Morphology and charge transport in conjugated polymers

Dissertation

zur Erlangung des Grades

"Doktor der Naturwissenschaften"

am Fachbereich Physik

der Johannes Gutenberg-Universität Mainz

Victor Rühle

geb. in Böblingen

Max-Planck-Institut für Polymerforschung

Mainz, Juni 2010

Vorsitzender:

1. Berichterstatter:
2. Berichterstatter:

Tag der mündlichen Prüfung: 6. September 2010

Abstract

To assist rational compound design of organic semiconductors, two problems need to be addressed. First, the material morphology has to be known at an atomistic level. Second, with the morphology at hand, an appropriate charge transport model needs to be developed in order to link charge carrier mobility to structure.
The former can be addressed by generating atomistic morphologies using molecular dynamics simulations. However, the accessible range of time- and length-scales is limited. To overcome these limitations, systematic coarse-graining methods can be used. In the first part of the thesis, the Versatile Object-oriented Toolkit for Coarse-graining Applications is introduced, which provides a platform for the implementation of coarse-graining methods. Tools to perform Boltzmann inversion, iterative Boltzmann inversion, inverse Monte Carlo, and force-matching are available and have been tested on a set of model systems (water, methanol, propane and a single hexane chain). Advantages and problems of each specific method are discussed.
In partially disordered systems, the second issue is closely connected to constructing appropriate diabatic states between which charge transfer occurs. In the second part of the thesis, the description initially used for small conjugated molecules is extended to conjugated polymers. Here, charge transport is modeled by introducing conjugated segments on which charge carriers are localized. Inter-chain transport is then treated within a high temperature non-adiabatic Marcus theory while an adiabatic rate expression is used for intra-chain transport. The charge dynamics is simulated using the kinetic Monte Carlo method.
The entire framework is finally employed to establish a relation between the morphology and the charge mobility of the neutral and doped states of polypyrrole, a conjugated polymer. It is shown that for short oligomers, charge carrier mobility is insensitive to the orientational molecular ordering and is determined by the threshold transfer integral which connects percolating clusters of molecules that form interconnected networks. The value of this transfer integral can be related to the radial distribution function. Hence, charge mobility is mainly determined by the local molecular packing and is independent of the global morphology, at least in such a non-crystalline state of a polymer.

Zusammenfassung

Bei der systematischen, Computer gestützten Entwicklung neuer organischer Halbleiter müssen zwei Kernpunkte behandelt werden: Zunächst muss deren Morphologie auf atomarer Ebene bestimmt werden. Anschließend wird ein Modell benötigt, welches Ladungsträgermobilitäten für diese Strukturen berechnen kann.
Ersteres kann durch Molekulardynamik Simulationen erfolgen, jedoch sind die erreichbaren Längen- und Zeitskalen stark eingeschränkt. Hier bieten systematische Vergröberungsmethoden einen Ausweg, welche im ersten Teil dieser Arbeit vorgestellt werden. Im Rahmen dieser Arbeit wurde ein Software-Paket ("Versatile Object-oriented Toolkit for Coarse-graining Applications") entwickelt, welches eine flexible Plattform für die einheitliche Implementierung von Vergröberungsmethoden bietet. Bisher wurden Anwendungen für die Methoden Boltzmann Inversion, iterative Boltzmann Inversion, Monte Carlo Inversion und Force-matching integriert und an vier Referenzsystemen (Wasser, Methanol, Propan und ein einzelnes Hexanmolekül) getestet. Die Vor- und Nachteile der verschiedenen Ansätze werden diskutiert.
Die Modellierung der Ladungsträgerdynamik ist in ungeordneten Systemen eng mit der Konstruktion geeigneter diabatischer Zustände verbunden, zwischen welchen Ladungstransport erfolgen kann. Im zweiten Teil dieser Arbeit wird ein Modell entwickelt, welches Ladungstransport in konjugierten Poylmeren beschreibt. Hierbei wird der Transport durch Springen zwischen konjugierten Segmenten (Bereiche im Polymer, auf welchen Ladungsträger lokalisieren können) beschrieben. Die Transportraten der Sprünge zwischen Molekülen wird mit der nicht-adiabatischen Marcus-Gleichung berechnet, wohingegen eine adiabatische Ratengleichung für Transport innerhalb der Polymerkette verwendet wird. Die Dynamik der Ladungsträger wird anschließend mit dem kinetischen Monte Carlo Algorithmus simuliert.
Im letzten Teil dieser Arbeit wird mit Hilfe des entwickelten Modells eine Relation zwischen Morphologie und Ladungsmobilität in neutralen und dotierten Zuständen des konjugierten Polymers Polypyrrol etabliert. Es wird gezeigt, dass bei kurzen Ketten die Mobilität der Ladungsträger kaum von der molekularen Ordnung abhängt. Zudem kann die Mobilität anhand des Schwellenwerts des Transferintegrals abgeschätzt werden, welches Moleküle zu einem einzelnen Cluster verbindet. Die Tatsache, dass das Transferintegral eng mit der radialen Verteilungsfunktion verknüpft ist, deutet darauf hin, dass die Ladungsmobilität, zumindest in solch einem nicht-kristallinen Zustand eines Polymers, überwiegend durch die lokale molekulare Packung gegeben ist und damit unabhängig von der globalen Ordnung ist.

Contents

Chapter 1

Organic electronics

Organic electronics promises the possibility of developing devices with electronic properties of inorganic materials and advantageous processing and mechanical properties of plastic materials. Appropriately designed compounds can be processed from solution, and thus cost-efficient techniques such as spin coating or ink-jet printing [1] can be employed. The use of plastic semiconducting materials allows for the design of flexible electronics. Typical applications are e-paper [2], bendable solar cells [3], rollable light sources and displays [4].

At present, four distinct classes of organic conductive materials are being scrutinized for organic electronics applications: (i) small organic molecules assembled in crystals, normally by vapor deposition. Typical examples here are polyacene or rubrene crystals [5, 6]; (ii) soluble small organic molecules which self-assemble in supramolecular structures (molecular wires). Discotic liquid crystals and some organic oligomers are typical representatives of such materials [7, 8, 9]; (iii) soluble conjugated polymers, such as derivatives of polythiophene, poly(p-phenylene vinylene) [10, 11]; (iv) finally, doped conjugated polymers such as polyacetylene and polypyrrole [12, 13, 14]. Some of the typical compounds are depicted in figure 1.1.

