1

Introduction

This thesis provides the essential information extracted out of XXX s computer simulation of molecules. If one includes the numerous explorative or discarded simulations, it sums up to XXX s, which required XXX days of computer processing time. The information offered by such simulations is the detailed picuture of the molecular system of interest with a resolution on the fs time scale $(1\,\text{fs} = 10^{-18}\,\text{s} = 1/1\,000\,000\,000\,000\,000\,\text{s})$ and on the nm length scale $(1\,\text{nm} = 10^{-9}\,\text{m} = 1/1\,000\,000\,000\,\text{m})$. Mechanistic pathways and favored configurational states of the molecules of interest can be infered from this detailed picture, which is often not accessible by experimental methods. Therefore computer simulation is an important tool to complement experiments with additional high-resolution information or even to predict experimental outcomes, thereby saving time and resources for experiments to be performed. XXX In the following, the history of models in chemistry is introduced before providing the theoretical basis and methods for computer simultions. The theoretical basis is given by classical mechanics and statistical thermodynamics whereas the methods include force-fields and algorithms governing the simulations.

1.1   HISTORY OF MOLECULAR MODELING

Since ancient times humans have seeked to understand their environment and the observed phenomena in the very same. They have invented models for the composition of the

surrounding matter and ideally describe its behaviour. The starting point were century-long disputations between representatives of atomistic and continuity theories. The atomistic theories were founded in the 5th century BC,[?] by Presocratics Leucippus and Democritus. They argued that matter is not infinitely divisible, but that there are least particles which are inalterable and indivible - the atoms. It was Democritus, who claimed[?][?]

> "by convention sweet and by convention bitter, by convention hot, by convention cold, by convention color; but in reality atoms and void".

This ground-breaking statement was far ahead of its time and was mostly rejected, especially by Aristotle with his work *De caelo*, stating that matter is continous, constisting out of five elements.[REF] As the church was adhered to the continuity theory of Aristotle, it was predominant in Europe until the greek atomisim experienced a revival with the rise of the mechanical atomism in the 17th century by philosophers like René Descartes, Pierre Gassendi and Robert Boyle.[?]

While Newton led the foundations of classical mechanics in his work *Principia*,[1] he also shared atomistic views as he stated that[?]

> the least parts of bodies to be - all extended, and hard and impenetrable, and moveable, and endowed with their proper inertia.

Before, the atomists were uncertain about the laws govering the movements of atoms. With Newton's three laws of motion, the dynamics of atoms could be in principle determined. The limitation, however, was that all the forces between the atoms need to be known.[?]

The early mechanical atomism was then superseded by the atomic theory of John Dalton in the early 19th century.[?] For the first time properties as the relative weight could be assigned to atoms. Chemical elements were already known, but now atoms could be ascibed to least units of elements. Likewise chemical compounds were found to be specific combinations of atoms.

The word "chemical structure" was termed in the mid 19th century by chemists like Archibald Scott Couper, Friedrich August Kekulé and Alexander Mikhailovich But-

Ierov.[REF] They have proposed the first chemical structures, thereby developing important concepts like valency, chemical bonds and substituents.[REF] Especially Kekulé played an important role in the structural eludication of the benzene ring, one of his hypotheses is illustrated in Fig. **??**.[REF] At about the same time, Johann Josef Loschmidt published a famous collection of 368 molecular structures, an example is given in Fig. **??**.[REF] His graphics included the spatial extent of different atoms in amazingly correct proportions.[REF]

The third spatial dimension came into play with August Wilhelm Hofmann, who created three-dimensional models of methane, chloroform and other small organic molecules with table crocket balls.[REF] Although the structures were not agreement with nowadays state of knowledge, he still contributed the color scheme for atoms (*e.g.* black for carbon, white for hydrogen) which is used until today. In the following years such models could be refined and improved with the advances of the theory and experimental techniques. Structural representations of larger and more complex (bio)molecules could be created.[REF Glowacki] Notable mile stones were the model of penicillin based on X-ray crystallographic data by Hodgkin et al, , the model of α-helixes by Pauling et al.and the famous DNA model by Crick and Watson.

