all based on electronegativity. The way the electronegativity is obtained and how the charge set is calculated from the electronegativity (iteratively, using charge balance, etc.) changes between the different schemes. All are rather fast and do, in principle, not require expensive electronic structure calculations.

Parameterized in force field Some force fields have a parameterization for the point charges incorporated which leads to a more consistent force field as explained earlier.

Multipole fitting A set of charge sites is used to describe the (induced) dipole moment and/or higher order electric moments. This can be environment specific for instance liquid water versus water ice.

ElectroStatic Potential fit Electrostatic potential, produced by the *ab initio* electron density, can be calculated at a grid of points. Atomic charges, centered at the nuclei, which would produce the same set of potentials, are then found by least-squares fitting. All the electron density can be treated equally in this: the only arbitrary decision is the size of the grid. The grid is often chosen on shells of surfaces between 1 and 2 times the VDW radii. ESP charges take longer to calculate, but they do converge as basis set size is increased. Two known problems are that core atoms that are at center of the molecule can get unphysically large charges and that the results depend on conformation of the molecule: equivalent atoms (for instance methyl hydrogens) get different charges. A solution is the Restrained ElectroStatic Potential fit, which is a fitting algorithm using quadratic restraints that keep the absolute charge of core atoms low and shows less conformational dependence.

3.3 Intermolecular interactions

The dispersion energy was first identified by London in 1930 and is quantum mechanical in nature. The contributions described in the previous sections were all due to classical electromagnetism. The electrons in an atom or molecule are in continual motion, even in the ground state. So, although on average the dipole moment of a spherically symmetrical system is zero, at any instant a temporary dipole moment can occur. This temporary dipole can induce a further temporary dipole in a neighboring atom or molecule and, as in the case of the inductive interaction, the net effect will be attractive. According to Drude's theory the dependence will be

$$\langle U \rangle_{\text{disp}} = -\left(\frac{D_6}{R^6} + \frac{D_8}{R^8} + \frac{D_{10}}{R^{10}} + \cdots \right) \tag{3.43}$$

The R^{-6} term is due to instantaneous dipole-induced dipole, higher order terms due to instantaneous quadrupole-induced quadropole etc. The factor D_6 is given by

$$D_6 = -\frac{3\alpha^2 \epsilon_1}{4(4\pi\epsilon_0)^2} \tag{3.44}$$

with ϵ_1 the first excitation energy of the species concerned.

One could summarize the non-bonded terms by

$$U_{\text{non bonded}} = U_{\text{rep}} + U_{\text{Coulomb}} + U_{\text{induction}} + U_{\text{disp}}. \tag{3.45}$$

Both the induction and the dispersion energy have a R^{-6}-dependence and therefore

$$U_{\text{non bonded}} = U_{\text{rep}} + U_{\text{Coulomb}} - \frac{C}{R^6} \tag{3.46}$$

We have covered the Coulomb contribution and we are only left with the repulsive part U_{rep}. The repulsion is due to the interpenetration of the closed electron shells and should have an exponential form. This is the case for the Buckingham or exp-6 potential

$$U_{\text{rep}} = A\exp(-BR). \tag{3.47}$$

However, in combination with the R^{-6} term the energy goes to minus infinity for small R. Although a large barrier exists before this regime is reached, it can be problematic for some applications. In this case often a damping function on the dispersion term is applied. The exponential term is computationally rather expensive and therefore different expressions are used as well. They all have in common that they fall off quickly with distance. A famous example is the Lennard-Jones potential

$$U_{\text{rep}} = \frac{C'}{R^{12}} \tag{3.48}$$

or

$$U_{\text{L-J}} = 4\epsilon\left(\left(\frac{\sigma}{R}\right)^{12} - \left(\frac{\sigma}{R}\right)^{6}\right). \tag{3.49}$$

Since the repulsion is term is the square of dispersion term, this is relatively cheap to calculate. Moreover, it remains repulsive at short distance in combination with the R^{-6} term.

3.4 Intramolecular interactions

The intramolecular interactions are usually described by three different terms accounting for the two-body, three-body and four-body interactions. The first represents stretching of covalent bonds, the second bending of angles and the third torsional motion of dihedral angles and out-of-plane movement (umbrella modes). For intermolecular interactions, force fields give the energy with respect to the molecules at infinite distance. For intramolecular interactions, the energy with respect to the "ground state energy" is given. This means for instance the bond energy is given with respect to the bond energy at the equilibrium distance. For this reason, force fields can compare different *conformers* and not *isomers*, since the energy in the bond itself is not included.

3.4.1 Stretching

An intuitive description of the stretching of bond is by using Hooke's law for the force on a spring with spring constant k_{s}

$$F_{\text{s}} = -k_{\text{s}}(l - l_{\text{eq}}) \tag{3.50}$$

where l_{eq} is the equilibrium length of the spring. The energy of this system is then

$$U(l) = -\int F(x)\mathrm{d}x = U(l_{\text{eq}}) + \frac{1}{2}k_{\text{s}}(l - l_{\text{eq}})^2 \tag{3.51}$$

As explained earlier, the energy reference is usually chosen such that $U(l_{\text{eq}}) = 0$ and $k_{\text{stretch}} = \frac{1}{2}k_{\text{s}}$

$$U_{\text{stretch}} = -k_{\text{stretch}}(l - l_{\text{eq}})^2 \tag{3.52}$$

Let us consider a diatomic molecule with a bond with length R along direction x, such that

$$R = x_2 - x_1. \tag{3.53}$$

According to Newton's second law the force on the first atom is

$$F_1 = m_1 \frac{\mathrm{d}^2 x_1}{\mathrm{d}t^2} = -k_\mathrm{s}(x_2 - x_1 - R_e) \tag{3.54}$$

with R_e the equilibrium distance. Similarly, the second atom

$$\frac{F_2}{m_2} = \frac{\mathrm{d}^2 x_2}{\mathrm{d}t^2} = \frac{k_\mathrm{s}}{m_2}(x_2 - x_1 - R_e), \tag{3.55}$$

which combines to

$$\frac{\mathrm{d}^2 R}{\mathrm{d}t^2} = \frac{\mathrm{d}^2 x_2}{\mathrm{d}t^2} - \frac{\mathrm{d}^2 x_1}{\mathrm{d}t^2} = k_\mathrm{s}\left(\frac{1}{m_1} + \frac{1}{m_2}\right)(R - R_e) = \frac{k_\mathrm{s}}{\mu}(R - R_e) \tag{3.56}$$

with

$$\frac{1}{\mu} = \left(\frac{1}{m_1} + \frac{1}{m_2}\right) \tag{3.57}$$

where μ is the reduced mass. What we have done in this example is to convert the Cartesian coordinates (x_1 and x_2) to internal coordinates (R), which is linear combination of the Cartesian coordinates. To this new coordinate belongs a reduced mass, instead of the original particle masses in the Cartesian system. Different types of internal coordinates (bonds, angles, dihedrals, etc) will have different functional forms to express the reduced mass. Often a mass-weighted coordinate system is used with

$$\mathbf{q}_i = \sqrt{m_i}\mathbf{r}_i \tag{3.58}$$

and mass-weighted forces which automatically takes into account the reduced masses and makes it easier to perform a normal mode analysis (Eq. (2.3)). The relationship between the k values and the frequencies now become

$$\nu = \frac{1}{2\pi}\sqrt{k}. \tag{3.59}$$

instead of

$$\nu = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}}. \tag{3.60}$$

A better representation for a bond is the Morse potential

$$U = D_\mathrm{e}\left(1 - \exp(-\beta(R - R_\mathrm{e}))\right)^2 \tag{3.61}$$

where D_e is the depth of the potential well and $\beta = \omega_\mathrm{e}\sqrt{\frac{\mu}{2D_\mathrm{e}}}$. This potential is however computationally more demanding and is therefore more often used for small systems than for larger molecules. Moreover, the harmonic potential and the Morse potential coincide for small deviations from the equilibrium distance.

3.4.2 Bending

For the three-body bending potential general forms are

$$U_\mathrm{ABC} = \frac{1}{2}k_\mathrm{ABC}\left(\theta_\mathrm{ABC} - \theta_\mathrm{e,ABC}\right)^2 \tag{3.62}$$

and

$$U_\mathrm{ABC} = \frac{k_\mathrm{ABC}}{2\sin^2\theta_\mathrm{e,ABC}}\left(\cos\theta_\mathrm{ABC} - \cos\theta_\mathrm{e,ABC}\right)^2, \tag{3.63}$$

or a combination with stretching

$$\frac{k}{2}(R - R_\mathrm{e})(\theta - \theta_\mathrm{e}). \tag{3.64}$$

3.4.3 Four-body terms

Consider four atoms sequentially bonded: A–B–C–D. The dihedral angle or torsional angle of the system is the angle between the A–B bond and the C–D bond as viewed along the B–C bond. Since the geometry repeats itself every 360°, the energy varies with the dihedral angle in a sine or cosine pattern. For systems A–B–C–D of lower symmetry, like butane, the torsional potential energy curve is more complicated, but a combination of sine or cosine functions will reproduce the curve: A general form for torsional motion is therefore

$$U = \frac{U_0}{2}(1 - \cos(n(\chi - \chi_e))) \tag{3.65}$$

with n a periodicity parameter. For ethane for instance the periodicity parameter will be three.

The contribution to the potential energy due out-of-plane/inversion/umbrella moment can either be expressed in terms of angle (ψ)

$$U = \frac{1}{2\sin^2\psi_{\mathrm{e}}}\left(\cos\psi - \cos\psi_{\mathrm{e}}\right)^2, \tag{3.66}$$

$$U = k_1\left(1 + k_2\cos(n\psi)\right) \quad \text{with } n \text{ a periodic parameter,} \tag{3.67}$$

or

$$U = k_1(1 + \cos(n\psi - k_2)) \tag{3.68}$$

or in terms of out-of-plane height ($h \propto \sin\psi$)

$$U = kh^2. \tag{3.69}$$

3.5 How to parameterize a force field

A force field consists of a description of the functional forms that are used which were discussed in the previous section together with a list of force constants. Force constants are required for all different bonds, angles etc. Often not only different elements are distinguished but also different atom types. Carbon for instance is often distinguished in four types: sp, sp^2 or sp^3 hybridization or aromatic carbon. Obviously the equilibrium distance and force constant change depending on the type of bond. Non-bonded terms discussed so far were mostly between two molecules, however they are often used as site-site potential in which all atom-atom combinations between two molecules are described by a non-bonded potential. Since the number of parameters becomes huge in this way, often combination rules are applied to obtain parameters for interaction pairs of different atom types for the interaction between atoms of the same type. In this way parameters for the O–C interaction can be obtained from the parameters of the C–C and O–O interactions. Typical combination rules are

$$C_{6,ij} = 2\left(\frac{R_i^*}{2} + \frac{R_j^*}{2}\right)^6\sqrt{\epsilon_i\epsilon_j} \tag{3.70}$$

and

$$C_{12,ij} = \left(\frac{R_i^*}{2} + \frac{R_j^*}{2}\right)^{12}\sqrt{\epsilon_i\epsilon_j} \tag{3.71}$$

with R_i^* is the minimum energy separation for two atoms of type i and ϵ_i the well depth;

$$C_{6,ij} = 4\left(\sigma_i\sigma_j\right)^3\sqrt{\epsilon_i\epsilon_j} \tag{3.72}$$

$$C_{12,ij} = 4\left(\sigma_i\sigma_j\right)^6\sqrt{\epsilon_i\epsilon_j} \tag{3.73}$$

and

$$C_{6,ij} = C\frac{\alpha_i \alpha_j}{\sqrt{\frac{\alpha_i}{N_i} + \frac{\alpha_j}{N_j}}} \tag{3.74}$$

$$C_{12,ij} = \frac{1}{2} C_{6,ij} \left(R_i + R_j\right)^6 \tag{3.75}$$

where α_i is the dipole polarizability of atom i, N_i the number of valence electrons and R_i the van der Waals radius.

"All-atom" force fields provide parameters for every type of atom in a system, including hydrogen, while "united-atom" force fields treat the hydrogen and carbon atoms in methyl and methylene groups as a single interaction center. "Coarse-grained" force fields, which are frequently used in long-time simulations of proteins, provide even more crude representations for increased computational efficiency.

There are different ways to obtain the force constants. One can optimize all parameters with respect to a set of experimental macroscopic test data, producing a "consistent" force field. The test set can be chosen to represent the system under investigation. Macroscopic quantities that can be used are for instance melting points, density, and heat capacity. The outcome is often good for a particular type of systems, or a particular property. However, extension to new systems can be difficult, since there is no direct link to 'physical reality' because the force constants have no physical origin. One can therefore not simply change or add individual force constants. On the other hand, this approach can account for bulk effects like an induced dipole in an effective way. The parameters are often determined using a specific set of charges and one should be careful to use the same point charge prescription.

Another approach is to mimic physical properties of individual elements or atom types, producing a "physical" force field. These properties can be taken from experimental data, or *ab-initio* calculations. The outcome will be 'reasonable' but never very good for a large range of different compounds and extension to new systems is relatively straightforward.

A third approach is to optimize all parameters with respect to energies obtained by *ab inito* calculations for a specific system. This gives very good results and when done carefully the parameters can have some physical meaning. The results are however hard to extend to new systems and constructing such a force field is computationally very demanding and will also require some human input.

Chapter 4

Geometry Optimization

A good starting point for each modeling effort is to do a geometry optimization, which is a search for the configuration that belongs to the local minimum on the potential energy surface, *i.e.*, finding a stationary point on the PES:

$$\frac{\partial E}{\partial q_1} = \frac{\partial E}{\partial q_2} = \cdots = \frac{\partial E}{\partial q_n} = 0. \tag{4.1}$$

A geometry optimization is a good first test of the accuracy of the force field that you are using. The optimized geometry should resemble the experimental structure which is for instance determined by X-ray crystallography or the structure that is obtained by theoretical methods of higher accuracy, which are computationally more expensive. Moreover, several modeling methods require to start on a potential energy minimum. For instance the calculation of vibrations and their frequencies by solving Eq. (2.3) only results in positive frequencies if the system is in a minimum. But also with Molecular Dynamics (Chapter 8) one can run into problems if the system is too high on the PES, which introduces an artificial amount of energy which allows the systems to cross barriers that would normally not be accessible. Kinetic Monte Carlo (Chapter 10) simulates the evolution of the system from one local minimum to the next. So also here it is required to first search for this minimum. In general, one can state that calculations of properties using a structure that is not in a minimum according to the potential energy surface that is used to calculate that properties can lead to unreliable results. One can therefore not do a geometry optimization with a high level of theory and then use a more approximate method to calculate properties.

This chapter will discuss several optimization methods. Since we search for a stationary point, it would be obvious to use the first and better even the second derivative of the energy. When using force fields this could in principle be done analytically, since a functional form of the energy is known. However, the functional form is not always known, in the case of electronic structure calculations for instance. Moreover, taking the analytical derivative of dihedrals is far from straightforward and one often wants to have a generic routine to optimize the structure which can be applied in combination with any force field. In those cases, the first derivative has to be determined numerically, which is computationally more demanding and challenging, especially for rough PESs.

Several geometry optimization routines have been proposed and one can generally distinguish three types of approaches

- methods that use no derivatives (only function values, such as the energy)
- methods that use first derivatives (slope, or force)
- methods that use second derivatives (curvature, or Hessian)

4.1 Derivative-free methods

In general, methods that use no derivatives spend the least amount of time at each point but require the most steps to reach the minimum. Methods such as the simplex minimization, simulated annealing, Markov-chain Monte Carlo and optimization by genetic algorithms fall into this category. Two methods will be discussed here. Markov-chain Monte Carlo will be introduced later in this course (Chapter 9).

4.1.1 Multivariate Grid Search

This is a very simple and intuitive method. It will require many steps to reach the minimum at high precision and the system will easily get stuck in a local minimum. It is a good method to use as a first optimization algorithm to get a crude estimate. More expensive methods per iteration can then be used to refine the structure. The algorithm uses the following steps:

1. Choose a suitable grid for the variables.

2. Choose a starting point $\mathbf{A}$ on the grid.

3. For each variable $q_1, q_2, ..., q_p$ evaluate the molecular potential energy U at the two points surrounding $\mathbf{A}$ (as determined by the grid size).

4. Select the new point for which U is a minimum, and repeat steps 3 and 4 until the local minimum is identified.

4.1.2 Simulated Annealing

Simulated annealing is an iterative algorithm. It is a special case of a Monte Carlo simulation which will be discussed in Chapter 9. A trial move is made by randomly changing some coordinates of the atoms/molecules. This trial move is excepted according to

$$\exp(-\Delta U/kT) > R(0,1) \tag{4.2}$$

with ΔU the difference in energy between the trial move and the current configuration, T a "synthetic temperature", k the Boltzmann constant and $R(0,1)$ a random value in the interval [0,1]. The "synthetic temperature" decreases during the optimization. This process mimics the process undergone by misplaced atoms in a metal when its heated and then slowly cooled. The cooling scheme and cooling rate have an important influence on the outcome and should be tailored for each application. Adaptive simulated annealing algorithms exists that address this problem by connecting the cooling schedule to the search progress. The algorithm is good in finding a local minimum for a system with many free parameters and a rather rough energy landscape with many very shallow minima. In the initial part it can efficiently move out of a bad geometry. When lowering the temperature, it gradually moves to the local minimum.

4.2 First-order methods

Methods that use only the first derivatives are sometimes used in computational chemistry, especially for preliminary minimization of very large systems. The first derivative tells the downhill direction and also suggests the step size when stepping down the hill. Large steps on a steep slope; small steps on flatter areas that are hopefully near the minimum. Methods such as steepest descent and a variety of conjugate gradient minimization algorithms belong to this group.