Thin organic semiconducting layers are typically used in three device types: organic light emitting diodes [15], field effect transistors [16] and solar cells [17]. These are depicted in figures 1.2 and 1.3. The properties of a semiconducting layer have to be adjusted for a specific application. The advantage of organic materials is that synthesis can be used to tune molecular properties, e.g. the band gap, light absorption spectra, etc. For the majority of applications two requirements are essential: high charge carrier mobilities and stability of materials. Indeed, significant efforts have been invested in improving charge mobilities and reported to be successful [18, 19]. It has been concluded, however, that optimizing the electronic structure is not sufficient,

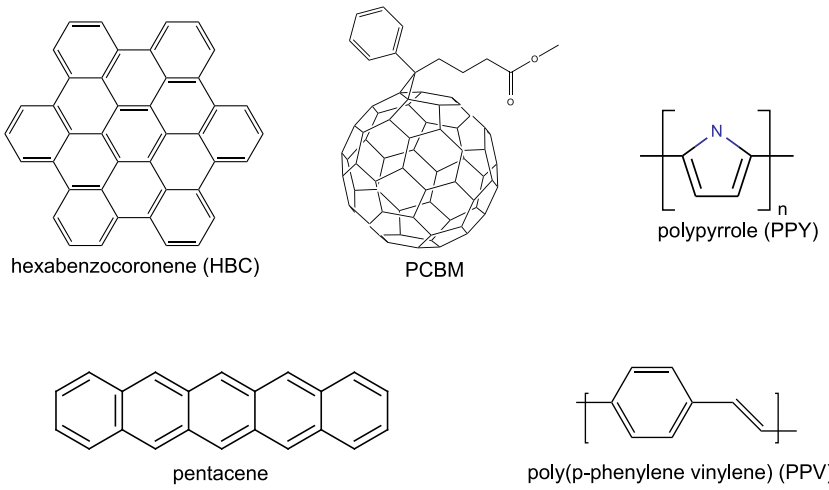


Figure 1.1: **Typical compounds used in organic electronics.**

since the material morphology, which heavily depends on the processing technique and the chemical structure, can alter charge mobility by several orders of magnitude [20, 21].
In addition, the global ordering of molecules is important for devices. For example, amorphous materials are more suitable OLEDs since in this case charge carriers sample more sites which increases the probability of electron-hole recombination [15]. Furthermore, OLEDs operate at high charge carrier densities and therefore require stable materials, which is easier to achieve with small, non-soluble molecules. In contrast, well-structured materials, for example crystals or molecules that self assemble in ordered monolayers, are preferred for OFETs, where defect free conducting layers and high charge carrier mobilities are required [16].
In the following section, the functionality of a bulk heterojunction solar cell is discussed in detail. Based on this discussion, the requirements for the molecular assembly on disparate length scales are outlined, emphasizing the need for multiscale simulation techniques.

1.1 Organic solar cells

In solar cells, absorption of light creates excitons (Coulomb bound electron-hole pairs), which subsequently dissociate into free charge carriers and can then be transported to the contact electrodes. In inorganic semiconductors, the binding energy of an exciton (the energy needed to separate hole and electron) is much smaller than the thermal energy at room temperature.

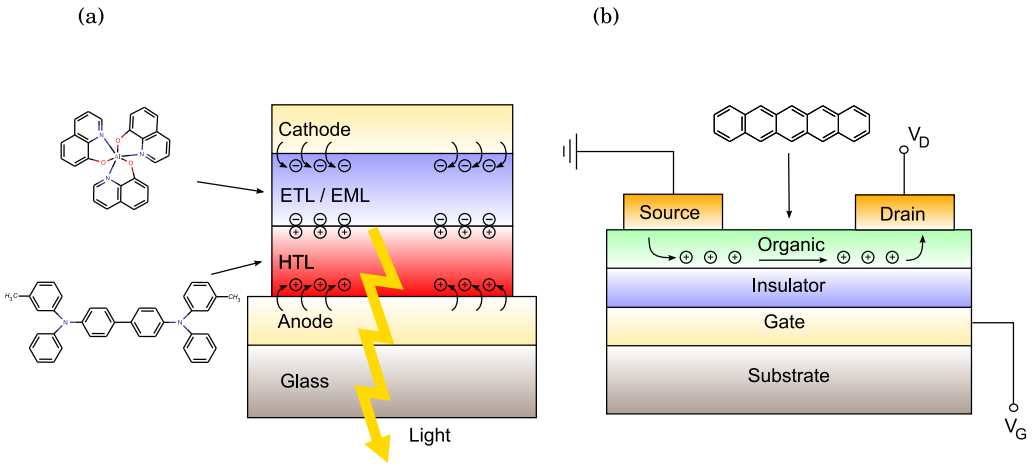


Figure 1.2: **Two types of organic semiconductors devices.** (a) Organic light emitting diode (OLED). The diode is manufactured as a heterolayer structure, with a hole transporting layer (HTL) and electron transporting layer (ETL). Holes and electrons are transported to the interface and recombine in the emission layer (EML). (b) Organic field-effect transistor (OFET). Depending on an p- or n-type transistor, a hole or electron transporting material is used, respectively. Adapted from ref. [22].

Therefore, charges can be separated due to thermal fluctuations and dissimilar work functions of the electrodes. In contrast, the low dielectric constant in organic solar cells results in strongly bound electron-hole pairs. The binding energy in organic semiconductors is of the order of several $k_B T$ and an additional mechanism is needed to drive charge separation [23].

An effective method to achieve this is via a donor-acceptor interface using compounds with different highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) energies. The difference in energy has to be at least the binding energy of the exciton. If an exciton reaches the interface, the electron will be transferred to the acceptor and the hole will reside in the donor. After separation, free charge carriers travel to the electrodes and are injected into an external circuit.

Several factors influence the efficiency of an organic solar cell, such as light absorption, exciton and charge carrier mobilities. While thick photoactive layers are required to enhance absorption of light, excitons must still be able to reach the interface within their lifetime. The typical diffusion length of 10 nm requires fine-grained structures with a large interface/bulk ratio. This issue is addressed in bulk heterojunction solar cells [24] as depicted in fig-

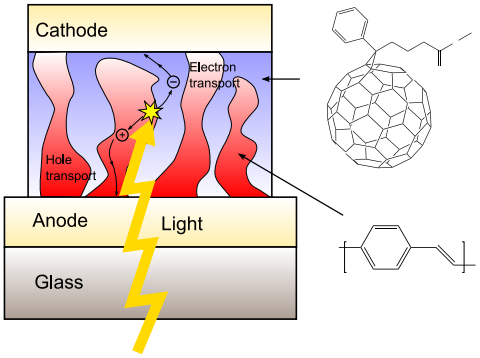


Figure 1.3: **Organic photovoltaic cell (OPVC).** Exciton binding energy is high in organic semiconductors. Bulk heterojunctions are needed to separate charges and produce a functioning device. Effective morphology control is one of the key factors in producing effective OPVCs. Adapted from ref. [22].

ure 1.3 by, for example, intermixing two conjugated polymers with different electron affinities or a conjugated polymer with a C_{60} derivative. Phase separation of the two compounds leads to a large interface/bulk ratio compared to a monolayer structure and excitons are formed near an interface [23]. At the same time, donor and acceptor domains still need to be connected to the electrodes in order to provide pathways for charges after they are separated. Self-assembling materials, such as discotic liquid crystals that form columnar phases [25] or block copolymers, can be used to improve morphology control during device manufacturing.

To summarize, apart from the electronic structure, both local molecular packing as well as domain alignment shall be controlled in order to obtain an efficient solar cell. The local packing of molecules directly affects exciton and charge carrier mobilities. Local packing is accessible via atomistic molecular dynamics simulations, where the initial configuration can be prepared with the help of experimental data, such as x-ray scattering or solid state NMR. On the other hand, global morphology (phase separation, nucleation and growth) cannot be addressed using atomistic simulations due to the limited range of time- and length-scales. Coarse graining techniques, which are described in sec. 2.2, can be used to overcome some of these limitations and to allow for simulations of larger scale morphologies.