FIGURE 1.1: *Examples of first molecular models.* Graph (a) shows a methane molecular model, created by August Wilhelm von Hofmann around 1860 and now part of the collection of the Royal Institution of London. *With kind permission of Henry Rzepa.*Graph (b) shows the first model of a complete protein, created and published 1958 by Kendrew et al.[REF]

Finally in 1958, the first and still coarse model of a complete protein was accomplished by Kendrew et al.[REF]It was a model of myoglobine made out of paint, plasticine and wood and is shown in Fig. **??**. A model of the structurally similar haemoglobine was

created by Perutz et al.[REF].

The advent of digital computers in the 1940s opened entirely new possibilities of building molecular models. Computers were not only used solve X-ray structures of the proteins mentioned above, but also to solve the equations of the physical models which were not solvable by hand anymore. Next to quantum chemistry these were classical mechanics simulations.[REF] After first monte carlo simulations of rigid spheres in 2 dimensions,[2] and molecular dynamics of hard spheres[3] in the 1950s, the first Lennard-Jones fluid was simulated in 1964.[4] Computer simulations added a completely new, but essential aspect to the molecular models: The alteration of the structure with time, resulting in vibrations, movements and conformational changes of molecules which allow reactions and are related to properties of chemical compounds. This crucial importance of the movements of atoms and molecules for their function and ultimately life was fameously phrased by Feynman who stated[5]

> "that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms."

Thus the new field of molecular simulations was born to contribute to the aim of understanding life in all its details on the atomistic level. Polyatomic molecules like of water[6] and proteins[7] were simulated shortly thereafter. In the following decades progress was made in developping force fields (interaction functions) and algortihms for reliable, efficient and accurate molecular dynamics simulations. Since the 1990s, it became an indispensable tool in many fields of science like biology, pharmaceutics, material science and chemistry. Due to the increasing computer power, larger systems are computable which increases the scope of applications. It not only helps vizualizing molecules for a better understanding, but is also a mean for calculating properties and replacing experiments.

1.2   CLASSICAL MECHANICS

The ground for classical mechanics was set by Newton,[1] and later reformulated by Lagrange and Hamilton.[REF] The theory of classical mechanics is not valid when dealing with high energies (*e.g.* high velocities, light-matter interactions) or low masses (*e.g.* electrons, elementary particles). In these cases, quantum mechanics and relativistic treatment of the system is advisable. However, in the opposite case of low energies and high masses, classical mechanics provides the accuracy usually needed while being simpler to process and to compute.

Classical systems are usually described by N point particles $i = 0, 1, \ldots, N-1$ with mass m_i in the Cartesian space, *i.e.* in x, y and z coordinates. Each particle are at a certain position $\mathbf{r}_i = \{x_i, y_i, z_i\}$ with a momentum $\mathbf{p}_i = m_i \mathbf{v}_i = m_i\{v_{x,i}, v_{y,i}, v_{z,i}\}$ where $\mathbf{v}$ is the velocity. Vectors $\mathbf{r}$ and $\mathbf{p}$ are be united in a position-momentum vector $\mathbf{x}_i = \{\mathbf{r}_i, \mathbf{p}_i\}$. The vectors $\mathbf{x}$ of all particles are the elements of the phase space vector $\mathbf{X} = \{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{N-1}\}$ which contains all information about the state of a system. The phase space is the associated $6N$-dimensional space which includes all possible states of $\mathbf{X}$, *i.e.* it contains all possible combinations of different positions and momenta of all particles i.

The associated energy is given by the so-called Hamilonian, which is a function of the phase space vector $\mathbf{X}$, and can be separated into a kinetic energy term $\mathcal{K}$ and a potential energy term $\mathcal{V}$ of the system,

$$\mathcal{H}(\mathbf{X}) = \mathcal{K}(\mathbf{p}) + \mathcal{V}(\mathbf{r}). \tag{1.1}$$

The kinetic energy is given by

$$\mathcal{K}(\mathbf{p}) = \sum_i \frac{\mathbf{p}_i^2}{2m_i}, \tag{1.2}$$

which is the sum over the kinetic energies of all particles. The potential energy term is (approximative) mathematical expression for the interaction of the particles with each

other and possibly with the environment. It can have different functional forms, depending on the nature of the system and the degree of approximation which is typically chosen according to the desired level of accuracy.