4.2.1 Steepest descent

This first-order derivative scheme for locating minima on molecular potential energy surfaces can be summarized as follows;

1. Calculate U for the initial structure.

2. Calculate U for structures where each atom is moved along the x, y and z-axes by a small increment. Movement of some atoms will lead to a small change in U, whilst movements of other atoms will lead to a large change in U. (The important quantity is clearly the gradient.)

3. Move the atoms to new positions such that the energy U decreases by the maximum possible amount ($\mathbf{q}^{(k)} = \mathbf{q}^{(k-1)} - \Delta q \mathbf{g}^{(k-1)}$). k indicates the iteration number and $\mathbf{g}$ is the gradient vector.

4. Repeat the relevant steps above until a local minimum is found.

The step length Δq is an input parameter. This can be set a fixed value or at each step one can determine the optimum value by performing a line search along the step direction to find the minimum function value along this line. The step direction is always orthogonal to the previous step direction, which makes the stepping algorithm rather inefficient.

4.2.2 Conjugate gradients

For conjugate gradients the step direction is not always orthogonal to the previous direction which results in a faster algorithm than steepest decent. Starting from point $q^{(k)}$, where k is the iteration count, we move in a direction given by the vector

$$\mathbf{q}^{(k+1)} = \mathbf{q}^{(k)} + \Delta q \mathbf{V}^{(k)} \tag{4.3}$$

where

$$\mathbf{V}^{(k)} = \mathbf{g}^{(k)} + \gamma^{(k)} \mathbf{V}^{(k-1)} \tag{4.4}$$

and $\mathbf{g}^{(k)}$ is the gradient vector at point $\mathbf{q}^{(k)}$ and $\gamma^{(k)}$ is a scalar given by

$$\gamma^{(k)} = \frac{(\mathbf{g}^{(k)})^{\mathrm{T}} \mathbf{g}^{(k)}}{(\mathbf{g}^{(k-1)})^{\mathrm{T}} \mathbf{g}^{(k-1)}} \tag{4.5}$$

or

$$\gamma^{(k)} = \frac{(\mathbf{g}^{(k)} - \mathbf{g}^{(k-1)})^{\mathrm{T}} \mathbf{g}^{(k)}}{(\mathbf{g}^{(k-1)})^{\mathrm{T}} \mathbf{g}^{(k-1)}} \tag{4.6}$$

T denotes the transpose of a matrix. Which expression for γ is superior depends on the functional form of the surface one wants to optimize. Again a line search can be applied to determine Δq.

4.3 Second-order methods

Methods that use both the first derivative and the second derivative can reach the minimum in the least number of steps because the curvature gives an estimation of where the minimum is. The simplest method in this category is a method based on the Newton-Raphson method for roots finding.

4.4 Newton-Raphson

This method finds the positions where the function gives result 0. For a function $f(x)$ with gradient $g(x)$, this results in the following algorithm:

1. Start in $x^{(1)}$ and determine $f^{(1)}$ and $g^{(1)}$.
2. The function can be linearly approximated by $y = g^{(1)}x + f^{(1)} - g^{(1)}x^{(1)}$
3. This has a root at $x^{(2)} = x^{(1)} - \frac{f^{(1)}}{g^{(1)}}$
4. Repeat steps 1-3

The starting point is crucial since it determines the root which is eventually found.

We can now apply this method to find stationary points by searching for roots of the gradient. We use a linear function for the approximation of the gradient which results in a quadratic function for the approximation of the function. The values for g and x in the next iteration can be calculated using

$$g^{(k)}(x) \approx H^{(k)}(x - x^{(k)}) + g^{(k)} = 0 \tag{4.7}$$

and

$$x^{(k+1)} = x^{(k)} - \frac{g^{(k)}}{H^{(k)}} \tag{4.8}$$

with $H^{(k)}$ the Hessian, second derivative, at the current point $x^{(1)}$. In the case of multiple variables:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - (\mathbf{H}^{(k)})^{-1}\mathbf{g}^{(k)}. \tag{4.9}$$

4.5 Numerical Hessian

For geometry optimization of a single molecule virtually all software packages use a method that needs a Hessian and/or its inverse. If the functional form of the PES is known, for instance in the case of force field where only the force constants change, one could derive an analytical expression for the gradient and the Hessian. However, in most cases the functional form is not known or too system specific to use in a general modeling program, and one has to determine the gradient and Hessian numerically.

Analytically differentiation is easier than integration; numerically it is the reverse. It is hard to get accurate values for the gradients and minimizers need accurate gradients to converge correctly. One can imagine that calculating the Hessian by taking the numerical derivative twice leads to an even worse approximation for the Hessian. Fortunately most minimization methods do not need precise Hessians and often approximates are applied. Quasi-Newton methods start with an approximate (inverse) Hessian and update this every iteration. There are several routines for this. One is the Davidon-Fletcher-Powell formula which updates the inverse Hessian $B_k = H_k^{-1}$ in the following way:

$$B_{k+1} = B_k - \frac{B_k y_k y_k^T B_k}{y_k^T B_k y_k} + \frac{\Delta x_k (\Delta x_k)^T}{y_k^T \Delta x_k} \tag{4.10}$$

with

$$y^{(k)} = g^{(k+1)} - g^{(k)} \tag{4.11}$$

with k the iteration step. The initial estimated Hessian can be the unit matrix or one can use internal coordinates. The individual diagonal elements of the Hessian can be identified as bond-stretching and bond-bending force constants, etc.

Chapter 5

Reaction path finding methods

It can be interesting to determine the reaction path connecting one potential energy minimum to another and the determine the (free) energy difference between the wells and the highest point on this path. Several methods have been developed to determine this minimum energy path (MEP). Exactly which one to take depends on the type of energy landscape (smooth, rugged or rough) and whether one knows both the initial and the final state or only the initial state.

5.1 Double-ended methods

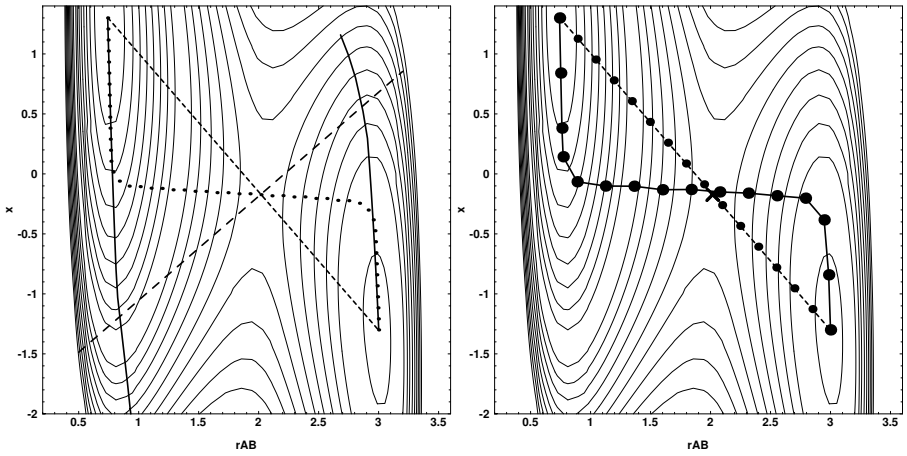
One can image that in the case of "double ended" searches — both the initial and final state are known — on a smooth surface, one simply climbs uphill along the reaction coordinate and minimize in all other directions to find the saddle point and the reaction path. This is the basis of the "Drag method". This is a simple method with many different names since it gets regularly "rediscovered". Figure 5.1a gives a schematic representation of this method. First the two minima are connected (short-dash line). Next the system is dragged along this line from one end to other by iteratively increasing the coordinate along this line and minimizing in all other degrees of freedom (long-dash line direction). In this way the system climbs the potential near the steepest ascent path (solid lines). In this case it not only misses the saddle point but even the other state. Dragging from the top left to bottom right state gives different result from the opposite drag. So there is clearly a strong hysteresis effect.

5.2 Nudged Elastic Band method

A solution to this can be the use of a chain-of-states method. An example of such a method is the Nudged Elastic Band method. The product and reactant state are connected by a string of replicas (or images) that are connected through springs. This string can be denoted by $[\mathbf{R}_0, \mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_N]$ where the endpoints are fixed and given by the initial and final states, $\mathbf{R}_0 = \mathbf{R}$ and $\mathbf{R}_N = \mathbf{P}$, but $N-1$ intermediate images are adjusted by an optimization algorithm which can be one of the methods discussed in Chapter 4. The most straightforward approach would be to minimize

$$\sum_{i=1}^{N-1} E(\mathbf{R}_i) + \sum_{i=1}^{N} \frac{k}{2}(\mathbf{R}_i - \mathbf{R}_{i-1})^2 \tag{5.1}$$

with respect to the intermediate images, $[\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{N-1}]$. The spring constant keeps the string "elastic". This gives however corner-cutting and down-sliding problems. Corner-cutting leads to irregular minimum energy paths. The down-sliding problem refers to the fact that images move downhill along the minimum energy path, away from the saddle point. This is leads



(a) (b)

Figure 5.1: *Schematic of the (a) drag and (b) NEB method for reaction path finding. Figures taken from [14].*

to a lowest resolution around the area of interest. This can be solved using:

$$\mathbf{F}_i^{\text{nudged}} = \mathbf{F}(\mathbf{R}_i)|_\perp + \mathbf{F}_i^s \cdot \hat{\boldsymbol{\tau}}_\parallel \hat{\boldsymbol{\tau}}_\parallel \tag{5.2}$$

with the spring force

$$\mathbf{F}_i^s \equiv k_{i+1}(\mathbf{R}_{i+1} - \mathbf{R}_i) - k_i(\mathbf{R}_i - \mathbf{R}_{i-1}) \tag{5.3}$$

and the force perpendicular to the elastic band

$$\mathbf{F}(\mathbf{R}_i)|_\perp = -\nabla V(\mathbf{R}_i) + \nabla V(\mathbf{R}_i) \cdot \hat{\boldsymbol{\tau}}_\parallel \hat{\boldsymbol{\tau}}_\parallel \tag{5.4}$$

where $\hat{\boldsymbol{\tau}}_\parallel$ is the tangent parallel the elastic band. Now, the force along the spring is solely determined by the spring constants and the force in the perpendicular directions by the PES. Using a geometry minimization method that uses at least first derivatives, the optimum position of band and therefore the MEP can be determined. This is shown in Figure 5.1b. One starts with images along the dashed line and the optimization in the first step occurs along the same coordinate as in the drag method, *i.e.*, perpendicular to the dashed line. In the next iterations it is however adjusted such that is always perpendicular to the spring. The images converge on the MEP with equal spacing if the spring constant is the same for all the springs. Typically none of the images lands at or even near the saddle point and the saddle point energy needs to be estimated by interpolation. An example of MEP from a NEB calculation is shown in Fig. 5.2. The MEP for dissociative adsorption of CH_4 on an Ir(111) surface has a narrow barrier compared with the length of the MEP. The NEB calculation completely misses this barrier. One can avoid this by using Climbing Image Nudged Elastic Band. The force for the highest energy image is now changed. The spring force of this image is zero and the parallel force is inverted

$$\mathbf{F}_{\text{max}}^{\text{nudged}} = -\nabla V(\mathbf{R}_{\text{max}}) + 2\nabla V(\mathbf{R}_{\text{max}}) \cdot \hat{\boldsymbol{\tau}}_\parallel \hat{\boldsymbol{\tau}}_\parallel. \tag{5.5}$$

This results in one image at the maximum of the MEP. This is shown again in Figure 5.2.

5.3 Transition Path Sampling

One can image that while the methods described above work well for smooth PESs or even for rugged potential surfaces, they do not for rough surfaces. In these cases, one cannot simply look

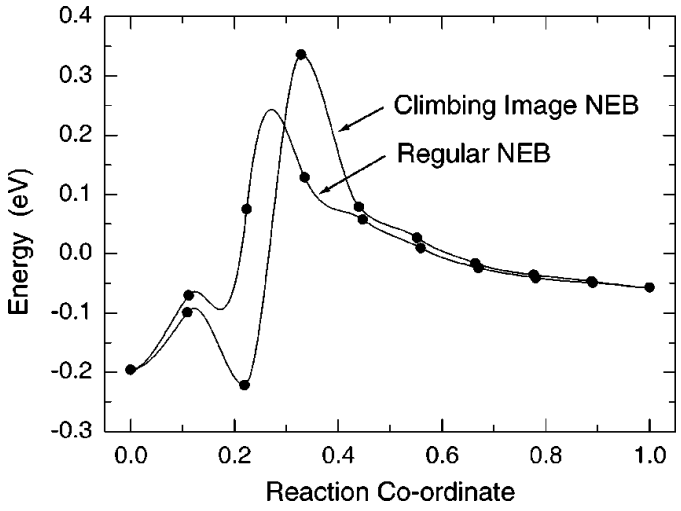


Figure 5.2: *Density functional theory calculations of the minimum energy path for CH_4 dissociative adsorption on a Ir(111) surface. A regular NEB calculation and a climbing image NEB calculation are compared, both involving 8 movable images. The regular NEB results in a low resolution of the barrier, and the interpolation gives an underestimate of the activation energy. The climbing image NEB brings one of the images right to the saddle point and gives the activation energy precisely with insignificant additional computational effort. Taken from [11].*

for the saddle point connecting the two wells, since there are many saddle points in between. An approach would be to use Transition Path Sampling (TPS). A good overview of the method can be found in [1]. Instead of determining one MEP, an ensemble of transition paths is created.

Starting from a certain state, one can perform dynamics simulations (see Chapter 8) which result in trajectories. These trajectories indicate the evolution of the system in time. Consider a system with two stable states A and B, then generally most trajectories starting in A will end in A as well. We are however usually only interested in trajectories that cross the barrier to B. TPS focuses on the most interesting part of the simulation, namely the trajectories resulting in crossings. An example could be an initially unfolded protein undergoing a transition to fold on itself. The aim of the TPS method is to reproduce precisely those folding moments. For such a complex systems there are generally many ways in which the transition can take place. Given an initial path, TPS provides some algorithms to perturb that path and create a new one. A powerful and efficient algorithm is the so-called shooting move [3]. Consider the case of a classical many-body system described by coordinates r and momenta p. Molecular dynamics generates a path as a set of (r_t, p_t) at discrete times t in $[0, T]$ where T is the length of the path. For a transition from A to B, (r_0, p_0) is in A, and (r_T, p_T) is in B. One of the path times is chosen at random, the momenta p are modified slightly into $p + \delta p$, where δp is a random perturbation consistent with system constraints, *e.g.*, conservation of energy and linear and angular momentum. A new trajectory is then simulated from this point, both backward and forward in time until one of the states is reached. Being in a transition region, this will not take long. If the new path still connects A to B it is accepted, otherwise it is rejected and the procedure starts again. This scheme is again a Monte Carlo scheme like in the case of simulated annealing. More detail about the Monte Carlo method and the theory behind this method is given in Chapter 8.

5.4 Single-ended methods

In some cases double-ended methods like NEB cannot be applied to find transition states, since they require knowledge of both the initial and final states. However, the final state is not always known. An example is in kinetic Monte Carlo simulations which will be discussed in Chapter 10. It can also be that one would like to explore the PES in the vicinity in the current state to have an unbiased way to explore the possible ways for the system to evolve. Several single-ended methods exist [24, 12], but the minimum-mode-following method [9] is the most frequently used. This method can be used when the transition state is assumed to be a single point on the potential energy surface (a first order saddle point). Olsen et al. have performed a comparative study of different methods for finding saddle points without knowledge of the final states[24]. They conclude that minimum-mode following methods are to be preferred if a more complete mapping of all existing (low-energy) saddle points is desired for large systems. The method starts by taking the system out of its potential energy minimum by making a small displacement. One can then move the system in the direction of the lowest-eigenvalue eigenvector of the Hessian which can be calculated exactly or approximated by a method such as the Dimer method [9], Lagrange Multipliers [21] or Lanczos method [16]. This moves the system uphill in the direction of the saddle point. By iterating the process the one can reach the saddle point when all eigenvalues but one are positive and the force on the system is low. From the saddle point one can make a displacement in the direction of the initial state and preform a geometry optimization to see whether the initial state is reached again. A displacement in the opposite direction followed by a geometry optimization can be made to obtain the final state. One needs to repeat this many times with different initial displacements to obtain good coverage of all close-lying saddle points.

Chapter 6

Statistical Thermodynamics and Transition State Theory

Geometry optimization results in the optimum structure at $T = 0$ K. Properties that are calculated on the basis of this structure are a good first starting point, but since we work experimentally mostly at $T > 0$ K, they can be quite far off. At $T > 0$ K the system may not be in its most optimum structure since also other states are accessible, which can have different properties leading to a different average macroscopic (measurable) property. The determination of this average property is the area of statistical mechanics. Statistical mechanics predicts the probability that a molecule is in a certain energy state (ground state vs. excited state).

The probability that a molecule is in a state i is

$$P_i = \frac{\exp(-\epsilon_i/kT)}{\sum_j \exp(-\epsilon_j/kT)} \tag{6.1}$$

with discrete energy levels and the molecular partition function

$$q = \sum_j \exp(-\epsilon_j/kT) \tag{6.2}$$

where T is the temperature and k is Boltzmann's constant. If $\epsilon_0 = 0$ then the partition function gives the number of accessible energy levels. In molecular mechanics one usually has a collection of interacting molecules and the mutual potential energy of the system is considered instead of the internal energy of the molecule. We now move from the microcanonical to the canonical ensemble with

$$P_i = \frac{\exp(-U_i/kT)}{\sum_j \exp(-U_j/kT)} \tag{6.3}$$

and

$$Q = \sum_j \exp(-U_j/kT). \tag{6.4}$$

To calculate a property of a system:

$$\langle A \rangle = \frac{\sum A_i \exp(-U_i/kT)}{Q} \tag{6.5}$$

So to get the average mutual potential energy

$$\langle U \rangle = \frac{\sum U_i \exp(-U_i/kT)}{Q} \tag{6.6}$$

We can achieve this by considering a large number of cells with the same thermodynamic properties (N, V, T, E, p and/or μ etc.) but different molecular arrangements. If N, V and T is kept constant between cells we speak about the *canonical ensemble*

$$P_i = \left(\frac{\exp(-U_i/kT)}{Q} \right)_{N,V,T} \tag{6.7}$$

with Q the canonical partition function. The *microcanonical ensemble* has constant N, E, and V, which means that there is no energy flow between the cells. This would apply to an isolated molecule. In the *isothermal-isobaric ensemble* N, p, and T are kept constant, which is a more common equilibrium situation. In the *grand canonical ensemble* V and T, and the total chemical potential μ remain constant. This can be used to describe a system with two phases with different chemical potential like solid and melt.