1.2 Mobility measurements

Several techniques have been developed to measure mobilities in organic semiconductors. They can be divided into groups based on how charge carriers are created (see figure 1.4). When comparing measurements of different

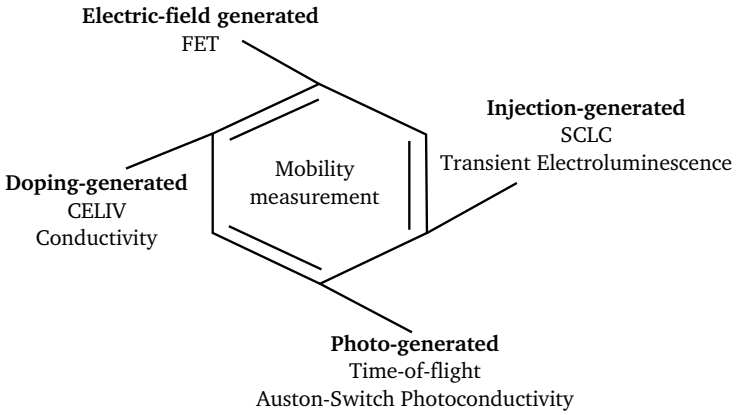


Figure 1.4: **Methods to measure mobilities in organic semiconductors.** Methods can be divided into groups based on how charge carriers are created. Injection of charge carriers from electrodes (interface), photoexcitation of charge carriers, free charge carriers due to doping of the sample e.g. during synthesis and electric field generated. Taken from ref. [26].

techniques, an essential point to consider is that these operate at different conditions. For example, conductivity measurements require the presence of intrinsic charge carriers as it is the case in doped systems. Other methods explicitly create free charge carriers and, depending on the type of creation (e.g. injection, photo-excitation), interface effects might have to be taken into account. In addition, the charge carrier density at which a measurement takes place as well as the path of charge movement can significantly affect the measured mobility. For example the time-of-flight (TOF) technique measures bulk mobilities at low charge carrier densities, a local mobility is measured in time resolved microwave conductivity experiments, and a surface mobility at high charge carrier densities is measured using a FET setup.
To highlight the differences which are important when simulation data is compared to experiments, two techniques, time of flight and field effect transistor measurements are discussed in the following sections.

1.2.1 Time of flight

A time of flight (TOF) experiment is the most direct way to measure charge carrier mobility. In this method, the mobility is determined by monitoring the time a charge needs to traverse the sample. The experimental setup is depicted in figure 1.5(a). The sample is embedded between two electrodes, one of which is semi-transparent. An external field is generated by applying a voltage to the electrodes. Charge carriers are created by a short laser pulse with a high enough photon energy to create free charge carriers. Subsequently, the charges propagate along the electric field and generate a displacement current as depicted in figure 1.5(b). This displacement current drops, as soon as the charges reach the second electrode. Knowing the duration t of time it takes for charges to cross the sample and sample thickness d, a drift velocity v and charge carrier mobility can be calculated as

$$\mu = \frac{v}{E} = \frac{d}{Et} = \frac{d^2}{Vt} \tag{1.1}$$

with the electric field denoted as E and the voltage applied to the electrodes as V.

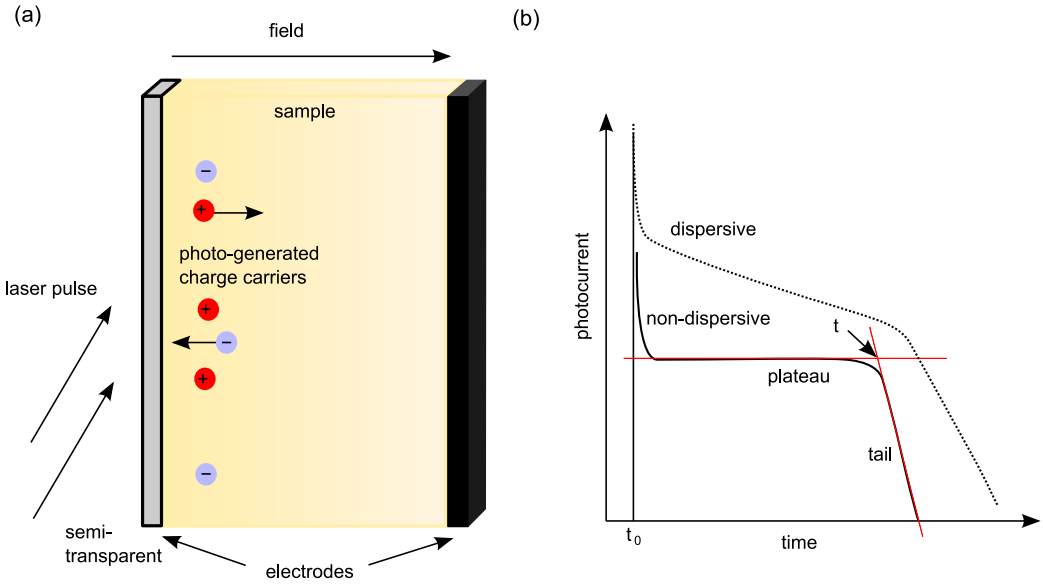


Figure 1.5: **Time of flight experiment** (a) Schematic of a time of flight experiment. (b) Mobilities are determined by monitoring the photocurrent and determining the time the charges require to cross the sample.

An advantage of the TOF technique is its insensitivity to interface effects at the contact electrodes. Charges are photo-generated within the sample,

while the current is measured capacitively. Hole and electron mobilities can be measured separately by changing the direction of the applied field.
To obtain a transient signal, which is comprised of a plateau followed by a rapid decrease shown by the solid curve in figure 1.5(b), charge carriers must not be dispersed. This allows the evaluation of a transient time at the intersection of the tangents, depicted in red. However, in highly disordered materials, transport may be dispersive and an ideal transient sharp cannot be obtained (see dotted curve in figure 1.5b). In this case, a transient time can still be determined from a double logarithmic plot of the photocurrent. However, the obtained mobility is system size dependent due to insufficient time for thermal equilibration of charge carriers [27, 28, 29, 30].

1.2.2 Field effect transistor measurements

Another method to determine mobilities is to use a field-effect transistor setup (see figure 1.2). As it was pointed out by Horowitz [31], in transistors the I-V (current-voltage) characteristics in the linear regime can be described by

$$I_{SD} = \frac{W}{L}\mu \left(V_G - V_T\right) V_{SD} \,, \tag{1.2}$$

and in the saturated regime by

$$I_{SD} = \frac{W}{2L}\mu C \left(V_G - V_T\right)^2 \,, \tag{1.3}$$

where I_{SD} and V_{SD} denote the current and voltage bias between source and drain, respectively, V_G the gate voltage, and V_T the threshold voltage at which the current starts to rise. The capacitance of the gate dielectric is described by C, and the width and length of the conducting channel are given by W and L, respectively [32]. The typical current-voltage characteristics of a FET are depicted in figure 1.6.
FET measurements show two main differences to TOF measurements. In TOF, mobilities perpendicular to the electrodes are measured, whereas mobilities within the plane of the substrate are measured in FET. The transport occurs in a narrow channel of a few tenth of nm and is affected by structural defects within the organic layer at the interface as well the surface topology and polarity of the dielectric. These effects are less pronounced for the case of amorphous materials, but can play a significant role in crystals or columnar phases in liquid crystalline systems.
Another difference concerns the charge carrier densities. Photoexcitation in a TOF experiment only creates a small amount of charge carriers. Therefore,