The time evolution of the system in Hamiltonian mechanics is given by the two equations. The first equation describes how the positions evolve in time,

$$\dot{\mathbf{r}}_i = \frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t} = \frac{\partial \mathcal{H}}{\partial \mathbf{p}_i}. \tag{1.3}$$

By inserting the definition of the kinetic energy, Eq. 1.2, one obtains $\dot{\mathbf{r}}_i = \partial \mathcal{H}/\partial \mathbf{p}_i = \partial \mathcal{K}/\partial \mathbf{p}_i = \mathbf{p}_i/m_i = \mathbf{v}_i$, which tells us that the positions change according to the velocities of the particles. The second equation describes the time evolution of the momenta,

$$\dot{\mathbf{p}}_i = \frac{\mathrm{d}\mathbf{p}_i}{\mathrm{d}t} = -\frac{\partial \mathcal{H}}{\partial \mathbf{r}_i}, \tag{1.4}$$

The time derivative of the momentum of particle i is equal to its acceleration a_i scaled by its mass, $\dot{\mathbf{p}}_i = m_i \dot{\mathbf{v}}_i = m_i a_i$. Therefore the second equation is another formulation of Newton's famous second law $\mathbf{f}_i = m_i \mathbf{a}_i$, where $\mathbf{f}_i$ is the force, given by the negative derivative of the Hamiltonian function with respect to the position $\mathbf{f}_i = -\frac{\partial \mathcal{H}}{\partial \mathbf{r}_i}$.

It is important to note, that the total energy of a system governed by Eqs. 1.3 and 1.4 is constant, *i.e.* $\mathcal{H}$ in Hamiltonian mechanics is a constant of motion.

1.3   CLASSICAL STATISTICAL MECHANICS

Statistical mechanics[8–12] is a branch of physics where properties of a macrosopic system are linked to the microscopic behaviour of the system. The microscopic behaviour of a macroscopic system is governed by the way the particles of the system interact, *i.e.* by their interaction function $\mathcal{V}$ in Eq. 1.1. In classical statistical mechanics, the interaction functions are formulated in the framework of classical mechanics and are part of the Hamiltonian of the system. Because of the immense size of macroscopic systems ($1\,\mathrm{mol} = 6.022^{23}$ particles) and the nontrivial microscopic interactions between particles

6

in real systems, an analytical treatment according to classical mechanics to calculate the properties of interest is not feasible. However, by statistically evaluating the microscopic details, we can still calculate the macroscopic properties of a system.

The statistical evaluation is performed on the basis of *ensembles*, which are a collection of microscopic states (*microstates*) of the same system obeying the same microscopic interactions and fulfill certain macroscopic boundary conditions. The macroscopic property of interest can be calculated by averaging over the systems of the ensemble. According to the ergodic hypothesis, [13] this property can also be calculated by observing the evolution of one system over a sufficient period of time. A microstate is defined by a point $\mathbf{x} = \{\mathbf{r}, \mathbf{p}\}$ in the phase space of the system, which describes all the coordinates and momenta of the particles in the system. Typically, a system treated by statistical mechanics has many degrees of freedom, *i.e.*for high-dimensional $\mathbf{x}$. Ensembles encompass all microstates which obey macroscopic boundary conditions, which are usually three constant thermodynamic observables, among them at least one extensive quantity. Common ensembles are the microcanonical ensemble with constant number N of particles, volume V and energy E, the canonical ensemble with constant number N of particles, volume V and temperature T, the isothermal-isobaric (Gibbs) ensemble with constant number N of particles, pressure P and temperature T and the grand canoncial ensemble with constant chemical potential μ, volume V and temperature T. In the following, we consider the microcanonical and the canonical ensemble, but the discussion can be straighforwardly extended to other systems.

1.3.1 THE MICROCANONICAL ENSEMBLE

The microcanonical ensemble is the simplest ensemble, that of an isolated system with a constant number N of particles in a container of constant volume V and at constant energy E.

To be part of the ensemble, the microstates defined by the phase space vector $\mathbf{x}$ have to fulfill the condition

$$\mathcal{H}(\mathbf{x}) = E. \tag{1.5}$$

All accessbile microstates $\mathbf{x}$ of the ensemble lie on a constant-energy hypersurface in the phase space and are equally probable. This is known as the *assumption of equal a priori probability*.