In Molecular Mechanics, the energy is a function of the coordinates of the atoms, our force field, and no discrete energy levels are used. Parameter space is defined by the (internal) coordinates of the systems and all points in parameter space contribute to the partition function. For this reason often an integration over parameter space is used instead of a sum over the discrete energy levels:

$$Q = \frac{1}{N!} \int \exp\left(-\frac{U(\mathbf{r}^N)}{kt} \right) \mathrm{d}\mathbf{r}^N \tag{6.8}$$

The factor $1/N!$ is for indistinguishable particles; for particles that can be distinguished this term drops from the equation. The average mutual potential energy can be determined by solving:

$$\langle U \rangle = \frac{\int U(\mathbf{r}^N) \exp\left(-\frac{U(\mathbf{r}^N)}{kT} \right) \mathrm{d}\mathbf{r}^N}{Q} \tag{6.9}$$

6.1 The average mutual potential energy of two dipoles

In Section 3.2 the angle dependent dipole-dipole mutual potential energy was determined

$$(4\pi\epsilon_0) U_{\mathrm{AB}} = -\frac{p_A p_B}{R^3} \left(2\cos\theta_A \cos\theta_B - \sin\theta_A \sin\theta_B \cos\phi \right) \tag{6.10}$$

One could employ Eq. (6.9) to obtain the average potential energy assuming a parallel configuration. In this case, the angles have the constraints $\theta_A = \theta_B$ and $\phi = 0$ and the potential energy becomes

$$U_{\mathrm{AB}} = \frac{p_A p_B}{4\pi\epsilon_0 R^3} \left(1 - 3\cos^2\theta \right). \tag{6.11}$$

Combining this with Eq. (6.9) leads to the average potential energy due to thermal fluctuations of θ

$$\begin{aligned} \langle U_{\mathrm{AB}} \rangle_{\mathrm{dip...dip}} &= \frac{\int U_{\mathrm{AB}} \exp(-U_{\mathrm{AB}}/kT) \mathrm{d}\tau}{\int \exp(-U_{\mathrm{AB}}/kT) \mathrm{d}\tau} \\ &= \frac{\int_\phi \mathrm{d}\phi \int_{\theta=0}^{\pi} U_{\mathrm{AB}} \exp(-U_{\mathrm{AB}}/kT) \sin\theta \mathrm{d}\theta}{\int_\phi \mathrm{d}\phi \int_{\theta=0}^{\pi} \exp(-U_{\mathrm{AB}}/kT) \sin\theta \mathrm{d}\theta}. \end{aligned} \tag{6.12}$$

If we assume that U_{AB} is smaller than kT, we can make the approximation

$$\exp(-U_{\mathrm{AB}}/kT) \approx 1 - U_{\mathrm{AB}}/kT. \tag{6.13}$$

The average energy now becomes

$$\langle U_{\mathrm{AB}} \rangle_{\mathrm{dip...dip}} \approx \frac{\int_{\theta=0}^{\pi} C \left(1 - 3\cos^2\theta \right) \left(1 - C \left(1 - 3\cos^2\theta \right) / kT \right) \sin\theta \mathrm{d}\theta}{\int_{\theta=0}^{\pi} (1 - C \left(1 - 3\cos^2\theta \right) / kT) \sin\theta \mathrm{d}\theta} \tag{6.14}$$

with $C = \frac{p_A p_B}{4\pi\epsilon_0 R^3}$. Substitution of $x = \cos\theta$, which implies that $\mathrm{d}x = -\sin\theta \mathrm{d}\theta$, leads to

$$\langle U_{\mathrm{AB}} \rangle_{\mathrm{dip...dip}} \approx \frac{\int_{x=-1}^{1} C\left(1-3x^2\right)\left(1-C\left(1-3x^2\right)/kT\right)\mathrm{d}x}{\int_{x=-1}^{1}\left(1-C\left(1-3x^2\right)/kT\right)\mathrm{d}x} = -\frac{4C^2}{5kT}$$
$$= \frac{-4p_A^2 p_B^2}{5kT(4\pi\epsilon_0)^2}\frac{1}{R^6}. \quad (6.15)$$

If not only θ but all angles are independently taken into account, this results in

$$\langle U_{\mathrm{AB}} \rangle_{\mathrm{dip...dip}} = \frac{-2p_A^2 p_B^2}{3kT(4\pi\epsilon_0)^2}\frac{1}{R^6} \quad (6.16)$$

which is called the Keesom interaction energy. The expression shows that the energy of a Keesom interaction, *i.e.*, the interaction between two dipoles which are constantly rotating and never get locked into place, depends on the inverse sixth power of the distance, unlike the interaction energy of two spatially fixed dipoles, which depends on the inverse third power of the distance as we have discussed earlier.

For the dipole-dipole mutual potential energy, the integrals in

$$\langle U \rangle = \frac{\int U(\mathbf{r}^N)\exp\left(-\frac{U(\mathbf{r}^N)}{kT}\right)\mathrm{d}\mathbf{r}^N}{Q} \quad (6.17)$$

could be solved analytically using some approximations. This is however often not the case and the integrals need to be solved numerically. This problem together with the arrival of the first computers lead to the development of the Monte Carlo technique, which will be discussed in the Chapter 9.

6.2 Transition State Theory

In Chapter 5 we discussed ways to determine a reaction path bringing the system from state A to state B. This section will discuss a way to extract a rate from this information: (Classical) Transition State Theory (TST). This theory can be derived using statistical mechanics. It is also known as "Absolute Rate Theory" and "Activated Complex Theory". It was developed by Pelzer & Wigner [26], Eyring [5], and Evans & Polyani [4]. The basic idea is that the transition state theory rate is the product of the probability of being in the transition state and the flux out of the transition state

$$k^{\mathrm{TST}} = (\text{probability of being in TS}) \times (\text{flux out of TS}). \quad (6.18)$$

The assumption here is off course that the process of interest can be described on a single PES, which might not always be the case for chemical reactions. TST holds quite well for non-reactive processes, *e.g.*, conformational changes or diffusion of species, where the vibration and event time scale are decoupled. In this case, we can assume we have a Markov chain and we can apply statistical mechanics. A Markov chain assumes that probability to go to the next state only depends on current state and not on a previous state. This will be further discussed in Chapter 10.

Figure 6.1a draws again a schematic representation of a 2D PES with two minima, A and B, and the MEP connecting the two state in red. We can now identify an one-dimensional dividing surface which contains the Transition State, that represents a bottle neck for going from an initial to a final state. This bottle neck can be due to an energy barrier and/or entropy. This dividing surface is indicated by the dashed line in Figure 6.1a.

To arrive at the TST rate we need to get the probability to be at the dividing surface times the flux to leave the dividing surface. A dividing surface has dimensionality $3N-1$ instead of $3N$.

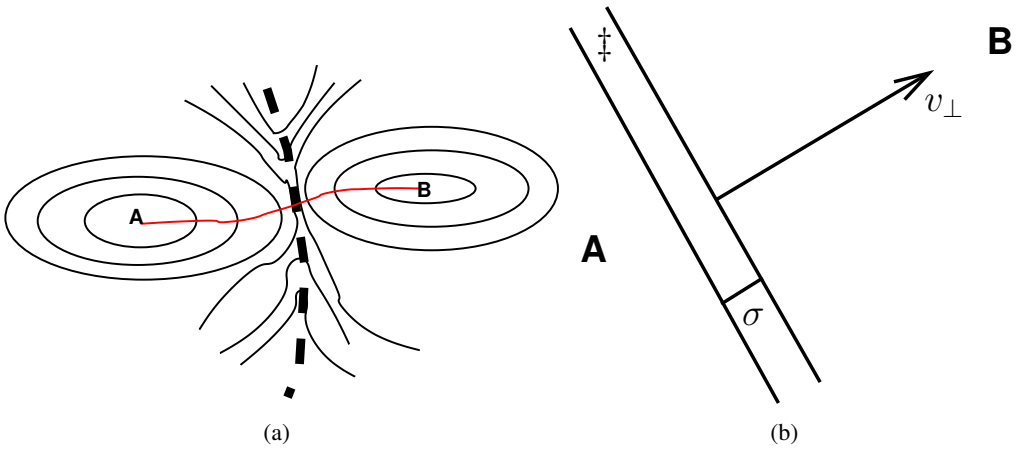


Figure 6.1: *(a) Schematic representation of a 2D PES with two minima A and B connected by a minimum energy path in red and separated by a dividing surface (dashed line). (b) Dividing surface of width σ and the perpendicular velocity bringing the system from A to B, $v_\perp$.*

We define a width σ along the reaction coordinate to again obtain a subspace of dimensionality $3N$ (see Figure 6.1b). For small σ, we obtain the probability really at the dividing surface. We are however not too sensitive to our choice of σ, since it is inversely proportional to the flux that leaves the dividing surface ($\langle v_\perp \rangle / \sigma$). The velocity $v_\perp$ is along the reaction coordinate and perpendicular to the dividing surface. The sign ‡ indicates the dividing surface.

In a Cartesian system, we can determine energy at a certain position $\mathbf{r}$ in parameter space with a certain velocity $\mathbf{v}$ according to

$$E(\mathbf{r},\mathbf{v}) = E_{\text{kin}} + E_{\text{pot}} = \sum_j^N \frac{1}{2} m_j v_j^2 + V(\mathbf{r}^{3N}) = \sum_j^N \frac{1}{2} m_j (v_{j,x}^2 + v_{j,y}^2 + v_{j,z}^2) + V(\mathbf{r}^{3N}) \quad (6.19)$$

We can make a transformation such that one of the axes lies along the reaction coordinate. All other $3N-1$ coordinates are orthogonal to this coordinate. Since we do not have individual coordinates for each mass, we cannot work with the individual masses of the particles but with reduced masses of the transformed coordinate system as we saw earlier in Chapter 3

$$E(\mathbf{q},\mathbf{v}_t) = E_{\text{kin}} + E_{\text{pot}} = \sum_i^{3N} \frac{1}{2} \mu_i v_{ti}^2 + V(\mathbf{q}^{3N}). \quad (6.20)$$

The total kinetic energy should remain the same. In general, the probability that $(\mathbf{q})$ is in some subspace, S, of A is

$$P_S = \frac{\int_S \exp(-E(\mathbf{q},\mathbf{v})/k_B T) \mathrm{d}\mathbf{q}^{3N}}{\int_A \exp(-E(\mathbf{q},\mathbf{v})/k_B T) \mathrm{d}\mathbf{q}^{3N}} \equiv \frac{Q_S}{Q_A} \quad (6.21)$$

with Q a configurational integral. We can now apply this to determine the TST rate

$$\begin{aligned} k^{\mathrm{TST}} &= (\text{probability of being in TS}) \times (\text{flux out of TS}) \\ &= \left(\frac{\sigma \int_{\ddagger} \exp(-E(\mathbf{q},\mathbf{v})/k_B T) \mathrm{d}\mathbf{q}^{3N-1}}{\int_A \exp(-E(\mathbf{q},\mathbf{v})/k_B T) \mathrm{d}\mathbf{q}^{3N}} \right) \times \left(\frac{\langle v_\perp \rangle}{\sigma} \right) \\ &= \frac{Q_\ddagger}{Q_A} \frac{\int_0^\infty v_\perp \exp(-\frac{1}{2}\mu_\perp v_\perp^2/k_B T) \mathrm{d}v_\perp \int \exp(-\frac{1}{2}\mu_i v_i^2/k_B T) \mathrm{d}\mathbf{v}^{3N-1}}{\int_{-\infty}^\infty \exp(-\frac{1}{2}\mu_\perp v_\perp^2/k_B T) \mathrm{d}v_\perp \int \exp(-\frac{1}{2}\mu_i v_i^2/k_B T) \mathrm{d}\mathbf{v}^{3N-1}} \\ &= \frac{Q_\ddagger}{Q_A} \sqrt{\frac{k_B T}{2\pi\mu_\perp}} \end{aligned} \tag{6.22}$$

Notice that we only integrate over the positive velocities in the term

$$\int_0^\infty v_\perp \exp(-\frac{1}{2}\mu_\perp v_\perp^2/k_B T) \mathrm{d}v_\perp \tag{6.23}$$

since we only want to include the movement leaving the transition state going to state B. One should realize that since

$$k^{\mathrm{TST}} = (\text{probability of being in TS}) \times (\text{flux out of TS}) \tag{6.24}$$

the Transition State Theory Rates only depends on the initial state and the transition state and NOT on the product state.

Let us use a more formal notation. Define the dividing surface subspace as the points $\mathbf{q}$ that satisfy

$$f(\mathbf{q}) = 0 \tag{6.25}$$

then the TST rate constant can be written as

$$k^{\mathrm{TST}} = \frac{1}{Z_A} \int \mathrm{d}\mathbf{p} \int \mathrm{d}\mathbf{q} \exp\left(-\frac{E(\mathbf{p},\mathbf{q})}{k_B T} \right) \delta(f(\mathbf{q})) \left(\frac{\partial f(\mathbf{q})}{\partial \mathbf{q}} \cdot \frac{\mathbf{p}}{m} \right) H\left(\frac{\partial f(\mathbf{q})}{\partial \mathbf{q}} \cdot \mathbf{p} \right). \tag{6.26}$$

Z_A in this expression is the partition integral over both configuration and momentum space and the Dirac delta function, $\delta(f(\mathbf{q}))$, ensures that only points at the dividing surface contribute to the integral. The next term gives the normal component of the velocity which includes both the contributions pointing towards A and B. The Heaviside function $H(x)$

$$H(x) = \int_{-\infty}^{x} \delta(t) \mathrm{d}t \tag{6.27}$$

which gives

$$H(x) = \frac{1 + \mathrm{sgn}(x)}{2} = \begin{cases} 0, & x < 0 \\ \frac{1}{2}, & x = 0 \\ 1, & x > 0. \end{cases} \tag{6.28}$$

ensures that only the velocities in the direction of B are included.

Transition state theory assumes that once the system reaches the dividing surface this always results in a crossing of the barrier. In reality this is not the case. Depending on the shape of the PES, recrossings can be quite important. It can be shown that TST always overestimates the rate of escape from a given initial state if all possible processes are included. This is mainly because of recrossing. This leads to a variational principle which can be used to find the optimal dividing surface.

6.2.1 Harmonic Transition State Theory

Often the harmonic approximation is used to go from the general form of the classical transition state rate to a form that is easier to apply. This approximation assumes a quadratic form of the potential close to the minimum and saddle point:

$$V_A(\mathbf{q}) \approx V_{\min} + \sum_{i=1}^{3N} \frac{1}{2}\kappa_{A,i} q_{A,i}^2 \tag{6.29}$$

and

$$V_\ddagger(\mathbf{q}) \approx V_{\mathrm{SP}} + \sum_{i=1}^{3N-1} \frac{1}{2}\kappa_{\ddagger,i} q_{\ddagger,i}^2. \tag{6.30}$$

The harmonic TST rate now becomes

$$\begin{aligned} k^{HTST} &= \sqrt{\frac{k_B T}{2\pi\mu_\perp}} \frac{Q_\ddagger}{Q_A} \\ &= \sqrt{\frac{k_B T}{2\pi\mu_\perp}} \frac{\int_\ddagger \exp(-(V_{\mathrm{SP}} + \sum_{i=1}^{3N-1} \frac{1}{2}\kappa_{\ddagger,i} q_{\ddagger,i}^2)/k_B T) \mathrm{d}\mathbf{x}_\ddagger}{\int_A \exp(-(V_{\min} + \sum_{i=1}^{3N} \frac{1}{2}\kappa_{A,i} q_{A,i}^2)/k_B T) \mathrm{d}\mathbf{x}_A} \\ &= \sqrt{\frac{k_B T}{2\pi\mu_\perp}} \frac{\exp(-V_{\mathrm{SP}}/k_B T) \Pi_{i=1}^{3N-1} \sqrt{2\pi k_B T/\kappa_{\ddagger,i}}}{\exp(-V_{\min}/k_B T) \Pi_{i=1}^{3N} \sqrt{2\pi k_B T/\kappa_{A,i}}} \\ &= \frac{\Pi_{i=1}^{3N} \nu_{A,i}}{\Pi_{i=1}^{3N-1} \nu_{\ddagger,i}} \exp(-(V_{\mathrm{SP}} - V_{\min})/k_B T) \end{aligned} \tag{6.31}$$

using

$$\nu = \frac{1}{2\pi}\sqrt{\frac{\kappa}{\mu}}. \tag{6.32}$$

This has the general form

$$k^{HTST} = \nu \exp(-E_a/kT) \tag{6.33}$$

which is the same as the empirical Van 't Hoff-Arrhenius equation.

Chapter 7

Properties

In Chapter 6, we discussed the general approach to obtain expectation values of properties by statistical averaging. This chapter will discuss some standard properties that can be measured during Molecular Dynamics and Metropolis Monte Carlo runs.