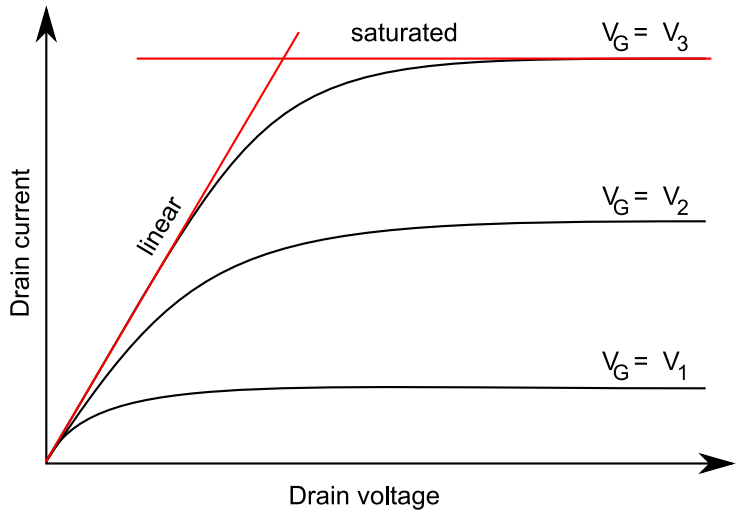


Figure 1.6: **Current-voltage characteristics of a FET.** $V_1 < V_2 < V_3$. Changing the gate voltage V_G alters the resistance of the conducting channel. For drain voltages V_D much smaller than the gate voltage V_G, the drain current V_D has ohmic behavior and is proportional to V_D. Increasing V_D creates a significant asymmetrically-shaped charge distribution in the conducting channel and the drain current saturates.

the respective measurements always occur in the limit of low charge carrier density. In contrast, charge carrier densities in FET measurements are much higher. Since mobility strongly depends on the charge carrier density, the measured mobilities can be of several orders of magnitude higher than those obtained during a TOF experiment [33] and should be extrapolated to the limit of low charge carrier density when these techniques are compared [10].

1.3 Rational compound design

As has been discussed previously in this chapter, several aspects can affect the performance of a device. Hence, the screening of new materials for use in organic electronics can be a cumbersome task. In experiments the first step is synthesis. Then, to evaluate the properties of the new compound, its morphology is characterized, a set of measurements (e.g. conductivity, mobility, absorption spectrum) is collected or it is directly analyzed in a specific application by fabrication of a complete device. However, it can be difficult in experiments to control specific properties (e.g. the morphology), in order to systematically analyze their effect on the measurements (e.g. the mobility). Both chemical structure and processing conditions can affect the

material morphology. In this situation, the overall device performance is determined by the electronic structure of a single molecule, as well as the long- and short-range ordering of molecules.
Computer simulations offer an additional perspective on the processes that occur during charge transport in these materials. For example, charge transport parameters can be calculated for a subset of the system (e.g. a pair of molecules) as well as global morphologies can be prepared and analyzed. However, computer simulations are often limited to small system sizes, short timescales and need experimental data as input. Therefore, a symbiosis of theory and experiments is needed in order to understand and systematically optimize charge transport in organic semiconductors.
In this context, the main focus of this thesis is to develop a model to investigate the effect of material morphology on charge carrier mobility using computer simulations. Two problems shall be addressed: first, the material morphology has to be known at an atomistic level. Second, with the morphology at hand, an appropriate charge transport model needs to be developed in order to link charge carrier mobility to structure.
The theoretical background of morphology simulations is described in chapter 2. Here, molecular dynamics may be used, although the accessible range of time- and length-scales of atomistic simulations is limited. However, systematic coarse-graining methods may be applied to overcome these limitations. To link charge mobility to the obtained morphology, a description which was originally developed to study disordered assemblies of small molecules has been extended for treatment of conjugated polymers. The theory for this framework is covered in chapter 3.
The coarse-graining methods and the model to calculate charge mobility are implemented in a software package, whose design and features are outlined in chapter 4. The coarse-graining functionality is illustrated for a set of reference systems (SPC/E water, methanol, liquid propane, and a single chain of hexane) in chapter 5, where advantages and problems of each specific method are discussed. Finally, the entire model is employed in chapter 6 to establish a relation between morphology and charge carrier mobility in the neutral and doped states of polypyrrole.

Chapter 2

Morphology simulations

In this thesis, a combination of methods from computational physics and chemistry is applied to study the effect of morphology on charge transport parameters in conjugated polymers. Here, an outline of the methods which are essential for the understanding of the topics discussed is given. For an overview of density functional theory (DFT) or molecular dynamics (MD), the reader is referred to text books [34, 35]. Well-tested implementations of these methods are provided in standard packages: Gaussian [36] was used for quantum chemical calculations and GROMACS [37] for atomistic and coarse-grained simulations.
First, methods to simulate large-scale material morphologies are outlined. The description starts with introducing the concept of force-fields for atomistic simulations and the basic principles needed for force-field refinement as performed in sec. 6.1 to obtain the atomistic model for polypyrrole. Then, coarse-graining methods are described, which enable the accessible time- and length-scales to be extended in particle-based simulations.

2.1 Force-field development

Practically all of the problems addressed in molecular modeling are too complex to be treated in full detail, that is solving the Schrödinger equation for electronic and nuclear degrees of freedom. A common simplification is the Born-Oppenheimer approximation, which assumes that the dynamics of electrons is much faster than the dynamics of the nuclei and, therefore, the dynamics of the entire system can be described by only the nuclear coordinates while electrons are assumed to be in their respective ground state associated with these coordinates. Quantum chemical calculations determine the forces on nuclei as the derivatives of the total energy, i.e. the combined energy of

electrons and nuclei with respect to the nuclear coordinates. However, this is still a computationally expensive task. The next level of simplification is to use empirical potentials to calculate forces on the nuclei instead of explicitly treating electronic structure. The system is then propagated in time by integration of the classical equations of motion. The corresponding method is called molecular mechanics, and the functional form of the potential energy surface, together with its parameters, is referred to as a force-field.

A force-field consists of contributions from *bonded* and *non-bonded* interactions. In this work the focus is on force-fields which describe bonded intra-molecular interactions via bond stretching, angle bending, and torsion potentials (see figure 2.1), and neglect cross-terms for their cross-coupling. Typical examples are the Optimized Potential for Liquid Simulations - All Atom (OPLS-AA) [38] or AMBER [39] force-fields. A force-field which has cross terms is, for example, MM3 [40].

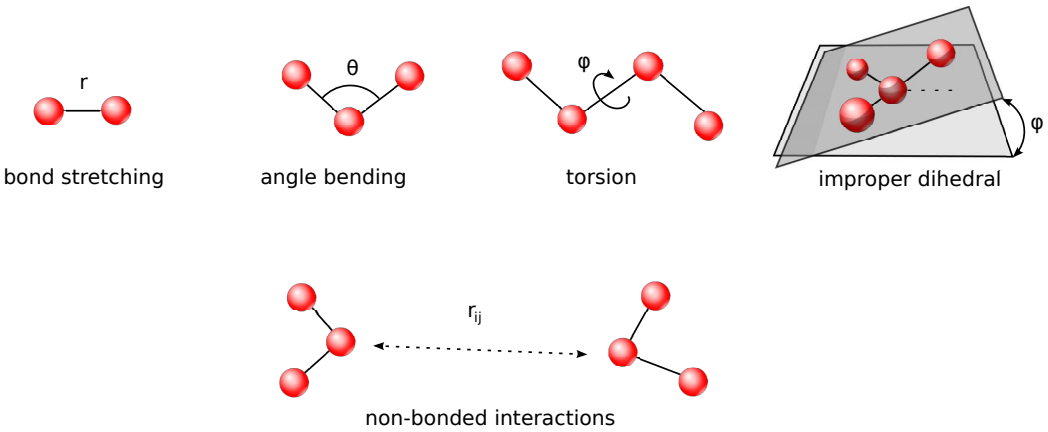


Figure 2.1: **Interaction types in force-fields.** Bond stretching, angle bending, torsion and improper dihedrals represent *bonded* interactions. Torsions are used to describe rotations around a bond while improper dihedrals are used to keep atoms within a plane. Coulomb and van-der-Waals interactions are normally used to describe *non-bonded* interactions.