A measure of the amount of phase space available to the system is the *partition function*. The microcanonical partition function Ω is calculated as an integral over the phase space

$$\Omega(N, V, E) = \frac{E_0}{N! h^{3N}} \int \mathrm{d}\mathbf{p} \mathrm{d}\mathbf{r} \delta(\mathcal{H}(\mathbf{r}, \mathbf{p}) - E) \tag{1.6}$$

$$= \frac{E_0}{N! h^{3N}} \int \mathrm{d}\mathbf{x} \delta(\mathcal{H}(\mathbf{x}) - E), \tag{1.7}$$

where the prefactor $E_0 N!^{-1} h^{-3N}$ accounts for indistinguishable particles $(N!^{-1})$, unit conversion to a dimensionless quantity (E_0) and Heisenberg's uncertainty relation (h^{-3N}). The partition function is the fundamental quantity in statistical mechanics. All thermodynamic properties of the system can be derived from it.

The entropy S of a system can be related to the number of microstates and therefore the partition function via the Boltzmann relation

$$S(N, V, E) = k_{\mathrm{B}} \ln \Omega(N, V, E), \tag{1.8}$$

where the proportionality factor $k_{\mathrm{B}} = 8.3146 \, \mathrm{J} \, \mathrm{K}^{-1} \mathrm{mol}^{-1}$ is the Boltzmann factor. When taking the total differential of the entropy $S(N, V, E)$,

$$\mathrm{d}S = \left(\frac{\partial S}{\partial E}\right)_{N,V} \mathrm{d}E + \left(\frac{\partial S}{\partial V}\right)_{N,E} \mathrm{d}V + \left(\frac{\partial S}{\partial N}\right)_{V,E} \mathrm{d}N$$

$$= \frac{1}{T} \mathrm{d}E + \frac{P}{T} \mathrm{d}V - \frac{\mu}{T} \mathrm{d}N, \tag{1.9}$$

and equating the coefficients,

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_{N,V}, \qquad \frac{P}{T} = \left(\frac{\partial S}{\partial V}\right)_{N,E}, \qquad \frac{\mu}{T} = \left(\frac{\partial S}{\partial N}\right)_{V,E}, \tag{1.10}$$

one can relate the thermodynamic variables T, P, and μ to $\Omega(N, V, E)$ by combining

Eq. **??** with Eq. 1.10:

$$\frac{1}{k_{\mathrm{B}}T} = \left(\frac{\partial \ln \Omega}{\partial E}\right)_{N,V}, \quad \frac{P}{k_{\mathrm{B}}T} = \left(\frac{\partial \ln \Omega}{\partial V}\right)_{N,E}, \quad \frac{\mu}{k_{\mathrm{B}}T} = \left(\frac{\partial \ln \Omega}{\partial N}\right)_{V,E}. \quad (1.11)$$

The normalized probability $P(\mathbf{x})$ of finding a system in a given microstate $\mathbf{x}$ can also be related to the inverse partition function

$$P(\mathbf{x}) = \frac{E_0}{N!h^{3N}} \frac{\delta(\mathcal{H}(\mathbf{x} - E)}{\Omega(N,V,E)}. \quad (1.12)$$

The normalized probability defines the distribution of states in equilibrium. Accordingly, any thermodynamic observable $Y(N,V,E)$ can be calculated by multiplying the instantaneous quantity $\mathcal{Y}(\mathbf{x})$ of a microstate with its probability and integrating over the whole phase space, *i.e.*

$$\begin{aligned} Y(N,V,E) &= \langle \mathcal{Y}(\mathbf{x}) \rangle \\ &= \int \mathrm{d}\mathbf{x}\, P(\mathbf{x})\mathcal{Y}(\mathbf{x}) \\ &= \frac{E_0}{N!h^{3N}\Omega(N,V,E)} \int \mathrm{d}\mathbf{x}\, \mathcal{Y}(\mathbf{x})\delta(\mathcal{H}(\mathbf{x} - E)). \end{aligned} \quad (1.13)$$

While being the simplest ensemble, the use of the microcanocial ensemble is very limited because it is very difficult to realize the the condition of a constant total energy in reality. In the following, the canonical ensemble is presented which has more practical use.

1.3.2 THE CANONICAL ENSEMBLE

The canonical ensemble is an ensemble of systems with a constant number N particles in a container of constant volume V at temperature T, *i.e.* the systems are in contact with a heat bath allowing the exchange of energy in order to keep the temperature of the system constant.