7.1 Radial Distribution Function

One of the standard properties to probe is the radial distribution function $g(r)$. We take a typical atom and draw two spheres of radii r and $r + dr$. We then count the number of atoms whose centers lie between these two spheres, and repeat the process for a large number N of atoms. If the result for atom i is $g_i(r)dr$, then the radial distribution function is defined as

$$g(r)dr = \frac{1}{N}\sum_{i=1}^{N} g_i(r)dr. \tag{7.1}$$

This process then has to be repeated for many complete shells over the range of values of r thought to be significant. In the case of an ideal gas, we would expect to find the number of particles to be proportional to the volume enclosed by the two spheres, which is $4\pi r^2 dr$. This gives $g(r) = 4\pi r^2$, a simple quadratic curve.

Consider now a simple cubic solid with lattice constant a. Each atom is surrounded by 6 nearest neighbors at a distance a, 12 next-nearest neighbors at a distance $\sqrt{2}a$, 8 next-next nearest neighbors at a distance $\sqrt{3}a$, 6 two cells away and so on. We would therefore expect to find a discrete radial distribution function. The height of each peak is proportional to the number of atoms a distance r from any given atom. The radial distribution function can also be normalized with respect to an ideal gas. Eq. (7.1) is then multiplied by $V/(4\pi r^2 N)$).

The radial distribution function is an important measure because several key thermodynamic properties, such as potential energy and pressure can be calculated from it. It plays a central role in theories of liquids. It is therefore often determined for, for instance, Lennard-Jones systems. On the other hand, neutron and X-ray scattering experiments on simple fluids, light-scattering experiments on colloidal suspensions, and diffraction experiments on solids give information about $g(r)$. The simulated $g(r)$ can therefore also be used as a test of the accuracy of the model by comparing them to experiments.

For a 3-D system where particles interact via pairwise potentials, the potential energy of the system can be calculated as follows

$$U = 2N\pi\rho \int_0^\infty r^2 u(r) g(r) \mathrm{d}r \tag{7.2}$$

where N is the number of particles in the system, ρ is the number density, $u(r)$ is the pair potential. The pressure of the system can also be calculated by relating the second virial coefficient to

$g(r)$. The pressure can be calculated through

$$P = \rho k_T - \frac{2}{3}\pi\rho^2 \int_0^\infty dr \frac{du(r)}{dr} r^3 g(r). \tag{7.3}$$

Note that the results of potential and pressure will not be as accurate as directly calculating these properties because of the averaging involved with the calculation of $g(r)$.

7.2 Pair Correlation Functions

For a homogeneous sample one has the distribution function

$$g(r) = \frac{V}{4\pi r^2 N} \frac{\mathrm{d}n(r)}{\mathrm{d}r}$$

In a mixture of species A and B, one can calculate the pair correlation function in a similar way

$$g_{\mathrm{AB}}(r_{\mathrm{AB}}) = \frac{V^2}{4\pi r^2 N_{\mathrm{A}} N_{\mathrm{B}}} \frac{\mathrm{d}n_{\mathrm{B}}(r_{\mathrm{AB}})}{\mathrm{d}r_{\mathrm{AB}}} \tag{7.4}$$

The terms "Radial Distribution Function" and "Pair Correlation Functions" are used interchangeable and also the exact normalization can vary. The pair correlation functions $g_{\mathrm{AB}}(r_{\mathrm{AB}})$ and $g_{\mathrm{AA}}(r_{\mathrm{AA}})$ are of special importance, as they are directly related (via a Fourier transform) to the structure factor of the system and can thus be determined experimentally using X-ray diffraction or neutron diffraction.

7.3 Diffusion

In MD time is defined and we can determine dynamic equilibrium properties, unlike in Metropolis MC where time is not defined. The terms "dynamic" and "equilibrium" seem to be in contradiction, but it turns out that time-dependent behavior of a system that is only slightly perturbed w.r.t. equilibrium is completely described by the dynamic equilibrium properties of the system. The diffusion coefficient is an example of such property. This can be determined in both MD and kinetic Monte Carlo simulations. If the diffusion of a molecule in a system consisting of molecules of the same species the term self-diffusion is used.

A typical way to quantify diffusion is to determine the mean square displacement $\left\langle |\boldsymbol{r}(t) - \boldsymbol{r}(0)|^2 \right\rangle$ in time. Here the traveled distance with respect to a time origin of one or more species is determined and averaged over either many species or, in the case of one species, over many time origins (trajectories). So in a simulation, we can obtain the mean square displacement by

$$\left\langle |\boldsymbol{r}(t) - \boldsymbol{r}(0)|^2 \right\rangle = \frac{1}{N} \sum_{i=1}^{N} |\boldsymbol{r_i}(t) - \boldsymbol{r_i}(0)|^2 \tag{7.5}$$

or by

$$\left\langle |\boldsymbol{r}(t) - \boldsymbol{r}(0)|^2 \right\rangle = \frac{1}{M} \sum_{j=1}^{M} |\boldsymbol{r}(t'(j) - t'_0(j)) - \boldsymbol{r}(t'_0(j))|^2. \tag{7.6}$$

with M the number of time origins and $t'(j) - t'_0(j) = t$. In the latter case, we divide our long simulation into M shorter intervals and we average over these M intervals. In both cases, we should be careful with periodic boundary conditions.

The diffusion constant, D, can be linked to the slope of the mean square displacement as a function of time

$$\frac{\partial \left\langle |\boldsymbol{r}(t) - \boldsymbol{r}(0)|^2 \right\rangle}{\partial t} = 2dD \tag{7.7}$$

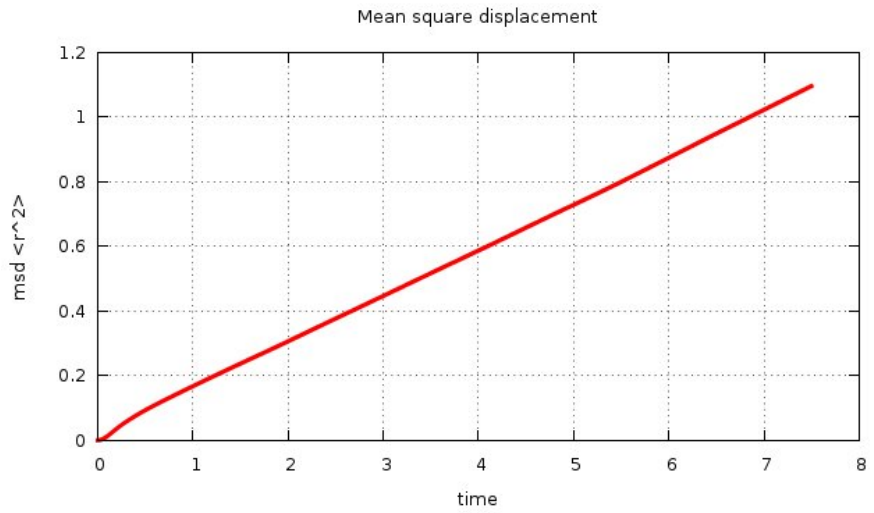


Figure 7.1: *Example of the mean square displacement as a function of time for a Lennard-Jones fluid.*

where d is the dimensionality of the system. So $d = 2$ for surface diffusion and $d = 3$ for bulk diffusion. Typically after some time, a linear regime is reached where the mean square displacement increases linearly in time. This can be seen in Figure 7.1. In this linear regime, we can use

$$\lim_{t\to\infty} \left\langle |\boldsymbol{r}(t) - \boldsymbol{r}(0)|^2 \right\rangle = 2dDt \tag{7.8}$$

We could also express the displacement in terms of velocity

$$\boldsymbol{r}(t) - \boldsymbol{r}(0) = \int_0^t \mathrm{d}t' \boldsymbol{v}(t') \tag{7.9}$$

In one dimension:

$$\left\langle |x(t) - x(0)|^2 \right\rangle = \left\langle \left(\int_0^t \mathrm{d}t' v_x(t') \right)^2 \right\rangle = \int_0^t \int_0^t \mathrm{d}t' \mathrm{d}t'' < v_x(t') v_x(t'') > \tag{7.10}$$

The mean square displacement is thus related to the velocity autocorrelation function $< v_x(t') v_x(t'') >$.

Chapter 8

Molecular Dynamics

Molecular Dynamics is a technique for computing the equilibrium and transport properties of a classical many-body system. Although the method simulates equilibrium conditions, time is defined, unlike in Metropolis Monte Carlo. This seems a contradiction but it allows to determine for instance the diffusion constant of individual atoms or molecules, even though the full system is in equilibrium.

Molecular Dynamics describes the classical motion of the atoms using Newton's equation of motion; quantum effects are therefore ignored. Where the fundamental expression for Monte Carlo techniques is the master equation, for Molecular Dynamics this is Newton's equation of motion:

$$\mathbf{F}_{\mathrm{A}} = m_{\mathrm{A}} \frac{\mathrm{d}^2 \mathbf{R}_{\mathrm{A}}}{\mathrm{d}t^2} = m_{\mathrm{A}} \frac{\mathrm{d}\mathbf{v}_{\mathrm{A}}}{\mathrm{d}t}. \tag{8.1}$$

We know

$$\mathbf{F}_{\mathrm{A}} = -\nabla U_{\mathrm{A}} \tag{8.2}$$

and so based on the potential and an initial velocity and position, we can calculate the motion of the system.

In equilibrium the velocities follow a Maxwell-Boltzmann distribution:

$$f_{\mathbf{v}}(v_x, v_y, v_z) = \frac{N_i}{N} = \frac{\exp(-\frac{mv_i^2}{2kT})}{\sum_j \exp(-\frac{mv_j^2}{2kT})} = \frac{\exp(-\frac{mv_i^2}{2kT})}{M} \tag{8.3}$$

with

$$\begin{aligned} M &= \int \exp(-\frac{mv^2}{2kT}) \mathrm{d}^3 v \\ &= \int \exp(-\frac{mv_x^2}{2kT}) \mathrm{d}v_x \int \exp(-\frac{mv_y^2}{2kT}) \mathrm{d}v_y \int \exp(-\frac{mv_z^2}{2kT}) \mathrm{d}v_z \\ &= \left(\frac{2\pi kT}{m}\right)^{\frac{3}{2}} \end{aligned} \tag{8.4}$$

Each component of the velocity vector has a normal distribution with mean

$$\mu_{v_x} = \mu_{v_y} = \mu_{v_z} = 0 \tag{8.5}$$

and standard deviation

$$\sigma_{v_x} = \sigma_{v_y} = \sigma_{v_z} = \sqrt{\frac{kT}{m}}. \tag{8.6}$$

So the vector has a 3-dimensional normal distribution, also called a "multinormal" distribution, with mean

$$\mu_{\mathbf{v}} = \mathbf{0} \tag{8.7}$$

and standard deviation

$$\sigma_{\mathbf{v}} = \sqrt{\frac{3kT}{m}}. \tag{8.8}$$

To obtain the distribution of the total velocity $v = \sqrt{v_x^2 + v_y^2 + v_z^2}$ we should integrate over $v^2 \sin\phi \, \mathrm{d}v \, \mathrm{d}\phi \, \mathrm{d}\theta$ instead of $\mathrm{d}v_x \, \mathrm{d}v_y \, \mathrm{d}v_z$

$$f_v(v) = \frac{\int \exp(-\frac{mv^2}{2kT}) v^2 \sin\phi \, \mathrm{d}\phi \, \mathrm{d}\theta}{\int \exp(-\frac{mv^2}{2kT}) v^2 \sin\phi \, \mathrm{d}\phi \, \mathrm{d}\theta \, \mathrm{d}v} = \sqrt{\frac{2}{\pi}} \left(\frac{m}{kT}\right)^3 \exp\left(-\frac{mv^2}{2kT}\right) v^2 \tag{8.9}$$

We can also relate temperature to the average velocities of particles

$$\left\langle \frac{1}{2} m v_\alpha^2 \right\rangle = \frac{1}{2} kT \tag{8.10}$$

with α = x, y or z. In an MD simulation this relation is used to determine the temperature at each point in the simulation. Since the kinetic energy fluctuates, so does the instantaneous temperature

$$T(t) = \sum_i^N \frac{m_i v_i^2(t)}{k N_\mathrm{f}} \tag{8.11}$$

with N_f the number of degrees of freedom (= $3N - 3$ for a system of N particles with fixed total momentum)

MD simulations all roughly follow the same algorithm

1. Read the parameters that specify the conditions of the run: initial temperature, number of particles, time step.
2. Initialize the system: initial position and velocities.
3. Compute the forces on all particles.
4. Integrate Newton's equations of motion.
5. After some equilibration time, quantities are measured every so many steps.
6. After completion of the central loop, averages of measured quantities are computed and printed, and stop

In this section we will discuss this algorithm step by step. They are repeated until we have computed the time evolution of the system for the desired length of time. Again the sampling of desired properties (step 5) need to be done between uncorrelated evaluations.

8.1 Initialization

Particles are placed in the simulation box to be compatible with the structure that we want to simulate, *i.e.*, they should not overlap. Easiest way is to place them on a lattice or they can be the results from a geometry optimization. This is comparable to off-lattice Metropolis Monte Carlo, where similar initialization procedures are used. Unlike Metropolis Monte Carlo, each atoms does not only have to be assigned a position but a velocity as well. As explained before these velocities should ultimately follow a Maxwell-Boltzmann distribution but they are often not initialized with the correct distribution. The equilibration period is then not only used to allow the system to relax its structure, but also to obtain the correct velocity distribution. For the initialization of the velocities often the following prescription is applied. To each velocity component of every particle a value is attributed that is drawn from a uniform distribution in the

[-0.5,0.5]. All velocities are then shifted, such that the total momentum is zero and the resulting velocities are scaled to adjust the mean kinetic energy to the desired value

$$v_{\text{scaled}} = v\sqrt{\frac{T_{\text{desired}}}{T(0)}} \tag{8.12}$$

Several integrators do not use velocities but positions at the current and the previous time step. To initialize we need to "make up" previous positions which is done through:

```
1       xp(i) = x(i) - v(i)*dt
```

This is rather crude, but good enough for our initialization.

8.2 Force Calculation

This is the most time-consuming part of most Molecular Dynamics simulations: the calculation of the force acting on every particle. If we consider a model system with pairwise additive interactions, we have to consider the contribution to the force on particle i due to the nearest image of all other particles in the system: $N(N-1)/2$ pairs. This implies that, if we use no tricks, the time needed for the evaluation of the forces scales N^2.

So let us consider a dimer which interacts through a L-J potential

$$U_{\text{AB}} = 4\epsilon\left(\left(\frac{\sigma}{R_{\text{AB}}}\right)^{12} - \left(\frac{\sigma}{R_{\text{AB}}}\right)^{6}\right). \tag{8.13}$$

The force is then

$$\mathbf{F}_{\text{AB}} = 24\epsilon\left(2\left(\frac{\sigma}{R_{\text{AB}}}\right)^{12} - \left(\frac{\sigma}{R_{\text{AB}}}\right)^{6}\right)\frac{\mathbf{R}_{\text{AB}}}{R_{\text{AB}}^2}. \tag{8.14}$$

An algorithm to calculate the force for this system is

```
1  function [f,en] = force(r,Np,box,sigma,epsilon,rc2,ecut)
2                                  % determine the force and energy
3   e = 0 ;                        % set initial energy
4   for i = 1:Np                   %     and forces to zero
5     f(i) = 0;
6   end
7   for i = 1:Np-1                 % loop over all pairs
8     for j = i+1:Np
9       dr = r(i,:) - r(j,:);
10      dr = dr-box*floor(dr/box); % periodic boundary conditions
11      r2 = dot(dr);
12      if (r2 < rc2)              % smaller than cutoff
13        r2i = 1/r2;
14        r6i = (sigma*sigma*r2i)^3;
15        r12i = r6i*r6i;
16        ff =  r2i*(2*r12i-r6i);  % L-J potential
17        f(i,:) = f(i,:) + ff*dr;
18        f(j,:) = f(j,:) - ff*dr;
19        e = e + r12i-r6i -ecut;  % substract energy at cutoff
20      end
21    end
22  end
23  e = 4*epsilon*e                % to save CPU these factors
24  f = 24*epsilon*f               % are taken out of the loop
```

Here we made use of the fact that the first term is the square of the second term and also returns in the force. In this way we only have to take the power once, which is computationally expensive. Multiplying a value with itself is also computationally cheaper than taking the square.

8.3 Periodic Boundary Conditions

In the previous section we came across Periodic Boundary Conditions (PBC). In the first atomistic simulations, a closed simulation box was used. This gave however several problems: the atoms near the edge feel a very different total force than the particles in the center and if an open box in which atoms are allowed to cross the edge is used, the density changes during the simulation. This gives a non-constant number particles and the grand canonical ensemble applies instead of the canonical ensemble. Periodic Boundary Conditions solve these problems and these are general used in molecular simulations. Molecules that move out on one side, move in on the opposite end. Molecules also interact through this periodic boundary. This can be seen in the code in the previous exercise (line). PBC are very easy and straightforward to implement for orthorhombic cells; it can however be more tricky for triclinic cells.

Eventhough PBC is rather straightforward, it introduces some issue that one should be aware off. Beginners sometimes think they are observing a problem when

- the molecule(s) does not stay in the center of the box, or
- it appears that (parts of) the molecule(s) diffuse out of the box, or
- holes are created, or
- broken molecules appear, or
- their unit cell was a rhombic dodecahedron or cubic octahedron but it looks like a slanted cube after the simulation, or
- crazy bonds all across the simulation cell appear.

This could be due to how your visualizer, handles PBC.

When calculating the energy one should consider PBC, especially when determining the cutoff distance. The minimum image convention states that, in a periodic system, the longest cutoff may be no longer than half the shortest box vector. This requirement is based on the periodicity of the system, such that if a particle sees its own periodic image, duplicate forces are calculated and the results are incorrect.