All atomistic simulations in this work are based on the OPLS-AA force-field [38], for which the functional form of the potential energy surface $U(\{\boldsymbol{r}_i\})$

is given by

$$
\begin{aligned}
U\left(\{\boldsymbol{r}_i\}\right) = & \sum_{bonds} \frac{1}{2} k_b \left(r - r_0\right)^2 + \frac{1}{2} \sum_{angles} k_\theta \left(\theta - \theta_0\right)^2 \\
& + \sum_{torsions} \left\{ \frac{V_1}{2}[1 + \cos(\varphi)] + \frac{V_2}{2}[1 - \cos(2\varphi)] \right. \\
& \left. + \frac{V_3}{2}[1 + \cos(3\varphi)] \right\} \\
& + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left\{ 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{r_{ij}} \right\}, \qquad (2.1)
\end{aligned}
$$

where bonds r and angles θ are modeled by harmonic potentials and torsion potentials are given by the first four terms of a Fourier series. The last part is a sum over all pairs of particles to account for non-bonded interactions which are calculated based on the separation r_{ij} of the two particles.
To study conjugated molecules, it can be useful to add an additional term to eq. 2.1 for improper dihedrals (see figure 2.1). The difference to proper torsions is that the latter are used to describe rotations around a bond, while improper dihedrals are constructed to keep atoms in a plane. However, the evaluation as well as the implementation of the force calculation is identical except for the functional form of the potential. While proper torsions are usually described by periodic potentials, a harmonic potential is used for improper dihedrals.
Non-bonded interactions are usually divided into two groups. Van-der-Waals interactions combine London dispersion forces for attraction and steric repulsion, both of which are short ranged and can be treated using a cut-off. Electrostatic interactions between partially charged atoms are long ranged. Thus, the direct evaluation of the Coulomb potential using a cut-off is computationally ineffective and can even lead to artifacts for periodic systems [41]. An alternative is to calculate the long-ranged contributions by projecting charges onto a grid. In this case, the long-ranged contributions are calculated in Fourier space, using a technique similar to Ewald summation [41].
An important aspect of all force-fields is their transferability. The challenge is to find a minimum set of parameters in eq. 2.1, which is transferable to compounds which were not used in the initial parametrization procedure. It turns out that a solely element-based definition of parameters is not sufficient. Instead, it is necessary to define atom types that reflect both the specific element and its bonding situation. The parameters are then defined based on these atom types and identical values of parameters can be used for similar compounds. In particular, dihedral potentials often need reparametrization

for their specific bonding situations. A common procedure to refine dihedral potentials is to fit the energy profile obtained in quantum chemical calculations (e.g. density functional theory). However, this energy profile represents the total energy of the system including contributions from other interactions, such as bonds, angles, improper dihedrals, Coulomb and Lennard-Jones interactions. Hence, the corresponding force-field based energy contributions have to be subtracted before fitting in order to avoid double counting of energies. This is described in sec. 6.1, in which the atomistic force-field for polypyrrole is developed. An additional point concerning transferability is that force-fields are normally parametrized for a certain state point, and their validity for different conditions is not guaranteed.
Although the concept of force-fields has been introduced based on interactions between individual atoms, it is not restricted to atomistic simulations but can in general be applied to all particle-based simulations. Examples are the coarse-grained simulations as described in the next section. Forces are evaluated in a similar way, however, tabulated potentials are used instead of the exact functional form of eq. 2.1. The coarse-graining techniques can, therefore, be seen as a systematic way to parametrize the force-field for a coarse representation of the system.

2.2 Coarse-graining

Computational materials science deals with phenomena covering a wide range of length- and time-scales, from Ångstrøms (typical bond lengths) and femtoseconds (bond vibrations) to micrometers (crack propagation) and milliseconds or more (a single polymer chain relaxation). Depending on the characteristic time- and length-scales involved, the description of the system can vary from first principles and atomistic force-fields to coarse-grained models and continuum mechanics. The role of bottom-up coarse-graining, in a broad sense, is to provide a systematic link between these levels of description.
Here the focus is on coarse-graining techniques that link two *particle-based* descriptions with a different number of degrees of freedom. The system with the larger number of degrees of freedom is denoted as the *reference* system. The system with the reduced number of the degrees of freedom is referred to as the *coarse-grained* system (see figure 2.2). An example which is treated in detail in this work is an all-atom (reference) and a single site (coarse-grained) models of water. Other examples can be readily found in the literature [42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53].
The prime task of systematic coarse-graining is to derive a potential energy function of the coarse-grained system, which reproduces the properties of

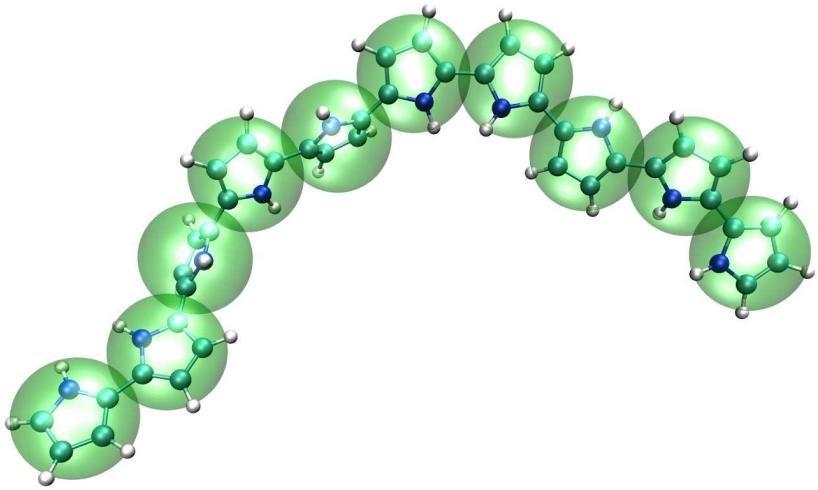


Figure 2.2: **Coarse grained model for polypyrrole.** Coarse-graining combines groups of atoms into beads, here one bead per repeat unit. The role of systematic coarse-graining is to find a set of potentials such that the coarse-grained model reproduces selected properties of the reference system.