The main difference to the microcanonical ensemble is, that, due to the heat exchange with the environment, the total energy is not conserved. The total energy of the systems

obeys a Boltzmann distribution, $\exp(-\beta\mathcal{H}(\mathbf{x}))$. The canonical partition function is an integral over the Boltzmann factor of all possible microstates

$$Q(N, V, T) = \frac{1}{h^{3N} N!} \int d\mathbf{p} \int d\mathbf{r} \exp(-\beta\mathcal{H}(\mathbf{p}, \mathbf{r})) \tag{1.14}$$

where $\beta = (k_{\mathrm{B}}T)^{-1}$, k_{B} is the Boltzmann's constant, and h is Planck's constant. $\mathcal{H}$ is the classical Hamiltonian of the system, which is a sum of the kinetic energy $\mathcal{K}$ and potential energy $\mathcal{U}$ of the system

$$\mathcal{H}(\mathbf{r}, \mathbf{p}) = \mathcal{U}(\mathbf{r}) + \mathcal{K}(\mathbf{p}). \tag{1.15}$$

The total energy of the ensemble is now the ensemble averaged Hamiltonian energy and is now related to the partition function *via*

$$E(N, V, T) = \langle\mathcal{H}(\mathbf{r}, \mathbf{p})\rangle. \tag{1.16}$$

$$= \int d\mathbf{p} \int d\mathbf{r}\, P(\mathbf{r}, \mathbf{p})\mathcal{H}(\mathbf{r}, \mathbf{p}) \tag{1.17}$$

$$= -\frac{1}{Q(N, V, T)} \frac{\partial}{\partial\beta} Q(N, V, T) \tag{1.18}$$

$$= -\frac{\partial}{\partial\beta} \ln Q(N, V, T), \tag{1.19}$$

$$\tag{1.20}$$

The entropy is defined by

$$S(N, V, T) = C - k_{\mathrm{B}} \int d\mathbf{p} \int d\mathbf{r}\, P(\mathbf{r}, \mathbf{p}) \ln P(\mathbf{r}, \mathbf{p}) \tag{1.21}$$

It is convenient to define the Helmholtz free energy A as

$$A(N, V, T) = -k_{\mathrm{B}}T \ln Q(N, V, T) \tag{1.22}$$

$$= -\beta^{-1} \ln Q(N, V, T) \tag{1.23}$$

1.4   MOLECULAR DYNAMICS SIMULATION

Molecular dynamics (MD) simulations apply the classical mechanics (Sect. **??**) to molecular systems which are usually organic or biological molecules in solvent. In the following, the framework of MD is briefly introduced.

1.4.1 THE SYSTEM

The molecular system to be simulated typically consists out of a solute(s), *e.g.* an organic molecule, a protein, a saccharide, a nucleic acid or a combination of those. The environment of this solute is either not modeled (vacuum), as implicit solvent (the average influence of solvent molecules is taken into account by a continuous medium) or as explicit solvent (solvent molecules are explicitly added to the system). The number of atoms can be on the order of several $100\,000$ of atoms, which leads to system sizes on the order of $10\,\text{nm}$. The times simulated are on the order of nanoseconds to milliseconds per day computer time, depending next to the available computing resources on the size of the system, the approximations done and the algorithms used.

The corresponding atoms to be simulated are described as N point particles having masses $\{m_i\}$ and charges $\{q_i\}$. Often, a number of atoms are combined to one point particle, which is referred to as united atom[14] (*e.g.* a methyl group) or coarse graining bead[REF] (*e.g.* a whole ester functional group). The electron and other elementary particle degrees of freedom are neglected, only their average influence is modeled by means of the masses and charges. Therefore, MD simulations cannot describe electronically excited states, which includes chemical reactions or light-matter interaction. In the case of explicit solvation, periodic boundary conditions are usually applied to avoid surface effects. The system is modeled in a space-filling polyhedron, *e.g.* a cube or a octahedron and an atom leaving the boundary on one side enters again on the other side. Additionally, the system is allowed to interact with copies of itself, representing continuous matter.

1.4.2 The Interaction Function (Force Field)

The potential energy $\mathcal{V}$ (Sect. **??**) is a mathematical expression to describe the nature of the system. In MD, this function is referred to as force field, which defines the interaction between the point particles in the system. The force field consists of many terms, which can be classified as covalent and non-bonding terms.[15] Most of them are two-body terms, *i.e.* they include only the coordinates of two point particles, some are three- and four body terms and higher-body terms can be included in exceptional cases.