8.4 Ewald summation

Coulomb interactions go with r^{-1} which decays rather slowly. This means that there is still an appreciable interaction at the cutoff. The interaction within one unitcell is

$$E = \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|} = \frac{1}{2} \sum_{i=1}^{N} \sum_{\substack{j=i \\ j \neq i}}^{N} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|}, \tag{8.15}$$

where we either sum over all pairs or do a full sum over all particles in the box. This can be extended to several simulation cells

$$E = \frac{1}{2} \sum_{\substack{\mathbf{n} \\ j \neq i \text{ if } \mathbf{n}=0}} \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j| + \mathbf{n}}. \tag{8.16}$$

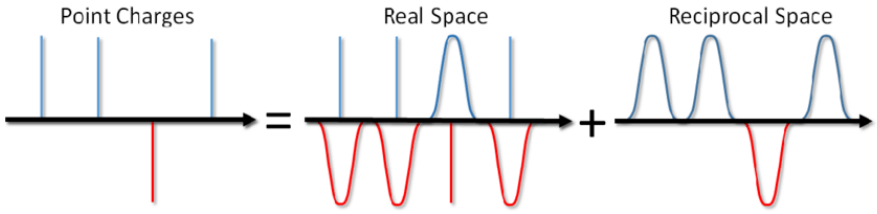


Figure 8.1: *Schematic depiction of Ewald summation. A shielding charge is added around point charge that leads to faster decay in real space. The charge is then subtracted again in reciprocal space.*

The first sum runs over other replications of the simulation cells with $\mathbf{n} = \{n_x L_x, n_y L_y, n_z L_z\}$, L_i the length of the simulations box in direction i, and n_i the number of simulation boxes extended over the periodic boundary in direction i. Eq. (8.16) converges very slowly and it is conditionally convergent since it depends on the order in which you do the summation. For simulation cells with a net charge, it converges either to plus or minus infinity.

Ewald summation is one technique to circumvent these problems. It adds a shielding charge around each point charge. This is typically a Gaussian charge distribution of opposite sign of the point charge. This is schematically depicted in Figure 8.1. The effect is that the interaction decays much faster because of the shielding. The shielding charge is then subtracted again, but the contribution by periodic Gaussian charge distributions can be more easily determined in reciprocal space. Mathematically, the following splitting is made

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1 - f(r)}{r}. \tag{8.17}$$

A usual choice is $f(r) = \mathrm{erf}(\alpha r)$, with erf the error function and α the Ewald splitting parameter. Because of the error function, the direct part now converges rapidly and is not too dependent on the cutoff value. Apart from the original Ewald formulation different schemes exist, but they all relay on the same concept and have three terms of interaction

$$E_{\text{total}} = E_{\text{direct}} + E_{\text{reciprocal}} + E_{\text{correction}}. \tag{8.18}$$

The correction term typically contains the self energy.

8.5 Verlet list

Efficient techniques exist to speed up the evaluation of both short-range and long-range forces in such a way that the computing scales as N, rather than N^2. One of this is to make neighbour lists or cell lists or a combination of the two. The idea behind these lists is that atoms move less than 0.2 Å/time step. So a large fraction of the neighbours, particles which fall in the cutoff range r_{cut}, remains the same during one time step, and it seems wasteful to recalculate which they are every single time. In a Verlet list we keep track of the indices of all atoms within a distance $r_{\text{m}} > r_{\text{cut}}$ which is schematically drawn in Figure 8.2. We only perform force and energy calculations on these pairs. The Verlet lists need to be updated every N_{m} time steps with $r_{\text{m}} - r_{\text{cut}} > N_{\text{m}} \overline{v} \Delta t$ to ensure that no particles have moved from beyond r_{m} to r_{cut}. Thus, we will spend of order N^2 time to compute the Verlet lists, but are rewarded with N_{m} iterations of order Nn^2 where n is the number of particles that are within r_{m}. Optimizing the choice of r_{m}, it can be shown that the simulation does not scale with N^2 in computing time but with $N^{5/3}$. It is even better to keep a list with atom movements and only update if movement is larger than $r_{\text{m}} - r_{\text{cut}}$. This method is less sensitive to fluctuations in velocity and temperature changes. Verlet lists are also often applied in Metropolis Monte Carlo.

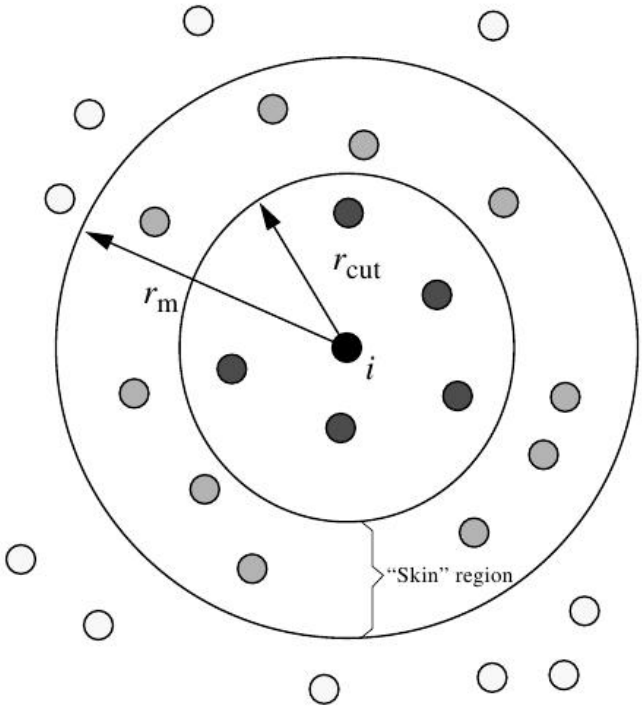


Figure 8.2: *Figure showing the general concept of a Verlet list. A list is maintained of the species within r_m and only these species are tested to be within r_{cut}.*

An other way of reducing the force calculation is by using cell lists which scale with N. This approach is similar to a Verlet list but now the simulation box is subdivided into cells. Particles only interact with particles in its own cell or adjacent cells.

8.6 Integration

Integration, together with the force calculation, makes up the core of the simulation and they are by far the most time consuming parts. A well-known, simple algorithm to integrate Newton's equations of motion is the Verlet algorithm, which is based on the following concept. We start with a Taylor expansion of the coordinate of a particle around time t

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{\mathbf{a}(t)\Delta t^2}{2} + \frac{\mathrm{d}\mathbf{a}(t)}{\mathrm{d}t}\frac{\Delta t^3}{6} + \mathcal{O}(\Delta t^4) \tag{8.19}$$

$$\mathbf{r}(t-\Delta t) = \mathbf{r}(t) - \mathbf{v}(t)\Delta t + \frac{\mathbf{a}(t)\Delta t^2}{2} - \frac{\mathrm{d}\mathbf{a}(t)}{\mathrm{d}t}\frac{\Delta t^3}{6} + \mathcal{O}(\Delta t^4) \tag{8.20}$$

If we sum these to two expressions, we get

$$\mathbf{r}(t+\Delta t) + \mathbf{r}(t-\Delta t) = 2\mathbf{r}(t) + \mathbf{a}(t)\Delta t^2 + \mathcal{O}(\Delta t^4) \tag{8.21}$$

or

$$\mathbf{r}(t+\Delta t) \approx 2\mathbf{r}(t) - \mathbf{r}(t-\Delta t) + \frac{\mathbf{f}(t)\Delta t^2}{m} \tag{8.22}$$

In this way we can obtain the new position using the position in the last two steps and force containing an error of order Δt^4.

As can be seen from Eq. (8.22) the Verlet algorithm does not use the velocity to compute the new positions. One can get an estimate of the velocity using a subtraction of similar Taylor expansions

$$\mathbf{r}(t+\Delta t) - \mathbf{r}(t-\Delta t) = 2\mathbf{v}(t)\Delta t + \mathcal{O}(\Delta t^3) \tag{8.23}$$

$$v(t) = \frac{\mathbf{r}(t+\Delta t) - \mathbf{r}(t-\Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2) \tag{8.24}$$

The velocity is only accurate to order Δt^2 and therefore also the kinetic energy.

Higher order algorithms can easily be constructed by adding or subtracting Taylor expansions of higher order. These algorithm will depend on the position in more time steps and require therefore more memory to store this information. Since the error will be of higher order in Δt, larger time steps can be taken with these higher order algorithms. Verlet, on the other hand, is a not good with large time steps, but requires very little storage. Although, high order integrators show better behaviour on short time scales, they show drift on longer time scales. This could mean that it can be that energy is not conserved on long timescales.

Since Newton's equations of motion are time reversible, therefore the algorithms that are used in MD should be as well. Many integrators are not, or only for infinitely small time steps. Verlet is time reversible. We can show this by propagating with $-\Delta t$

$$\mathbf{r}(t-\Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t+\Delta t) + \frac{\mathbf{f}(t)\Delta t^2}{m}. \tag{8.25}$$

If we rearrange this we find Eq. (8.22) back. The Euler integrator

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{\mathbf{f}(t)\Delta t^2}{2m} + \mathcal{O}(\Delta t^3) \tag{8.26}$$

is not time reversible.

$$\mathbf{r}(t-\Delta t) = \mathbf{r}(t) - \mathbf{v}(t)\Delta t + \frac{\mathbf{f}(t)\Delta t^2}{2m} \tag{8.27}$$

$$\mathbf{r}(t) = \mathbf{r}(t-\Delta t) + \mathbf{v}(t)\Delta t - \frac{\mathbf{f}(t)\Delta t^2}{2m} \tag{8.28}$$

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}(t+\Delta t)\Delta t - \frac{\mathbf{f}(t+\Delta t)\Delta t^2}{2m} \tag{8.29}$$

Algorithms should not only be time reversible but also conserve volume. All trajectories that correspond to a particular energy E are contained in a volume Ω in phase space; the volume in which they are contained in phase space after integration should be equal. Non-area preserving algorithms usually expand in time and have therefore serious energy drift issues. Verlet is area preserving, the Euler algorithm is not. Non-time reversible algorithms are often not area preserving either.

Another often used integration algorithm is Leap Frog. Here the velocities between two times are defined as

$$\mathbf{v}(t-\Delta t/2) \equiv \frac{\mathbf{r}(t) - \mathbf{r}(t-\Delta t)}{\Delta t} \tag{8.30}$$

$$\mathbf{v}(t+\Delta t/2) \equiv \frac{\mathbf{r}(t+\Delta t) - \mathbf{r}(t)}{\Delta t} \tag{8.31}$$

which leads to

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}(t+\Delta t/2)\Delta t. \tag{8.32}$$

If we equal the positions obtained by the Verlet algorithm to the Leap Frog positions, we can get an expression for the velocity at $+\Delta t/2$

$$\begin{aligned}
\mathbf{r}(t) + \mathbf{v}(t+\Delta t/2)\Delta t &= 2\mathbf{r}(t) - \mathbf{r}(t-\Delta t) + \frac{\mathbf{f}(t)\Delta t^2}{m} \\
\mathbf{v}(t+\Delta t/2) &= \frac{\mathbf{r}(t) - \mathbf{r}(t-\Delta t)}{\Delta t} + \frac{\mathbf{f}(t)\Delta t}{m} \\
\mathbf{v}(t+\Delta t/2) &= \mathbf{v}(t-\Delta t/2) + \frac{\mathbf{f}(t)\Delta t}{m}
\end{aligned} \tag{8.33}$$

Since the velocities, which determine the kinetic energy, are determined at a different time from the positions, which determines the potential energy, the total energy cannot be determined.

8.7 Simulations in different ensembles

This setup is to work in a N, V, E ensemble (microcanonical). Often, it is much more realistic to use a N, V, T or N, P, T. Here we will only discuss the canonical ensemble. For constant temperature simulations we should first realize what this means. In statistical physics, a certain temperature is always related to a Boltzmann distribution of energy. The distribution of the kinetic energy is directly related to the velocities:

$$\langle E_{\text{kin}} \rangle = \frac{1}{2} m \langle v^2 \rangle = \frac{3}{2} kT \tag{8.34}$$

If we would force the temperature of our simulated system to be constant throughout the simulation this would also remove all velocity and kinetic energy dispersion. The relative variance in the kinetic energy is

$$\frac{\sigma_{v^2}^2}{\langle v^2 \rangle^2} = \frac{\langle v^4 \rangle - \langle v^2 \rangle^2}{\langle v^2 \rangle^2} \tag{8.35}$$

For this we use the Maxwell-Boltzmann distribution

$$\begin{aligned} \langle v_x^2 \rangle &= \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} \int v_x^2 \exp(-\frac{mv_x^2}{2kT}) \mathrm{d}v_x \int \exp(-\frac{mv_y^2}{2kT}) \mathrm{d}v_y \int \exp(-\frac{mv_z^2}{2kT}) \mathrm{d}v_z \\ &= \frac{kT}{m} \end{aligned} \tag{8.36}$$

and so $\langle v^2 \rangle = 3\frac{kT}{m}$. In a similar way we can obtain $\langle v^4 \rangle = 15(\frac{kT}{m})^2$ using $v^4 = (\sum_\alpha v_\alpha^2)^2$. This leads to

$$\frac{\sigma_{v^2}^2}{\langle v^2 \rangle^2} = \frac{\langle v^4 \rangle - \langle v^2 \rangle^2}{\langle v^2 \rangle^2} = \frac{2}{3} \tag{8.37}$$

If we would use the kinetic energy per particle as a measure of the instantaneous temperature:

$$T_k = \sum_i^N \frac{m v_i^2(t)}{k N_{\text{f}}} \tag{8.38}$$

and

$$\langle T_k \rangle = \frac{m}{k N_{\text{f}}} N \langle v^2 \rangle \tag{8.39}$$

and

$$\langle T_k^2 \rangle = \frac{m^2}{k^2 N_{\text{f}}^2} (N \langle v^4 \rangle + N(N-1) \langle v^2 \rangle^2 \tag{8.40}$$

we would find that, in a canonical ensemble, this temperature fluctuates. Its relative variance is

$$\begin{aligned} \frac{\sigma_{T_k}^2}{\langle T_k \rangle_{NVT}^2} &= \frac{\langle T_k^2 \rangle_{NVT} - \langle T_k \rangle_{NVT}^2}{\langle T_k \rangle_{NVT}^2} \\ &= \frac{N \langle v^4 \rangle + N(N-1) \langle v^2 \rangle^2 - N^2 \langle v^2 \rangle^2}{N^2 \langle v^2 \rangle^2} = \frac{2}{3N} \end{aligned} \tag{8.41}$$

So even in the canonical ensemble the temperature fluctuates. Our MD simulation should reflect this.

Several "thermostats" have been suggested to keep the temperature constant. Often used examples are the Andersen thermostat and the Nosé-Hoover thermostat. Thermostats are usually

controlled by a thermostat parameter. This parameter should be chosen carefully to ensure that no temperature drift occurs but that at the same time the temperature fluctuations scale with $2/(3N)$.

The Andersen Thermostat couples the system to a heat bath represented by stochastic impulsive forces that act occasionally on randomly selected particles. The coupling strength is determined by the frequency of the stochastic collisions, ν

$$P(t) = \nu \exp(-\nu t) \tag{8.42}$$

where $P(t)\mathrm{d}t$ is the probability that the next collision will take place in interval $[t, t + \mathrm{d}t]$ When a particle undergoes a collision with the heat bath, it is assigned a new velocity drawn from the Maxwell-Boltzmann distribution at the desired temperature.

Chapter 9

Metropolis Monte Carlo

As described in Chapter 6 a property of a system can be determined by

$$\langle A \rangle = \frac{\int A \exp(-E(\mathbf{r}, \mathbf{v})/k_B T) \mathrm{d}\mathbf{r}^{3N}}{\int \exp(-E(\mathbf{r}, \mathbf{v})/k_B T) \mathrm{d}\mathbf{r}^{3N}} \tag{9.1}$$

applying statistical mechanics. Since this cannot be done analytically in many cases, there was a need to be able to do this numerically. This, together with the first electronic computer, lead to the development of the Monte Carlo method. The first electronic computer, ENIAC, was designed to compute firing tables during the second World War. John von Neumann thought this computer could also be used to make a computational model for thermonuclear reactions. Together with Nicholas Metropolis and Stanislaw Ulam he developed a statistical approach to solve the problem of neutron diffusion in fissionable material. This approach uses *pseudo*-random numbers, hence the name "Monte Carlo". Ulam's uncle liked to gamble and would borrow money from relatives saying he 'just had to go to Monte Carlo'. Enrico Fermi independently developed the Monte Carlo method using the FERMIAC which was a mechanical device to study neutron transport.

Currently, the method is being used for many different applications, which can be roughly classified in three groups: integration (I), evolution (E), and optimization (O). It is being used in many different fields/areas of research. A few examples are

1. Ensemble averages in statistical physics (I)
2. Radiative transfer (I+E)
    - light from an astronomical object to the telescope
    - light transport in biological tissue
3. Finance (stock market behavior) (E)
4. Evolution of a physical system (E)
    - Evolution of chemical abundances in a cloud
    - Evolution of a particle on a surface
    - Crystallization
5. Particle physics (I+E)
6. Geometry optimization of a system (O)
7. Transport (E)
8. ....

The type of application is indicated in the parentheses. As becomes clear from this list of examples, all three types of applications are used in chemistry and physics. Simulated annealing, which was discussed in Chapter 4, is a special form of Monte Carlo simulation used for optimization. The next chapter will introduce kinetic Monte Carlo which is a way to study the evolution of a system using the principles of Monte Carlo. This chapter will mainly deal with integration, which is still the most used application. We will refer to the application of this technique to solve Eq. (9.1) as Metropolis Monte Carlo, for reasons that will become clear later in this chapter. This technique is an equilibrium method to determine ensemble averages. Kinetic Monte Carlo and Metropolis Monte Carlo have a different background and different capabilities. For optimization routines this is less the case. A piece of advise is to always make sure you understand what type of Monte Carlo method is used or in a conversation with a fellow modeler to ensure that you are on the same page and discuss the same method. This can prevent unnecessary, complicating confusion.