interest (e.g. distribution functions, pressure, free energy) of the reference system. To do this, several coarse-graining approaches can be used. From the point of view of implementation, these approaches can be divided in *iterative* and *non-iterative* methods. *Boltzmann inversion* is a typical example of a non-iterative method [42]. In this method, which is exact for *independent* degrees of freedom, coarse-grained interaction potentials are calculated by inverting the distribution functions of the coarse-grained system. Another example of a non-iterative method is *force matching*, where the coarse-grained potential is chosen in such a way that it reproduces the forces on the coarse-grained beads [54, 46]. Configurational sampling [55], which matches the potential of mean force, also belongs to this category. Boltzmann inversion and force matching only require a trajectory for a reference system. Once that is known, coarse-grained potentials can be calculated for any mapping scheme. Note that Boltzmann inversion often uses a "special" trajectory which is designed to decouple the degrees of freedom of interest, e.g. a single polymer chain in vacuum with appropriate exclusions [42].
Iterative methods refine the coarse-grained potential by re-iterating coarse-grained simulations and calculating corrections to the potential on the basis of the reference and coarse-grained observables (e. g. radial distribution function or pressure). The simplest example is the iterative Boltzmann inversion method [56], which is an iterative analogue of the Boltzmann inversion

method. More sophisticated (in terms of the update function) is the inverse Monte Carlo approach [57].
One can also classify systematic coarse-graining approaches by micro- and macroscopic observables they use to derive the coarse-grained potential, such as structure-based [57, 58, 42], force-based [54, 59, 46], potential-based [60], and free energy based approaches [61], where the name identifies the observable used for coarse-graining. Note that hybrids of these methods are also possible [44, 53].
With a rich zoo of methods plus their combinations[1] available at hand, the question arises whether there is an optimal method for a specific class of systems. On a more fundamental level one might wonder whether the different methods provide the same coarse-grained potential and whether it is possible to formulate a set of (even empirical) rules favoring one method over another. It is obvious that this is a difficult task to treat analytically, especially for realistic systems. To assess the quality of a particular coarse-graining technique one needs to apply all available methods to a certain number of systems and to compare and quantify the degree of discrepancy between the coarse-grained and reference descriptions. This is, however, cumbersome in particular due to the absence of a single package where all these methods are implemented with the same accuracy and same level of technical detail.
In this thesis such a coarse-graining package is developed. First, the basic ideas behind each method are described in this chapter, paying special attention to the technical issues one has to overcome when implementing them. The design of the software package is described in chapter 4. In chapter 5, the implemented methods are illustrated and compared by coarse-graining systems of different complexity: a three-site SPC/E water, methanol, propane, and hexane.

2.2.1 General considerations

Here, the general concepts of distribution based coarse-graining are recapitulated. This is mainly a brief summary of the paper by Noid *et al.* [62]. The discussion starts with the definition of a high resolution (atomistic) and a low resolution (coarse-grained) model for a given system. First, a link between these two models via the mapping operator is introduced, and second, conditions for a consistent coarse-grained model are given.
The instantaneous configuration of an atomistic system is described by its

[1] Bonded- and non-bonded interactions can be treated separately using different methods, e.g. Boltzmann inversion for bonded and force-matching for non-bonded interactions.

Cartesian coordinates $\boldsymbol{r}^n$ and momenta $\boldsymbol{p}^n$

$$\boldsymbol{r}^n = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n\}, \tag{2.2}$$
$$\boldsymbol{p}^n = \{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n\} \tag{2.3}$$

of the n atoms in the system. At a coarse-grained level, the coordinates and momenta are specified by the positions $\boldsymbol{R}^N$ and momenta $\boldsymbol{P}^N$ of coarse-grained sites

$$\boldsymbol{R}^N = \{\boldsymbol{R}_1, \ldots, \boldsymbol{R}_N\}, \tag{2.4}$$
$$\boldsymbol{P}^N = \{\boldsymbol{P}_1, \ldots, \boldsymbol{P}_N\}. \tag{2.5}$$

Note that capitalized symbols are used for the coarse-grained sites while lower-case letters describe the atomistic system.
The mapping operator establishes a link between the atomistic and coarse-grained representations of the system. Here only the special case of linear mapping schemes is discussed, where the mapping operator can be written as a vector $\boldsymbol{c}_I$ for each bead I

$$\boldsymbol{R}_I = \sum_{i=1}^{n} c_{Ii}\boldsymbol{r}_i, \tag{2.6}$$

$$\boldsymbol{P}_I = M_I\dot{\boldsymbol{R}}_I = M_I\sum_{i=1}^{n} c_{Ii}\dot{\boldsymbol{r}}_i = M_I\sum_{i=1}^{n}\frac{c_{Ii}}{m_i}\boldsymbol{p}_i. \tag{2.7}$$

for all $I = 1, \ldots, N$. Note that the mapping operator reduces the number of degrees of freedom and several atomistic configurations $\{r_i\}$ can lead to identical coarse-grained coordinates R_I. Due to translational invariance, that is if an atomistic system is translated by a constant vector the corresponding coarse-grained system is also translated by the same vector, the mapping operator must fulfill the following conditions:

$$\sum_{i=1}^{n} c_{Ii} = 1\,. \tag{2.8}$$

In some cases it is useful to define coarse-grained mapping in a way that some atoms belong to several coarse-grained beads [63]. In this case, two types of atoms for each of the N coarse-grained beads can be defined [62], namely *specific* and *involved* atoms, as depicted in figure 2.3. For each site I, a set of *involved* atoms $\mathcal{I}_I$ is defined as

$$\mathcal{I}_I = \{i | c_{Ii} \neq 0\}. \tag{2.9}$$

An atom i in the atomistic model is involved in a coarse-grained site I if this atom provides a nonzero contribution to the sum in eq. 2.7. A set of *specific* atoms is defined as:

$$\mathcal{S}_I = \{i | c_{Ii} \neq 0 \text{ and } c_{Ji} = 0 \text{ for all } J \neq I\}. \tag{2.10}$$

In other words, atom i is specific to site I if it is only involved in site I and not in any other sites.

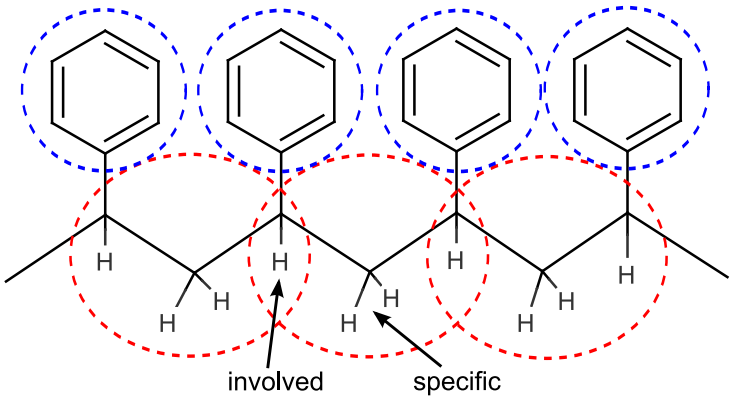


Figure 2.3: **Specific and involved atoms.** Involved atoms are shared by two or more coarse-grained beads while specific atoms only belong to a single coarse grained bead. The mapping scheme for polystyrene is described in ref. [63].

Further discussion is based on the Hamiltonian for the atomistic and coarse-grained systems. In the atomistic case, the Hamiltonian can be written as

$$h\left(\boldsymbol{r}^n, \boldsymbol{p}^n\right) = \sum_{i=1}^{n} \frac{1}{2m_i} \boldsymbol{p}_i^2 + u\left(\boldsymbol{r}^n\right) \tag{2.11}$$

The equilibrium probability density p_{rp} of the atomistic system is given by

$$p_{rp}\left(\boldsymbol{r}^n, \boldsymbol{p}^n\right) = p_r\left(\boldsymbol{r}^n\right) p_p\left(\boldsymbol{p}^n\right) , \tag{2.12}$$

$$p_p\left(\boldsymbol{p}^n\right) \propto \exp\left(-\sum_{i=1}^{n} \frac{\boldsymbol{p}_i^2}{2m_i}\right) . \tag{2.13}$$

The Hamiltonian $H\left(\boldsymbol{R}^n, \boldsymbol{R}^n\right)$ and distribution functions $P_R\left(\boldsymbol{R}\right)$, $P_P\left(\boldsymbol{P}\right)$ of the coarse-grained system are identical, except that the lower-case variables should be replaced with the capitalized ones.