The covalent terms describe the interactions within one molecule up do third neighbor interactions, *i.e.* between atoms separated by up to three chemical bonds. They include usually bond stretching terms, which are harmonic potentials controlling the distance between two atoms; angle bending terms, which are harmonic potentials controlling the angle spanned by three atoms; torsional dihedral terms, which are usually cosine series controlling the configuration of four consecutively bonded atoms; and improper dihedral terms, which are harmonic potentials controlling the configuration of four atoms of which three are bonded to a central atom. Other terms are possible, but seldomly used.

Interactions between atoms separated by more than two bonds (including pairs of atoms of different molecules) are described by non-bonded interaction terms. These terms are usually only two-body terms, depending on the distance between the two atoms. There are typically two different terms considered which describe the electrostatic and van-der-Waals interactions between the atoms. The electrostatic interactions are defined as Coulomb potential terms having a $1/r$ dependence where r is the distance of the atoms. These terms model the attraction of two atoms with opposite sign partial charges or the repulsion of atoms having both positive or both negative charges. The van-der-Waals interactions have the Lennard-Jones functional form modeling the repulsion due to the Pauli exclusion principle for the electrons for short distances (r^{-12} dependence) and attraction due to London dispersion forces (r^{-6} dependece) for longer distances. Because non-bonded interactions, especially the electrostatic interactions, are non-negligible at distances longer than the system sizes, special care is taken for long-range non-bonded interactions to avoid self-interaction between copies of the same atoms and to save

computation time. Most prominent methods for this are the reaction field[16] (RF) and the particle-mesh Ewald[REF] (PME) methods. In the RF method, a cutoff is applied which sets all the interaction functions to zero at distances larger than the cutoff distance. By only cutting off the electrostatic interactions, one would introduce discontinuities in the potential energy and therefore destroy the conservation of total energy. To avoid this, reaction-field term is added correcting for the missing dielectric environment at longer ranges. The PME ...XXX. Since the covalent terms take care of interactions between atoms separated by up to three bonds, the covalent interactions for first and second neighbours are usually excluded and scaled or special parameters are used for third neighbour non-bonded interactions.

All the force-field terms extensively make use of parameters like force constants, reference bond lengths, angles, torsions and improper dihedrals, partial charges and Lennard-Jones parameters. The number of parameters is on the order of the degrees of freedom of the parameterized molecules. To get an idea about this, a water (H_2O) molecule is in principle described by 10 parameters, including 4 Lennard-Jones parameters (2 for the oxygen, 2 for the hydrogen, the latter sometimes set to zero), 2 partial charges, two reference values and two force contants for the bond lengths and angle, respectively. These parameters have to be fitted to the molecular system in order to reproduce experimental properties in a MD simulation. Initial guesses of parmeters like force constants (energetic information) are taken from experimental spectroscopic measurements. For reference structural information (bond lengths, angles, torsions and improper dihedrals), X-ray crystallography and nuclear magnetic resonance is of great help. All experimentally derivable parameters can be refined or even completely replaced by quantum-mechanical (QM) calculations. However, the initial parameters taken from experiments or calculations have to be refined to make them compatible to each other and to the employed algortihms. Therefore force-field parameterization is a time consuming, difficult task, which is still the subject of ongoing research in the MD community.[REF]

1.4.3 Integration of the Equations of Motion

After having a force field describing the interaction of the particles and defining the Hamiltonian $\mathcal{H}$ (Eq. 1.1) of the system, the classical equations of motion Eqs. 1.3 and 1.4 can be integrated. In view of the large size of the system ($6N$ dimensions), the analytical integration of the equations of motion is impossible and has to be performed numerically instead. One employs a finite time step Δt, which has to be short enough to avoid quadrature errors. It is typically chosen at the order of $1/10$ of the shortest time scale of the system, given by bond vibrations with a femtoseconds time scale. Different algorithms aiming at reducing the integration error have been developed. One example is the leap-frog algorithm[17] which evolves the positions $\mathbf{r}_i$ and velocities $\mathbf{v}_i$ shifted by a half timestep:

$$\mathbf{v}_i(t + \frac{\Delta t}{2}) = \mathbf{v}_i(t - \frac{\Delta t}{2}) + \frac{\mathbf{f}_i(\mathbf{r}(t))}{m_i}\Delta t \tag{1.24}$$

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t + \frac{\Delta t}{2})\Delta t. \tag{1.25}$$

This leads to a vanishing Δt^2 term and therefore to a reduced error on the order of Δt^3.