9.1 Integration by Monte Carlo

As explained before, Monte Carlo started as a way to numerically solve integrals in statistical physics. Now let us start with considering the following integral:

$$I = \int_{x_1}^{x_2} f(x)\mathrm{d}x \tag{9.2}$$

In a first attempt to integrate this integral numerically, one could divide the interval $[x_1, x_2]$ into N equally spaced intervals with width Δx. The approximated value for I would then simply be

$$I = \sum_{n=0}^{N} f(x_1 + n\Delta x)\Delta x. \tag{9.3}$$

This will lead to the correct result for $N \to \infty$ or $\Delta x \to 0$, but it is computationally very demanding, especially if one is interested in many dimensions.

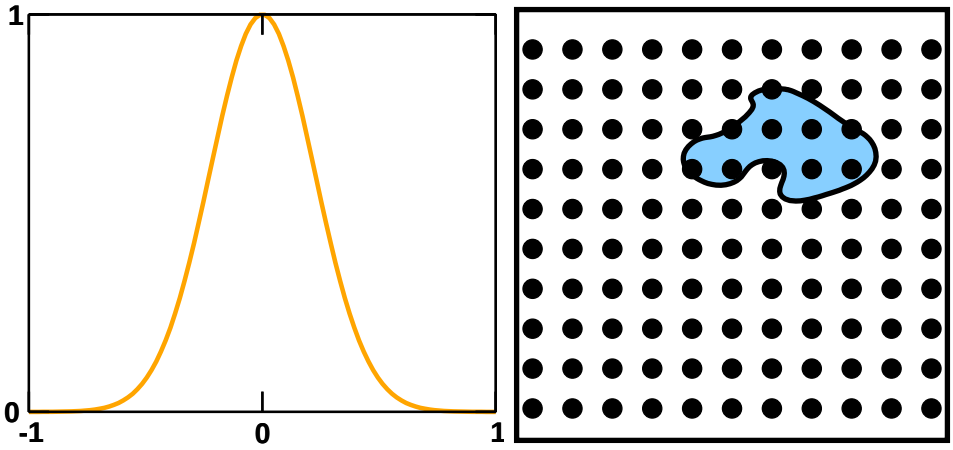
A second attempt would be by only considering M randomly chosen points. The approximation for I would then be

$$I = \frac{x_2 - x_1}{M} \sum_{n=1}^{M} f(x_n) \tag{9.4}$$

with x_M randomly picked on interval $[x_1, x_2]$. This will lead to the correct result for $M \to \infty$. Let us look at the example of integrating the area under the Gaussian plotted in Figure 9.1a. We draw a random number x_i where the x_i is uniformly distributed in $[-1, 1]$. We average the value y_i over all draws. Since the random number is uniformly distributed in our parameter space, we call this uniform sampling. Let us now consider a function with a two dimensional parameter space with its highest function values in one local area as depicted schematically in Figure 9.1b. To get an accurate determination of integral of this function we need to sample many points in the blue area. Uniform sampling will lead to many 'wasted' samples outside the blue area that do not contributed to I. This becomes worse with increasing dimensions.

9.2 Importance sampling

The solution to this problem is *importance sampling*. The aim is to have a higher probability to pick a point in parameter space that has a large contribution to I. We can rearrange our original



(a) (b)

Figure 9.1: *Two example functions to integrate by Monte Carlo: (a) a Gaussian and (b) a 2D surface with only a large contribution to the integral in the blue area.*

integral in the following way

$$I = \int_{x_1}^{x_2} f(x)\mathrm{d}x$$
$$I = \int_{x_1}^{x_2} \frac{f(x)}{w(x)} w(x)\mathrm{d}x \tag{9.5}$$

and we define

$$\frac{\mathrm{d}u(x)}{\mathrm{d}x} \equiv w(x). \tag{9.6}$$

The integral now becomes

$$I = \int_{u_1}^{u_2} \frac{f(x(u))}{w(x(u))}\mathrm{d}u. \tag{9.7}$$

You might wonder what we gain by this redefinition and rearrangement. This becomes clear when we make an estimate for the error in our sampling

$$\sigma^2 = \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\left\langle \left(\frac{f(x(u_i))}{w(x(u_i))} - < f/w >\right)\left(\frac{f(x(u_j))}{w(x(u_j))} - < f/w >\right)\right\rangle. \tag{9.8}$$

The angular brackets denote here the true average in the limit $M \to \infty$. If i and j are independent,

$$\sigma^2 = \frac{1}{M^2}\sum_{i=1}^{M}\left\langle \left(\frac{f(x(u_i))}{w(x(u_i))} - < f/w >\right)^2\right\rangle$$
$$\sigma^2 = \frac{1}{M}\left[< (f/w)^2 > - < f/w >^2\right]. \tag{9.9}$$

If $w(x)$ is chosen such that f/w is nearly constant, $< (f/w)^2 >$ approaches $< f/w >^2$ and σ becomes small even for small M. Different weighing functions w have been suggested. The most commonly used function was suggested by Metropolis et al [17] and leads to *Metropolis importance sampling*.

Since in statistical physics we often want to solve the following integral

$$\langle A \rangle = \frac{\int A(\mathbf{r}^N)\exp(-U(\mathbf{r}^N)/kT)\mathrm{d}\mathbf{r}^N}{Q}. \tag{9.10}$$

Notice that this not only contains an integral in the numerator. The denominator consists of unknown constant Q, the partition function, which is an integral as well. Let us choose

$$w(\mathbf{r}^N) = \frac{\exp(-U(\mathbf{r}^N)/kT)}{Q} \tag{9.11}$$

then

$$\langle A \rangle \approx \frac{1}{M} \sum_i^M A(\mathbf{r}_i) \tag{9.12}$$

The sampling points are now not uniformly distributed but according to the Boltzmann factor.

How to do this, since Q is not known? Metropolis and coworkers solved this problem by the following algorithm

1. Choose initial conditions
2. Make a small displacement to obtain a new configuration
3. Evaluate E for the new configuration
4. If $\Delta E < 0$, accept the new configuration, else accept with a $\exp(-\Delta E/kT)$ probability
5. Once every so-many iterations, determine A
6. Repeat from step 2
7. Average all A's

The reason why this works is that a point is chosen with respect to a previous point. So instead of using Eq. (9.11) which contains Q,

$$\frac{w(\text{trial})}{w(\text{current})} = \frac{\exp(-U(\text{trial})/kT)}{\exp(-U(\text{current})/kT)} \tag{9.13}$$

can be applied to determine the probability to pick the next point, which has no dependence on Q. The step size for the trial move is usually chosen such that the acceptance probability is ~ 0.5, although there is no real reason why this should be the most optimal choice.

The Metropolis algorithm produces a Markov chain of different configurations, which is effectively a random walk through parameter space where each steps only depends on the current position and not on previous steps. Typically, property A is not determined at each iteration to contribute to Eq. (9.12). This is because the two successive configurations are highly correlated. After $N/P_{acceptance}$ iterations two evaluations are usually considered not to be correlated, where N is the number of particle in the system. Often the term "cycle" is reserved for the combined iterations after which A is sampled, sometimes the term "cycle" is used for one iteration.

9.3 Umbrella sampling

Clearly the Metropolis algorithm only gets an accurate value for $\langle A \rangle$ if the system is *ergodic*, *i.e.*, there is a finite probability to reach any point in configuration space. Systems in which energy barrier separates two or more regions of configuration space may suffer from poor sampling, since the probability that the system leaves the part of configuration space where the simulation started is small. Umbrella sampling might help in these cases.

Considering a system, which state can described by an order parameter λ: low λ represents liquid, high λ solid. Both liquid and solid phases are low in energy, but are separated by a free energy barrier at intermediate values of λ. This prevents the simulation from adequately

sampling both phases. We can solve this by applying a bias potential and sample in a slightly different ensemble π, in which configurations are sampled with a probability proportional to

$$\pi(\mathbf{r}^N) = \frac{\exp(-U(\mathbf{r}^N)/kT + W(\mathbf{r}^N))}{\int \exp(-U(\mathbf{r}^N)/kT + W(\mathbf{r}^N))\mathrm{d}\mathbf{r}^N} \tag{9.14}$$

instead of

$$w(\mathbf{r}^N) = \frac{\exp(-U(\mathbf{r}^N)/kT)}{Q}. \tag{9.15}$$

$W(\mathbf{r}^N)$ is the so-called bias function or bias potential. In the case of the melting solid, $W(\lambda)$ may be chosen such that it takes high values at intermediate λ and low values at low/high λ, facilitating barrier crossing. The expectation value of property A can be rewritten in this ensemble

$$\begin{aligned}
\langle A \rangle &= \frac{\int A(\mathbf{r}^N)\exp(-U(\mathbf{r}^N)/kT)\mathrm{d}\mathbf{r}^N}{\int \exp(-U(\mathbf{r}^N)/kT)\mathrm{d}\mathbf{r}^N} \\
&= \frac{\int A(\mathbf{r}^N)\exp(-U(\mathbf{r}^N)/kT + W(\mathbf{r}^N))\exp(-W(\mathbf{r}^N))\mathrm{d}\mathbf{r}^N}{\int \exp(-U(\mathbf{r}^N)/kT + W(\mathbf{r}^N))\exp(-W(\mathbf{r}^N))\mathrm{d}\mathbf{r}^N} \\
&= \frac{\frac{\int A(\mathbf{r}^N)\exp(-U(\mathbf{r}^N)/kT+W(\mathbf{r}^N))\exp(-W(\mathbf{r}^N))\mathrm{d}\mathbf{r}^N}{\int \exp(-U(\mathbf{r}^N)/kT+W(\mathbf{r}^N))\mathrm{d}\mathbf{r}^N}}{\frac{\int \exp(-U(\mathbf{r}^N)/kT+W(\mathbf{r}^N))\exp(-W(\mathbf{r}^N))\mathrm{d}\mathbf{r}^N}{\int \exp(-U(\mathbf{r}^N)/kT+W(\mathbf{r}^N))\mathrm{d}\mathbf{r}^N}} \\
&= \frac{\langle A\exp(-W)\rangle_\pi}{\langle \exp(-W)\rangle_\pi}
\end{aligned} \tag{9.16}$$

We can now run our MC simulation with an additional bias function that lowers the barrier and average $A\exp(-W)$ and $\exp(-W)$ to obtain the expectation value of A in the canonical ensemble. The advantage of umbrella sampling is that one samples a larger area of parameter space if chosen the correct bias function which is however not straightforward.

9.4 Microscopic reversibility

Let us now return to the acceptance criterion. For Metropolis sampling, a trial move is always accepted when $\Delta U < 0$ and otherwise it is accepted with probability $\exp(-\Delta U/kT)$ where ΔU or

$$\tilde{k}_{c\to t} = \min(1, \exp\left[-(U(t) - U(c))/kT\right]) \tag{9.17}$$

where c denotes the current configuration and t the configuration after the trial move. For simulations in the canonical ensemble, the chance that we pick a certain configuration depends on its Boltzmann weight

$$P(\mathbf{r}^N) = \frac{\exp(-U(\mathbf{r}^N)/kT)}{Q}. \tag{9.18}$$

So the probability to pick configuration t with respect to picking configuration c is

$$\frac{P(t)}{P(c)} = \frac{\exp(-U(t)/kT)}{\exp(-U(c)/kT)}. \tag{9.19}$$

In order to achieve this, the acceptance probability from going from current state c to trial state t with respect to the acceptance probability of the reverse action should obey the same ratio. This is called *detailed balance* or *microscopic reversibility*

$$\frac{\tilde{k}_{c\to t}}{\tilde{k}_{t\to c}} = \frac{P(t)}{P(c)} = \exp(-(U(t) - U(c))/kT) = \exp(-\Delta U/kT). \tag{9.20}$$

We will come back to this expression at a later stage. The main point here is that we can freely choose one acceptance probability, but then the probability of the inverse action is dictated by detailed balance. Different choices have been proposed. Their quality depends on how fast convergence in averaging is obtained and how well they cover the full parameter space (ergodicity).

Chapter 10

Kinetic Monte Carlo

I will call the Monte Carlo technique applied for integration to obtain an equilibrium property of a system "Metropolis Monte Carlo". This in contrast to the kinetic Monte Carlo technique which will be discussed in the present chapter. The kinetic Monte Carlo technique is an evolutionary simulation technique. The exact start of the technique is not so clear as for Metropolis Monte Carlo. Different names are used including dynamic Monte Carlo.

Kinetic Monte Carlo is a method to solve the master equation. We will give a brief derivation of this equation in order to discuss the underlying assumptions to the derivation of this expression, the different ways in which this equation is solved, the approximations involved, and their advantages and disadvantages. A more thorough derivation can be found in textbooks such as Refs. 27, 6, 8.

Most molecular systems can be described by a state vector x and a time coordinate t. In this chapter, x is an abstract quantity which may be some local minimum on the potential energy surface (PES), which is schematically drawn in the left panel of Figure 10.1. Here the black circles represent different states. For some simulations, for instance in heterogeneous catalysis, a PES is however not explicitly considered and in those cases a state x can be represented by, *e.g.*, the species on a surface, their position, the temperature, etc., or x can simply contain the coordinates of all particles on the surface.

Let us call $p(x,t)$ the probability density for the system to be in state x at time t. The probability density can be written as

$$p(x,t) = \frac{\mathrm{d}P(x,t)}{\mathrm{d}x}, \tag{10.1}$$

where $P(x,t)$ is the probability that the system is in an infinitesimal volume element $\mathrm{d}x$ around x at time t. Since the probability to find the system in any x at a given time t is unity, we have

$$\int \mathrm{d}x p(x,t) = 1. \tag{10.2}$$

The definition of the probability density, Eq. (10.1), can be extended to an n'th order probability density function $p(x_n,t_n;x_{n-1},t_{n-1};\ldots;x_1,t_1)$ which gives the probability density for the system to be in x_1 at t_1 *and* in x_2 at t_2 etc. To describe the time evolution of the system the conditional probability density is introduced

$$p(x_m,t_m;\ldots;x_{n+1},t_{n+1}|x_n,t_n;\ldots;x_1,t_1) = \frac{p(x_m,t_m;\ldots;x_1,t_1)}{p(x_n,t_n;\ldots;x_1,t_1)}, \tag{10.3}$$

which gives the probability density to be in states x_{n+1} through x_m at time t_{n+1} through t_m *given that* it is in states x_1 through x_n at times t_1 through t_n. The higher order conditional probability densities are complicated functions and therefore one usually assumes a Markov

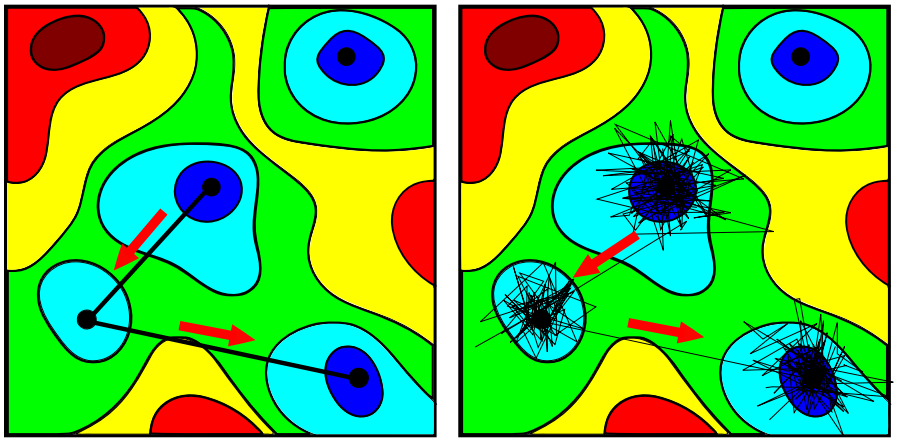


Figure 10.1: *Schematic representation of a potential energy surface (PES). The states, local minima on the PES, are indicated by black solid spheres. (left) The system moves from state to state or (right) the vibrational movement is included as well and the trajectories spend most of their time around the local minimum.*

Chain. To define a Markovian system we imply time-ordering ($t_3 > t_2 > t_1$) and impose the following restriction on the conditional probability density functions

$$p(x_n, t_n|x_{n-1}, t_{n-1}; \ldots; x_1; t_1) = p(x_n, t_n|x_{n-1}, t_{n-1}). \tag{10.4}$$

This means that the probability of arriving in x_n, t_n is only dependent on x_{n-1}, t_{n-1} and not on the states before t_{n-1}. A Markovian system is completely defined by an initial state, say x_1, t_1, and the first order conditional probability densities since any other probability density function can be constructed from these. For example

$$\begin{aligned} p(x_3, t_3; x_2, t_2; x_1; t_1) &= p(x_3, t_3|x_2, t_2; x_1; t_1)p(x_2, t_2; x_1, t_1) \\ &= p(x_3, t_3|x_2, t_2)p(x_2, t_2|x_1, t_1)p(x_1, t_1), \end{aligned} \tag{10.5}$$

where we used Eqs. (10.3) and (10.4).