The criteria for the two models to be consistent is that the distribution functions of the coarse grained model should match those of the mapped atomistic model,

$$P_r\left(\boldsymbol{R}^n\right) = p_r\left(\boldsymbol{R}^n\right) , \tag{2.14}$$
$$P_p\left(\boldsymbol{P}^n\right) = p_p\left(\boldsymbol{P}^n\right) . \tag{2.15}$$

Note that equations (2.14) and (2.15) specify the manybody distribution function. Both equations can be rewritten by applying the mapping operator to the atomistic distribution function. By integrating out atomistic degrees of freedom, the conditions for consistency in the configuration space can be obtained [62]

1. The mapping operator should be linear[2]

2. Each coarse-grained site shall have at least one atom which is specific to this site[3].

Naturally, additional requirements must be imposed on the coarse-grained potential. They are discussed in sec. 2.2.5, where the force-matching method is described. If the distribution of momenta, eq. 2.15, should match as well, the following additional requirements arise

1. All beads have to be specific (coarse-grained beads do not share atoms).

2. The coarse grained masses must satisfy the following relation,

$$M_I = \left(\sum_{i \in \mathcal{I}_I} \frac{c_{Ii}^2}{m_i} \right)^{-1} . \tag{2.16}$$

   This expression simplifies to the total mass if the center of mass is used for the position of the coarse-grained bead, i.e.

$$c_{Ii} = m_i / M_I , \tag{2.17}$$
$$M_I = \sum_{i \in \mathcal{I}_I} m_i . \tag{2.18}$$

[2]The conclusions below are valid only for linear mapping operators. A non-linear mapping scheme is, in principle, possible but is not considered here.

[3]This is required to perform the partial integration when averaging the atomistic degrees of freedom in the right-hand-side of eq. 2.14 [62].

A model that does not fulfill these requirements will have a different distribution of momenta compared to the mapped reference model. Note that the momenta of atoms due to intra-bead contributions are averaged out and only the effective contribution which acts on the mapping point remains. Special attention is required if the coarse-grained center of a bead is not in the center of mass. In this case, the mass of the coarse-grained bead is not the sum of all atom masses, but is given by eq. 2.16.
If the criteria above are fulfilled, the coarse-grained model can match the full manybody distribution function. However, in reality, the coarse-grained force-field does not have manybody terms, and only the pair correlation functions (e.g. radial distribution function) can be used to parametrize the coarse-grained potentials and higher correlation functions are ignored. For example, when coarse-graining water using iterative Boltzmann inversion (IBI), the radial distribution function matches perfectly but tetrahedral packing cannot be reproduced [64]. In what follows, specific techniques to derive coarse-grained potentials are outlined.

2.2.2 Boltzmann inversion

Boltzmann inversion is the simplest method one can use to obtain coarse-grained potentials [42]. It is mostly used for *bonded* potentials, such as bonds, angles, and torsions. Boltzmann inversion is structure-based and only requires positions of atoms.
The idea of Boltzmann inversion stems from the fact that in a canonical ensemble *independent* degrees of freedom q obey the Boltzmann distribution, i. e.

$$P(q) = Z^{-1} \exp\left[-\beta U(q)\right] \ , \tag{2.19}$$

where $Z = \int \exp\left[-\beta U(q)\right] \mathrm{d}q$ is a partition function, $\beta = 1/k_{\mathrm{B}}T$. Once $P(q)$ is known one can invert eq. 2.19 and obtain the coarse-grained potential, which in this case is a potential of mean force

$$U(q) = -k_{\mathrm{B}}T \ln P(q) \ . \tag{2.20}$$

Note that the normalization factor Z is not important since it would only enter the coarse-grained potential $U(q)$ as an irrelevant additive constant.
In practice, $P(q)$ is computed from the trajectory of the reference system which is sampled either by Monte Carlo, molecular dynamics, stochastic dynamics, or any other integrator that ensures a canonical distribution of states.
Boltzmann inversion is simple to implement, however one has to be careful with the rescaling of the probability P due to normalization as well as

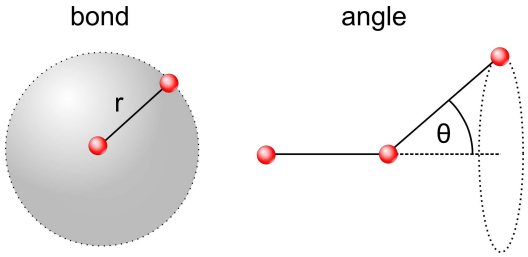


Figure 2.4: **Rescaling for bonds and angles.** Before Boltzmann inverting distribution functions for bonds and angles, they have to be normalized by the area of a spherical surface and the circumference of a rotation, respectively.

computational issues. The probability normalization can be explained on a particular example of coarse-graining of a single polymer chain by beads with bond, angle and torsion potentials. In this case the coarse-grained potential U depends on three variables, bond length $\boldsymbol{r}$, angle θ and torsion angle φ. Assuming, as before, a canonical distribution and independence of the coarse-grained degrees of freedom, one can write

$$P(r,\theta,\varphi) = \exp\left[-\beta U(r,\theta,\varphi)\right] , \tag{2.21}$$

$$P(r,\theta,\varphi) = P_r(r)P_\theta(\theta)P_\varphi(\varphi) . \tag{2.22}$$

If now the histograms for the bonds $H_r(r)$, angle $H_\theta(\theta)$, and torsion angle $H_\varphi(\varphi)$ are computed, they must be rescaled in order to obtain the normalized distribution functions (see figure 2.4).

$$P_r(r) = \frac{H_r(r)}{4\pi r^2} , P_\theta(\theta) = \frac{H_\theta(\theta)}{\sin\theta} , P_\varphi(\varphi) = H_\varphi(\varphi) . \tag{2.23}$$

The coarse-grained potential can then be calculated by Boltzmann inversion of the distribution functions

$$U(r,\theta,\varphi) = U_r(r) + U_\theta(\theta) + U_\varphi(\varphi) , \tag{2.24}$$

$$U_q(q) = -k_\mathrm{B}T \ln P_q(q),\ q = r,\theta,\varphi .$$

On the technical side, the implementation of the Boltzmann inversion method requires *smoothing* of $U(q)$ to provide a continuous force. Triangular or spline smoothing can be used for this purpose. Poorly and unsampled regions, that is regions with high $U(q)$, must be *extrapolated*. Since the contribution of these regions to the canonical density of states is small, the exact shape of the extrapolation is less important.
Another crucial issue is the cross-correlation of the coarse-grained degrees of freedom. Independence of the coarse-grained degrees of freedom is the main

assumption that allows factorization of the probability distribution, eq. 2.22, and the potential, eq. 2.24. Hence, one has to carefully check whether this assumption holds in practice. This can be done by performing coarse-grained simulations and comparing cross-correlations for all pairs of degrees of freedom in atomistic and coarse-grained resolution, e. g. using a two-dimensional histogram, analogous to a Ramachandran plot. This concept is illustrated in sec. 5.5 for a single molecule of hexane. Note that checking the linear correlation coefficient does not guarantee statistical independence of variables, for example $c(x, x^2) = 0$ if x has a symmetric probability density $P(x) = P(-x)$. This case is often encountered in systems used for coarse-graining [65, 66].