1.4.4 Free-Energy Calculations

One may calculate thermodynamic properties from molecular dynamics simulations in terms of simple averages of the instantaneous quantities thereby approximating the full thermodynamic ensemble, *i.e.* the integrals in Eq. **??** by a sum over a set of frames obtained from a simulation. Although we can express Eq. **??** of the free energy as an ensemble average by noting that $Q(N, V, T) = \langle \exp(\beta\mathcal{H}(\mathbf{X})) \rangle$,

$$A(N, V, T) = -k_{\mathrm{B}}T\langle \exp(\beta\mathcal{H}(\mathbf{X})) \rangle \tag{1.26}$$

This approach, however, is not eligible for the calculation of properties where no corresponding instantaneous quantities exist like the free-energy or the entropy. These quantities are inherent properties of the whole ensemble, not of single frames or configura-

tions. A lot of effort has been put in developing methods for calculating these properties from MD simulations. The approaches are quickly summarized in the following.

1.5   AIM OF THE THESIS

In Chapter **??**, XXX. In Chapter **??**, XXX. In Chapter **??**, XXX. In Chapter **??**, XXX.

References

[1] I. Newton, *The Principia: Mathematical principles of natural philosophy. (A new translation by I. B. Cohen and A. Whitman)*, University of California Press, Berkeley, USA (1999).

[2] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, *Equation of state calculations by fast computing machines*, J. Chem. Phys. 21 (1953), pp. 1087-1092.

[3] B.J. Alder and T.E. Wainwright, *Phase transition for a hard sphere system*, J. Chem. Phys. 27 (1957), pp. 1208-1209.

[4] A. Rahman, *Correlations in the motion of atoms in liquid argon*, Phys. Rev. 136 (1964), pp. A405-A411.

[5] R.P. Feynman, R.B. Leighton and M. Sands, *The Feynman lectures on physics Volume 1.*, Addison-Wesley, Boston, USA (1963).

[6] F.H. Stillinger and A. Rahman, *Improved simulation of liquid water by molecular dynamics*, J. Chem. Phys. 60 (1974), pp. 1545-1557.

[7] J.A. McCammon, B.R. Gelin and M. Karplus, *Dynamics of folded proteins*, Nature 267 (1977), pp. 585-590.

[8] H.L. Friedman, *A course in statistical mechanic*, Prentice Hall, Inc., Englewood Cliffs, New Jersey, USA (1985).

[9] R. Kubo, *Statistical Mechanics. An Advanced Course with Problems and Solutions Edition 8.*, Elsevier Science Publisher B.V., Amsterdam, The Netherlands (1990).

[10] D.A. McQuarrie, *Statistical mechanics*, University Science Books, Sausalito, CA, USA (2000).

[11] L.K. Nash, *Elements of statistical Thermodynamic Edition 2.*, Dover Publications Inc., Mineola, New York, USA (2006).

[12] K. Huang, *Introduction to statistical physics Edition 2nd ed..*, Chapman & Hall/CRC Taylor & Francis Group, Boca Raton, USA (2010).

[13] G. Gallavotti, *Ergodicity: a historical perspective. Equilibrium and nonequilibrium*, Eur. Phys. J. H 41 (2016), pp. 181-259.

[14] X. Daura, A.E. Mark and W.F. van Gunsteren, *Parametrization of aliphatic CH_n united atoms of GROMOS96 force field*, J. Comput. Chem. 19 (1998), pp. 535-547.

[15] C. Oostenbrink, A. Villa, A.E. Mark and W.F. van Gunsteren, *A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6*, J. Comput. Chem. 25 (2004), pp. 1656-1676.

[16] H.J.C. Berendsen, W.F. van Gunsteren, H.R.J Zwinderman and R.G. Geurtsen, *Simulations of proteins in water*, Ann. New York Acad. Sci. 482 (1986), pp. 269-285.

[17] R.W. Hockney, *The potential calculation and some applications*, Methods Comput. Phys. 9 (1970), pp. 135-211.