The Markov chain assumption is valid for memoryless processes. For transitions in molecular systems, we are typically interested in rare events, such as conformational changes, surface or bulk diffusion and desorption of species from a surface. This type of events occur at much longer timescales than the (lattice) vibrations and the event timescale and the vibrational timescale become decoupled. This makes rare events effectively memoryless, since all history about which states where previously visited is lost due to the vibrations between two transitions. This is schematically depicted in Figure 10.1. The residence time of the system in a local minimum of the potential energy surface is several orders of magnitude larger than the vibrational timescale. The system will therefore typically proceed according to a trajectory like the one drawn in the right panel of Figure 10.1. At the time that the system leaves the potential well, all information about the direction from which it initially entered this potential well, is lost. In kinetic Monte Carlo this is approximated by a trajectory as drawn in the left panel. In some cases, when the residence time in a state is short as compared to the vibrational timescale, this approximation breaks down. For these cases, Molecular Dynamics simulations, which use Newton's equations of motion to determine the molecular trajectories, are better suited. These types of simulations would result in trajectories plotted in the panel on the right. A drawback of this technique is however that their simulation timescales is much shorter. Chapter 8 will discuss this technique in more detail.

10.1 Derivation of the master equation

The transition probabilities of Markov processes obey the Chapman-Kolmogorov equation

$$p(x_3, t_3|x_1, t_1) = \int \mathrm{d}x_2 p(x_3, t_3|x_2, t_2) p(x_2, t_2|x_1, t_1). \tag{10.6}$$

This equation describes the transition from x_1, t_1 to x_3, t_3 through all possible values of x_2 at t_2. Let us now assume that the conditional probability densities are time independent and so we can write

$$p(x_2, t_2|x_1, t_1) = p_{\Delta t}(x_2|x_1),\ \Delta t = t_2 - t_1, \tag{10.7}$$

because there is now only a dependence on the time difference Δt. The expression $p_{\Delta t}(x_1, x_2)$ denotes the transition probability density from state x_1 to state x_2 within the time interval Δt. This can be expressed as a rate, k, by

$$k(x_2|x_1) = \frac{p_{\Delta t}(x_2|x_1)}{\Delta t}. \tag{10.8}$$

Using the Chapman-Kolmogorov equation (10.6) for $p_{\Delta t}$, we get

$$p_{\Delta t+\Delta t'}(x_3|x_1) = \int \mathrm{d}x_2 p_{\Delta t'}(x_3|x_2) p_{\Delta t}(x_2|x_1). \tag{10.9}$$

Inserting this into the Chapman-Kolmogorov equation results in

$$p_{\Delta t+\Delta t'}(x_3|x_1) = (1 - k_{\mathrm{tot}}(x_3)\Delta t') p_{\Delta t}(x_3|x_1) + \Delta t' \int \mathrm{d}x_2 k(x_3|x_2) p_{\Delta t}(x_2|x_1), \tag{10.10}$$

and so

$$\begin{aligned} \frac{p_{\Delta t+\Delta t'}(x_3|x_1) - p_{\Delta t}(x_3|x_1)}{\Delta t'} &= \int \mathrm{d}x_2 k(x_3|x_2) p_{\Delta t}(x_2|x_1) \\ &\quad - \int \mathrm{d}x_2 k(x_2|x_3) p_{\Delta t}(x_3|x_1), \end{aligned} \tag{10.11}$$

in which we have used the definition of k_{tot}. In the limit $\Delta t' \to 0$ we arrive at the master equation for the transition probabilities densities

$$\frac{\mathrm{d}}{\mathrm{d}\Delta t} p_{\Delta t}(x_3|x_1) = \int \mathrm{d}x_2 k(x_3|x_2) p_{\Delta t}(x_2|x_1) - \int \mathrm{d}x_2 k(x_2|x_3) p_{\Delta t}(x_3|x_1). \tag{10.12}$$

If we want to obtain a master equation for the state probability density itself, we can multiply this equation by $p(x_1, t)$, integrate over x_1. This will ultimately lead to the master equation

$$\frac{\mathrm{d}}{\mathrm{d}t} p(x_3, t) = \int \mathrm{d}x_2 k(x_3|x_2) p(x_2, t) - \int \mathrm{d}x_2 k(x_2|x_3) p(x_3, t). \tag{10.13}$$

On a discrete state space, which is usually used for molecular simulations, the equation is written in terms of probabilities in the following way:

$$\frac{\mathrm{d}}{\mathrm{d}t} P(x_i, t) = \sum_j \left(k(x_i|x_j) P(x_j, t) - k(x_j|x_i) P(x_i, t) \right). \tag{10.14}$$

The first part of the sum represents the increase in $P(x_i, t)$ because of events that enter state x_i from any other state x_j; the second part is the decrease in $P(x_i, t)$ due to all events leaving x_i. Most chemists will construct this expression intuitively. Here we made the formal derivation of the master equation to show the assumptions involved at the root of this expression which could be a severe restriction, for instance the Markov chain assumption.

10.2 Derivation of time evolution

Gillespie derived the KMC method to numerically simulate the stochastic time evolution of coupled chemical reactions starting from the master equation (Eq. (10.14))[7]. Using this equation, one can derive the time interval between two events, which is one of the fundamental expressions in kinetic Monte Carlo, since it describes how time progresses during a KMC simulation. If one starts from state i at time t (so $P(x_j, t) = 0 \; \forall \; j \neq i$), then for small Δt we have

$$\frac{\mathrm{d}}{\mathrm{d}t} P(x_i, t + \Delta t) = -\sum_j k(x_j|x_i) P(x_i, t + \Delta t) \tag{10.15}$$

from which we can calculate the probability that the system has not left state x_i at time $t + \Delta t$

$$P_{\text{not left}}(x_i, t + \Delta t) = \exp\left(- \int_t^{t+\Delta t} \mathrm{d}t' k_{\text{tot}}(x_i, t') \right), \tag{10.16}$$

where $k_{\text{tot}}(x_i, t)$ is the total rate leaving state x_i. If we assume stationary rates than Eq. (10.16) becomes

$$P_{\text{not left}}(x_i, t + \Delta t) = \exp\left(-k_{\text{tot}}(x_i)\Delta t\right). \tag{10.17}$$

Stationary rates are a reasonable assumption for systems of constant temperature. If the systems temperature varies, the rates change over time and the time interval Δt should be determined differently [13].

In practice, this means a KMC cycle starting from state x_i consists of three steps if all possible events are known: first a final state x_f is picked with probability

$$p_{i \to f} = \frac{k(x_f|x_i)}{k_{\text{tot}}(x_i)}, \tag{10.18}$$

using a random number. Next, the time is advanced where the values for Δt are chosen such that they follow the distribution dictated by Eq. (10.17). This can be numerically achieved by generating a uniform random number Z in range $(0, 1]$ and equating this to the probability that the reaction has not yet occurred,

$$Z = P_{\text{not left}}(x_i, t + \Delta t), \tag{10.19}$$

which leads to

$$\Delta t = -\frac{\ln(Z)}{k_{\text{tot}}(x_i)} \tag{10.20}$$

for stationary rates. Finally, the picked transition is evaluated and the transition rates leaving this new state, x_f, are determined. The consequence of Eqs. (10.18) and (10.20) is that at each given time all possible transition events leaving the current state and their corresponding rates should be known. This combination of events and rates is often referred to as "table of events".

10.3 Microscopic reversibility

A requirement for every rate is that it should fulfill macroscopic reversibility. In equilibrium, the master equation should result in a steady state

$$\sum_{x_f} \left(\tilde{k}(x_i|x_f) P_{\text{eq}}(x_f) - \tilde{k}(x_f|x_i) P_{\text{eq}}(x_i) \right) = 0, \tag{10.21}$$

where $\tilde{k}(x_f|x_i)$ denotes the equilibrium rates, which are equal the transition probabilities times some unit of time. The easiest way to fulfill this equation is to require detailed balance or microscopic reversibility, *i.e.*, the net flux between every pair of states is zero

$$\frac{\tilde{k}(x_i|x_f)}{\tilde{k}(x_f|x_i)} = \frac{P_{\text{eq}}(x_i)}{P_{\text{eq}}(x_f)}. \tag{10.22}$$

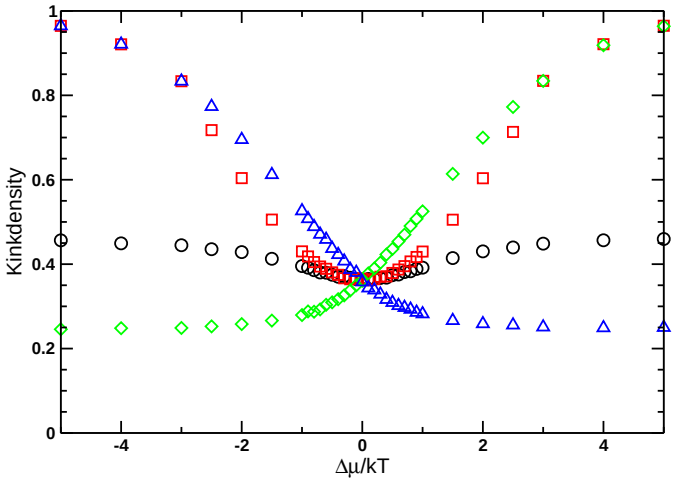


Figure 10.2: *Kink density of a step on a surface of a simple cubic crystal as a function of the driving force for crystallization determined by four different choices of transition probability. All satisfy microscopic reversibility and give the same result in equilibrium.*

For a canonical ensemble

$$\frac{P_{\text{eq}}(x_i)}{P_{\text{eq}}(x_f)} = \exp\left(-\frac{E(x_i) - E(x_f)}{kT}\right). \tag{10.23}$$

This was also obtained for Metropolis Monte Carlo. So again, one can choose one probability and the probability for the reverse process is fixed. In the case of Metropolis Monte Carlo, where we are in equilibrium, the choice of scheme only determines how many iterations are required for equilibration and how easy it is to sample the entire PES. The end result is the same. This can be seen in Figure 10.2 which shows the kink density of a step on a surface of a simple cubic crystal as a function of the driving force for crystallization. The system is in equilibrium at $\Delta\mu/kT = 0$. Different KMC results are shown using different choices for the rates, which are probabilities per unit time, but all obeying microscopic reversibility. The plot shows that the results coincide in equilibrium. Outside equilibrium this results differ significantly.

10.4 KMC algorithm

The theory behind KMC can be implemented into an algorithm in several ways. Here, we will discuss a few algorithms that are most commonly used. Figure 10.3 shows a flow diagram of the KMC algorithm as proposed by Nordmeyer and Zaera[22, 23] which is based on the original paper by Gillespie[7]. This algorithm is mainly developed to study the kinetics of gas adsorption on solid surfaces and it is used to simulate homogeneous surfaces, in other words the desorption and diffusion rates are site independent and also only one type of species is considered. The algorithm starts by inputing all rates, *i.e.,* creating the table of events. Then, the surface is initialized by creating an empty matrix, M, where each entry represents a site and the occupancy of this site can be changed by changing the number. In the next step the desorption, diffusion and deposition rate are calculated. Since a homogeneous surface is assumed this corresponds to

$$R_{\text{desorb}} = N k_{\text{desorb}}, \tag{10.24}$$
$$R_{\text{diff}} = N k_{\text{hop}}, \tag{10.25}$$
$$R_{\text{dep}} = N_{\text{sites}} k_{\text{acc}}, \tag{10.26}$$
$$k_{\text{total}} = R_{\text{desorb}} + R_{\text{diff}} + R_{\text{dep}}, \tag{10.27}$$

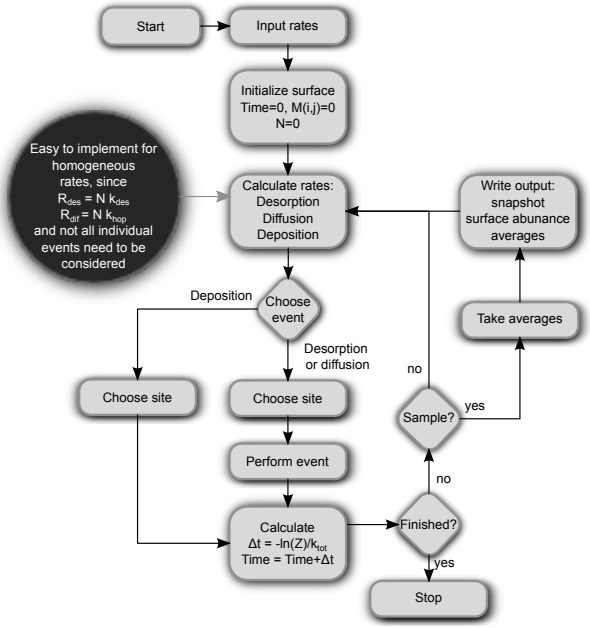


Figure 10.3: *Flow diagram of the KMC algorithm as proposed by Nordmeyer and Zaera[22, 23]*

where N is the number of species on the grain, N_{sites} the number of sites (the size of matrix M) and k_x the rate for an individual atom or molecule to undergo process x. Then one of the three possible events is chosen by picking a random number between 0 and k_{total}, a second random number is picked to choose the site where the event will occur and a third one to increment the simulated time according to Eq. (10.20), and if the termination criterion is met, the simulation will finish; otherwise the loop is entered again.

Upon reentering the simulation loop, the following iterations simply start by recalculating the desorption, diffusion and deposition rates. For homogeneous surfaces this can be done in a straightforward way and this step does not require much effort. Extending the model to systems with more species, would already be more involved and if one would like to treat inhomogeneous surfaces, smart updating routines should be applied to keep the computational cost of this step limited. In that case, however separating the determination of the overall event and the specific location becomes unnecessarily cumbersome and a different algorithm would be preferred, for instance the commonly-used n-fold way algorithm[2].

Figure 10.4 shows a flow diagram based on the n-fold way KMC algorithm[2]. The main difference between this algorithm and the algorithm of Figure 10.3 is that the event and the location are chosen in one step. This makes this algorithm more appropriate for systems where the surface is inhomogeneous or where the adsorbed species interact which each other, since not all sites are equally likely to have an event. Only two random numbers are needed per iteration in this routine, but the determination of which event will occur is more computationally intense as well as determining the total rate, since the table of events is longer. Updating for the next iteration can be optimized by only recalculating those entries in the table of events that refer to the changed species or, in the case of interacting species, to its neighbors. The total rate can be determined by subtracting the old rates at these entries and adding the new ones.

Montroll and Weiss designed an algorithm called continuous-time, random-walk algorithm to describe non-interacting random walkers on a surface[19]. Its flow diagram is depicted in Figure 10.5. The main difference between the previous two algorithms is that a separate time line for each individual random walker is determined. The time increment is therefore not determined by the total rate for the whole system, but the total rate for an individual random walker. The event times for each walker are then compared, together with the next deposition time, and the

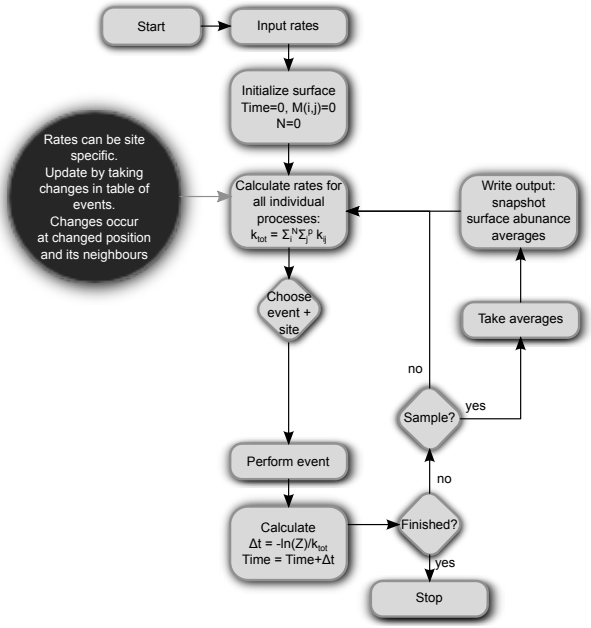


Figure 10.4: *Flow diagram of the n-fold way KMC algorithm[2].*

first event is then executed. Only one random number per iteration is needed; two in the case of deposition. In this way when updating the system for the next iteration only the random walker which moved in the previous iteration has to be considered. This random walker will get a new event time and will be placed in the species list which is sorted according to its event time. This algorithm works very well for non-interacting species. Obviously when species interact, a move of one random walker can affect the total rate of a neighboring random walker and, in turn, its event time.

10.5 Differences between Kinetic Monte Carlo and Metropolis Monte Carlo

As explained before, the kinetic Monte Carlo technique is a method to solve the master equation. One can show that the Metropolis Monte Carlo technique can solve this equation in equilibrium conditions, $\frac{\mathrm{d}}{\mathrm{d}t}P(x,t) = 0$. The master equation is therefore is the fundamental expression underlying both Monte Carlo methods. Let us now look at the differences between both Monte Carlo techniques.

Metropolis Monte Carlo calculates averages of systems in equilibrium and its algorithm is only interested in one trial event at each point in the simulation. The starting assumption for these type of simulations is that the system is in equilibrium, hence $\frac{\mathrm{d}}{\mathrm{d}t}P(x,t)$ is zero. The result is that time is not defined in this context, one uses probabilities instead of rates, and different probability schemes can be used, which has no effect on the outcome as long as detailed balance is obeyed. Kinetic Monte Carlo on the other hand uses rates which are crucial to determine the dynamics. The rates allow to simulate the evolution of a system in time, but the drawback is that the simulation runs from state to state and all accessible processes at a given cycle in the simulation need to be known. This difference in approach has a direct effect on the way the system is represented.

In kinetic Monte Carlo simulations, the system is often represented in terms of a *lattice-gas* model. Species are placed on a regular grid and only the occupancy of each grid point is registered. Each grid point can be seen as a local minimum on the potential energy surface. A clear advantage of a lattice-gas model is that one can work with a predefined event table. Since

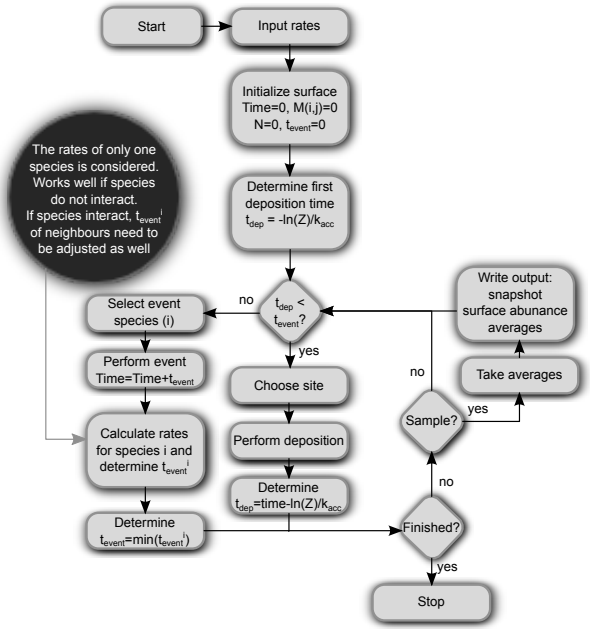


Figure 10.5: *Flow diagram of the continuous-time, random-walk KMC algorithm[19].*

Table 10.1: Summary of the differences between Metropolis and Kinetic Monte Carlo

	Metropolis MC	kinetic MC
Ensemble averaging	Yes	No
Time defined	No	Yes
State to state?	Not required	Required
Table of events?	Not required	Required
Equilibrium	Yes	Not required
Result depends on transition scheme	No	Yes

the molecules are confined to a lattice, only a limited amount of events is possible of which the rates can be determined at the start of the simulation. The limited amount of possible events in lattice-gas models allows one to cover very large timescales. The drawback is however that some of the molecular detail is lost in these models. One has to make guesses on the possible events which can be far from straightforward and one can easily miss important mechanisms. Furthermore, since the molecules are confined to lattice positions, using realistic molecular potentials is less straightforward.

In Metropolis Monte Carlo it is much more common to use an off-lattice model, where atoms can take every possible (x, y, z). The simulations do not run from state to state, *i.e.,* local minimum to local minimum, but the configurations of the system which are visited should follow a Boltzmann distribution in terms of (free) energy. This can be achieved by making a random position change (so not necessarily to a minimum on the potential energy surface) to obtain a trial move. This trial move is then accepted or not, depending *only* on the energy difference between the current and trial configuration. The differences between KMC and Metropolis MC are summarized in Table 10.1.

10.6 Off-lattice methods

One of the main restrictions of traditional KMC methods is the confinement of the system to a predefined lattice. The restriction of the atomic coordinates limits the amount of physical

detail contained by the simulation and makes the use of realistic interaction potentials far from straightforward. In the solid state, the lattice approximation is often well-justified for simple crystalline systems, but the situation already becomes more complicated for molecular crystals which often show some degree of disorder, like many hydrates and the hydrogen bond network in H_2O ices. Small site-to-site differences in these systems affect the rate constants and make the generation of the table of events a tedious task. For amorphous and liquid systems, the situation obviously becomes even worse.

A second problem with traditional KMC simulations is the need to define the table of events before the start of the simulation. As with the limitation to a lattice, this can be done for simple crystalline systems but quickly becomes more challenging as the complexity of the system increases and the transition mechanisms become less intuitive. Ideally, one would like to evaluate every process individually, as accurately as possible. This can be done in off-lattice KMC by applying force fields [20, 18, 15, 25] or even electronic structure methods [28] to calculate individual process rates. As every process is considered to be unique, this evaluation needs to be done on-the-fly, which considerably increases the computational costs. The advantage of not having any *a priori* assumptions about processes in the system is, however, often worth this additional effort. A few different approaches which take KMC simulations beyond the lattice approximation have been proposed. The continuum Kinetic Monte Carlo Method [29] is a particularly interesting off-lattice KMC method which incorporates diffusion of particles analytically into a continuous-time KMC scheme. Reaction rates and diffusion constants still need to be specified before the simulation starts though. The Adaptive Kinetic Monte Carlo [10] method is an on-the-fly, off-lattice KMC method which allows for very high atomistic detail, requiring no other input than an interatomic interaction potential. At each step in the KMC algorithm all possible events leaving the current state are obtained using single-ended transition path searches (Chapter 5) in combination with (harmonic) TST (Chapter 6).

Chapter 11

Free Energy Calculations

One of the aims of computer simulations is to predict first order phase transitions. A possible method to compute these are to have the system in one state and slowly heat the sample. However, since first order phase transitions usually have a huge barrier, hysteresis can occur. The temperature at which the transition would occur for heating would be much higher than the real transition temperature. Cooling the system would lead to a too low transition temperature. The height of the barrier for a first order phase transition is determined by the interfacial free energy between the two coexisting phases. One could remove the hysteresis by starting with a system which already has this interface between coexisting phases or by a method that does not require the formation of an interface all together.

One could study the relative stability by determining the free energy of the two phases. In a canonical ensemble, the Helmholtz free energy, $F \equiv E - TS$ is at a minimum in equilibrium. For an isothermal-isobaric ensemble, the Gibbs free energy $G \equiv F + PV$ is at a minimum. If we would wish to know which of two phases is more stable, we should simply compare the free energy of the two phases. A phase transition would boil down to a change in relative order of the free energy of the two phases. Remember how one would obtain the energy of system

$$\langle U \rangle = \frac{\int U(\mathbf{r}^N) \exp(-U(\mathbf{r}^N)/kT) \mathrm{d}\mathbf{r}^N}{Q}. \tag{11.1}$$

Using Metropolis Monte Carlo simulations we could solve this integral by averaging over U during the simulation

$$\langle U \rangle \approx \frac{1}{M} \sum_{i}^{M} U(\mathbf{r}_i). \tag{11.2}$$

By using importance Metropolis sampling, we could solve this integral without knowledge about Q which remains unknown. Now we need the Helmholtz free energy which is given by

$$F = -kT \ln Q(N,V,T) = -kT \ln \left(\frac{V^N \int \mathrm{d}\mathbf{p}^N \mathrm{d}\mathbf{r}^N \exp[-E(\mathbf{p}^N, \mathbf{r}^N)/kT]}{\Lambda^{dN} N!} \right). \tag{11.3}$$

This quantity depends directly on the partition function, Q, and the partition function cannot be directly obtained from a simulation (or experiment). The free energy can therefore not be obtained using an ensemble average.

One could reversibly change a system with known free energy to the system of interest by varying volume or temperature. During this change one could integrate the following equations

$$\left(\frac{\partial F}{\partial V} \right)_{NT} = -P \tag{11.4}$$

and

$$\left(\frac{\partial F/T}{\partial 1/T} \right)_{VN} = E. \tag{11.5}$$

For experiments an ideal gas is an obvious reference system. For simulations the reversible path does not have to be physical. One could for instance change from a Lennard-Jones fluid to a Stockmayer fluid which has an additional dipole-dipole term ($\propto r^{-3}$). This can be achieved by constructing the following energy function

$$U(\lambda) = (1-\lambda)U_{\mathrm{I}} + \lambda U_{\mathrm{II}} \tag{11.6}$$

where U_{I} is the Lennard-Jones potential and U_{II} the Stockmayer potential. By varying λ from 0 to 1, one can reversibly change between the two systems. The free energy difference between these systems is

$$\begin{aligned} F(\lambda=1) - F(\lambda=0) &= \int_0^1 \mathrm{d}\lambda \left(\frac{\partial F}{\partial \lambda}\right)_{N,V,T} = -\int_0^1 \mathrm{d}\lambda kT \frac{\partial}{\partial \lambda} \ln Q(N,V,T,\lambda) \\ &= -\int_0^1 \mathrm{d}\lambda \frac{kT}{Q} \frac{\partial Q(N,V,T,\lambda)}{\partial \lambda} \\ &= \int_0^1 \mathrm{d}\lambda \frac{\int \mathrm{d}\mathbf{r}^N (\frac{\partial E(\mathbf{r}^N,\lambda)}{\partial \lambda}) \exp(-E(\mathbf{r}^N,\lambda)/kT)}{Q} \\ &= \int_0^1 \mathrm{d}\lambda \left\langle \frac{\partial E(\mathbf{r}^N,\lambda)}{\partial \lambda} \right\rangle_\lambda \end{aligned} \tag{11.7}$$

We now have again an ensemble average of $\left\langle \frac{\partial E(\mathbf{r}^N,\lambda)}{\partial \lambda} \right\rangle_\lambda$ which we need to integrate over λ. In practice, we need in the order of 10 simulations at different values of λ to determine $\left\langle \frac{\partial E(\mathbf{r}^N,\lambda)}{\partial \lambda} \right\rangle_\lambda$ and evaluate this integral numerically. This method is obviously not restricted to the transition of a Lennard-Jones fluid to a Stockmayer fluid. Different reference systems have been suggested for different type of applications.

11.1 Einstein Integration

For solids, the ideal gas or the Lennard-Jones fluid is not a useful reference system, instead often the Einstein crystal is used. In this model, each atom in the lattice is an independent 3D quantum harmonic oscillator around a lattice position and all atoms oscillate with the same frequency. The energy of a particle attached is

$$\epsilon(r) = \frac{1}{2}\kappa r^2 \tag{11.8}$$

and

$$Q = \frac{1}{V^{\text{unit cell}}} \int_0^\infty \mathrm{d}r 4\pi r^2 \exp(-\epsilon(r)/kT) = \left(\frac{2\pi kT}{\kappa}\right)^{\frac{3}{2}} \tag{11.9}$$

Since Q is known also the free energy is known. To calculate the free energy of a crystal we need to make a reversible path from the Einstein crystal to the crystal of interest. This can again be done by introducing the order parameter λ

$$U(\lambda) = (1-\lambda)U_{\mathrm{Ein}} + \lambda U_{\mathrm{crys}} \tag{11.10}$$

We now integrate the system from $\lambda = 0$ to 1.

$$\Delta F = F_{\mathrm{crys}} - F_{\mathrm{Ein}} = \int_0^1 \mathrm{d}\lambda \left\langle \frac{\partial U(\mathbf{r}^N,\lambda)}{\partial \lambda} \right\rangle_\lambda = \int_0^1 \mathrm{d}\lambda \left\langle U_{\mathrm{Ein}} - U_{\mathrm{crys}} \right\rangle_\lambda \tag{11.11}$$

11.2 Widom test particle insertion method

The Widom test particle insertion method is a way to get the chemical potential of a particle. It is often applied to fluids. The chemical potential is given by

$$\mu_i = \left(\frac{\partial F}{\partial N_i}\right)_{V,T,N_{j\neq i}} \tag{11.12}$$

which we can estimate by

$$\mu = \frac{F(N+1,V,T) - F(N,V,T)}{N+1-N} = -kT \ln \frac{Q(N+1,V,T)}{Q(N,V,T)}. \tag{11.13}$$

The canonical partition is given by

$$Q(N,V,T) = \frac{V^N}{N!\Lambda^{3N}} \int \mathrm{d}\mathbf{r}^N \exp(-E(\mathbf{r}^N)/kT) \tag{11.14}$$

For reduced coordinates ($s_i = r_i/L; V = L^3$)

$$\begin{aligned}
\mu &= -kT \ln\left(\frac{V}{\Lambda^3(N+1)}\right) - kT \ln\left(\frac{\int \mathrm{d}\mathbf{s}^{N+1} \exp[-E(\mathbf{s}^{N+1})/kT]}{\int \mathrm{d}\mathbf{s}^N \exp[-E(\mathbf{s}^N)/kT]}\right) \\
&= -kT \ln\left(\frac{V}{\Lambda^3(N+1)}\right) - kT \ln\left(\frac{\int \mathrm{d}\mathbf{s}^{N+1} \exp[-E(\mathbf{s}^N)/kT] \exp[-\Delta E^+/kT]}{\int \mathrm{d}\mathbf{s}^N \exp[-E(\mathbf{s}^N)/kT]}\right) \\
&\approx kT \ln(\rho\Lambda^3) - kT \ln\left\langle \exp[-\Delta E^+/kT]\right\rangle_{N,V,T} \\
&= \mu_{\mathrm{IG}} + \mu_{\mathrm{ex}}
\end{aligned}$$

We should keep in mind that the ensemble average

$$\left\langle \exp[-\Delta E^+/kT]\right\rangle_{N,V,T} \tag{11.15}$$

is an unbiased average over all possible positions of the test particle and a Boltzmann average for our system of N particles at volume $V = L^3$ and temperature T. In practice, we simulate a system of N particles in volume V at temperature T using the conventional Monte Carlo algorithm. During this simulation, we keep track of the ensemble average of $\exp[-\Delta E^+/kT]$, in which ΔE^+ is the energy change when an additional particle is inserted into the system at a random position (without ever accepting such an insertion). Widom test particle method is not only possible in the NVT ensemble, but also in the Gibbs ensemble and the NPT ensemble. The resulting expression for the NPT ensemble is

$$\mu(P) = kT \ln(\rho P \Lambda^3) - kT \ln\left\langle \frac{PV}{kT(N+1)} \exp[-\Delta E^+/kT]\right\rangle_{N,P,T} \tag{11.16}$$

Bibliography

[1] Bolhuis, P. G., D. Chandler, C. Dellago, and P. L. Geissler (2002, October). TRANSITION PATH SAMPLING: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Ann. Rev. Phys. Chem. 53*, 291–318.

[2] Bortz, A. B., M. H. Kalos, and J. L. Lebowitz (1975, January). A New Algorithm for Monte Carlo Simulation of Ising Spin Systems. *J. Comp. Phys. 17*, 10.

[3] Dellago, C., P. G. Bolhuis, and D. Chandler (1998, June). Efficient transition path sampling: Application to Lennard-Jones cluster rearrangements. *J. Chem. Phys. 108*, 9236–9245.

[4] Evans, M. G. and M. Polanyi (1935). Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Trans. Faraday Soc. 31*, 875–894.

[5] Eyring, H. (1935, February). The Activated Complex in Chemical Reactions. *J. Chem. Phys. 3*, 107–115.

[6] Gardiner, C. W. (1994). *Handbook of stochastic methods for physics, chemistry and the natural sciences*. Berlin: Springer.

[7] Gillespie, D. T. (1976, December). A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *J. Comp. Phys. 22*, 403.

[8] Gillespie, D. T. (1992). *Markov Processes: An Introduction for Physical Scientists*. London: Academic Press Limited.

[9] Henkelman, G. and H. Jónsson (1999, October). A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J. Chem. Phys. 111*, 7010–7022.

[10] Henkelman, G. and H. Jónsson (2001, December). Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table. *J. Chem. Phys. 115*, 9657–9666.

[11] Henkelman, G., B. P. Uberuaga, and H. Jónsson (2000, December). A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys. 113*, 9901–9904.

[12] Heyden, A., A. T. Bell, and F. J. Keil (2005, December). Efficient methods for finding transition states in chemical reactions: Comparison of improved dimer method and partitioned rational function optimization method. *J. Chem. Phys. 123*(22), 224101.

[13] Jansen, A. P. J. (1995). *Comp. Phys. Comm. 86*, 1.

[14] Jónsson, H., G. Mills, and K. Jacobsen (1998). *Nudged elastic band method for finding minimum energy paths of transitions*. Singapore: World Scientific.

[15] Karssemeijer, L. J., A. Pedersen, H. Jónsson, and H. M. Cuppen (2012). Long-timescale simulations of diffusion in molecular solids. *Phys. Chem. Chem. Phys. 14*, 10844–10852.

[16] Malek, R. and N. Mousseau (2000, December). Dynamics of Lennard-Jones clusters: A characterization of the activation-relaxation technique. *Phys. Rev. E 62*, 7723–7728.

[17] Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953, June). Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys. 21*, 1087–1092.

[18] Middleton, T. F. and D. J. Wales (2004, May). Comparison of kinetic Monte Carlo and molecular dynamics simulations of diffusion in a model glass former. *J. Chem. Phys. 120*, 8134–8143.

[19] Montroll, E. W. and G. H. Weiss (1965, February). Random Walks on Lattices. II. *J. Math. Phys. 6*, 167–181.

[20] Much, F., M. Ahr, M. Biehl, and W. Kinzel (2002, August). A Kinetic Monte Carlo method for the simulation of heteroepitaxial growth. *Comp. Phys. Comm. 147*(1-2), 226–229.

[21] Munro, L. J. and D. J. Wales (1999, February). Defect migration in crystalline silicon. *Phys. Rev. B 59*, 3969–3980.

[22] Nordmeyer, T. and F. Zaera (1991, August). A Monte Carlo simulation for the precursor model of adsorption on a spatially homogeneous surface. *Chem. Phys. Lett. 183*, 195–198.

[23] Nordmeyer, T. and F. Zaera (1992, December). A theoretical study of the parameters affecting the kinetics of gas adsorption on solid surfaces. *J. Chem. Phys. 97*, 9345–9354.

[24] Olsen, R. A., G. J. Kroes, G. Henkelman, A. Arnaldsson, and H. Jónsson (2004, November). Comparison of methods for finding saddle points without knowledge of the final states. *J. Chem. Phys. 121*, 9776–9792.

[25] Pedersen, A., G. Henkelman, J. Schiøtz, and H. Jónsson (2009, July). Long time scale simulation of a grain boundary in copper. *New J. Phys. 11*(7), 073034.

[26] Pelzer, H. and E. Wigner (1932). *Z. Phys. Chem. Abt. B 15*, 445.

[27] van Kampen, N. G. (1992). *Stochastic Processes in Physics and Chemistry*. Amsterdam: Elsevier Science.

[28] Xu, L. and G. Henkelman (2008, September). Adaptive kinetic Monte Carlo for first-principles accelerated dynamics. *J. Chem. Phys. 129*(11), 114104.

[29] Zhang, X.-Q. and A. P. J. Jansen (2010, October). Kinetic Monte Carlo method for simulating reactions in solutions. *Phys. Rev. E 82*(4), 046704.