2.2.3 Iterative Boltzmann inversion

Iterative Boltzmann inversion (IBI) is a natural extension of the Boltzmann inversion method. Since the goal of the coarse-grained model is to reproduce the distribution functions of the reference system as accurately as possible, one can also iteratively refine the coarse-grained potentials using some numerical scheme. Depending on the update function, this can be done either by using iterative Boltzmann inversion [56] or the inverse Monte Carlo method [57, 58].
In iterative Boltzmann inversion, the coarse-grained potential is refined according to the following scheme

$$\begin{aligned} U^{(n+1)} &= U^{(n)} + \Delta U^{(n)} \,, \\ \Delta U^{(n)} &= k_\mathrm{B}T \ln \frac{P^{(n)}}{P_\mathrm{ref}} = U^\mathrm{ref}_\mathrm{PMF} - U^{(n)}_\mathrm{PMF} \,. \end{aligned} \tag{2.25}$$

One can easily see that convergence is reached as soon as the distribution function $P^{(n)}$ matches the reference distribution function P_ref, or, in other words, the potential of mean force, $U^{(n)}_\mathrm{PMF}$ converges to the reference potential of mean force. Note that eq. 2.25 is nothing else but a numerical scheme that allows to match the coarse-grained and the reference distribution functions. A possible choice for the initial guess of the iterative procedure is the potential of mean force, eq. 2.20. In the case of non-bonded interactions, it reads

$$U^{(1)}(r) = -k_\mathrm{B}T \ln g(r) \,, \tag{2.26}$$

where $g(r)$ is the radial distribution function. In fact, it can be seen as a first-order correction to the interaction potential with respect to a gas of non-interacting particles. In an ideal gas (non-interacting particles), the probability of finding two particles at a distance r is $P^{(0)} \propto 4\pi r^2$, which is equivalent to the statement that the radial distribution function of an

ideal gas is one. Substituting $P^{(0)}$ into eq. 2.25, the first iteration $U^{(1)} = -k_\mathrm{B}T \ln(P_\mathrm{ref}/4\pi r^2)$ is obtained, which is the potential of mean force, eq. 2.26.

IBI can be used to refine both bonded and non-bonded potentials. It is primarily used for simple fluids with the aim of reproducing the radial distribution function of the reference system in order to obtain non-bonded interactions [56]. It can have convergence problems for multicomponent systems, since it does not account for cross-correlation terms, that is the updates for P_AA, P_AB, and P_BB are not coupled (the subscript enumerates a single component in a multicomponent system). For such systems, the inverse Monte Carlo method is said to work better. The scheme can be stabilized by multiplying the update function, $\Delta U^{(n)}$, by a factor $\eta \in [0..1]$.

On the implementation side, IBI has the same issues as the inverse Boltzmann method, i. e. smoothing and extrapolation of the potential must be implemented.

It shall also be mentioned that, according to the Henderson theorem [67, 68], which is a classical analogue of the Hohenberg-Kohn theorem, the pairwise coarse-grained potential $U(r)$ is unique up to an additive constant and exists [69, 70], which, in principle, states that all structure-based iterative methods must converge to the same coarse-grained potential, provided that their aim is to exactly reproduce pair correlation functions of the reference system. As will be seen later, this is often not the case in practice, since small changes in the radial distribution function often lead to big changes in the pair potential, i. e. it is difficult to control systematic errors during the calculation of the potential update.

Another issue of coarse-graining is that coarse-grained models cannot reproduce all the statistical or thermodynamic properties of the reference system. Pressure, compressibility, or viscosity [71] are often very different from those of the reference system. In some cases, however, one can correct for some of these. For example, the viscosity can be adjusted by tuning the parameters of the thermostat [72] and the pressure can be corrected iteratively by adding a linear term to the non-bonded potential

$$\Delta U^\mathrm{pressure} = A\left(1 - \frac{r}{r_\mathrm{cut}}\right), \tag{2.27}$$

where A is either a constant, e. g. $-0.1\,k_\mathrm{B}T$ [56], or can be estimated from the virial expansion [64]. Compressibility and pressure, however, cannot be corrected simultaneously.

2.2.4 Inverse Monte Carlo

Inverse Monte Carlo (IMC) is another iterative procedure that refines the coarse-grained potentials until the coarse-grained model reproduces a set of reference distribution functions. It is very similar to IBI except that the update of the potential, ΔU, is calculated using rigorous thermodynamic arguments.
The name "inverse Monte Carlo" is somehow confusing and is due to the fact that the original algorithm was combined with Monte Carlo sampling of the phase space [57]. However, practically any sampling method can be used (e.g. molecular dynamics or stochastic dynamics) as long as it provides a canonical sampling of the phase space.
A detailed derivation of the IMC method can be found in ref. [57]. Here the more compact version for non-bonded interactions, which is outlined in ref. [68], is recapitulated, emphasizing technical problems encountered during implementation and application of the method.
The idea of IMC is to expand the potential update ΔU in a thermodynamically consistent way in terms of measurable statistical properties, e. g. radial distribution function $g(r)$. Considering a single-component system as an example, the Hamiltonian of the system can be written as

$$H = \sum_{i,j} U(r_{ij}) \ , \tag{2.28}$$

where $U(r_{ij})$ is the pair potential and it is assumed that all interactions depend only on the distance, r_{ij}, between particles i and j. Further it is assumed that this potential is short-ranged, i.e. $U(r_{ij}) = 0$ if $r_{ij} \geq r_{\text{cut}}$.
The next step is to tabulate the potential $U(r)$ on a grid of M points, $r_\alpha = \alpha\Delta r$, where $\alpha = 0, 1, \ldots, M$, and $\Delta r = r_{\text{cut}}/M$ is the grid spacing. Then the Hamiltonian, eq. 2.28, can be rewritten as

$$H = \sum_{\alpha} U_\alpha S_\alpha \ , \tag{2.29}$$

where S_α is the number of particle pairs with inter-particle distances $r_{ij} \in [r_\alpha, r_\alpha + \Delta r]$ which correspond to the tabulated value of the potential U_α.
On one hand, the average value of S_α is related to the radial distribution function $g(r)$

$$\langle S_\alpha \rangle = \frac{N(N-1)}{2} \frac{4\pi r_\alpha^2 \Delta r}{V} g(r_\alpha) \ , \tag{2.30}$$

where N is the number of atoms in the system ($\frac{1}{2}N(N-1)$ is then the number of all pairs) and Δr is the grid spacing, r_{cut}/M, V is the total volume of the

system. On the other hand, $\langle S_\alpha \rangle$ is a function of the potential U_α and hence can be expanded in a Taylor series with respect to small perturbations of U_α, ΔU_α

$$\Delta \langle S_\alpha \rangle = \sum_\gamma \frac{\partial \langle S_\alpha \rangle}{\partial U_\gamma} \Delta U_\gamma + \mathcal{O}(\Delta U^2) \ . \tag{2.31}$$

The derivatives $\partial \langle S_\alpha \rangle / \partial U_\gamma$ can be obtained by using the chain rule

$$\begin{aligned} A_{\alpha\gamma} &= \frac{\partial \langle S_\alpha \rangle}{\partial U_\gamma} \\ &= \frac{\partial}{\partial U_\gamma} \frac{\int dq S_\alpha(q) \exp\left[-\beta \sum_\lambda U_\lambda S_\lambda(q)\right]}{\int dq \exp\left[-\beta \sum_\lambda U_\lambda S_\lambda(q)\right]} \\ &= \beta \left(\langle S_\alpha \rangle \langle S_\gamma \rangle - \langle S_\alpha S_\gamma \rangle \right) \ . \end{aligned} \tag{2.32}$$

Equations 2.30, 2.31, and 2.32 allow to calculate the correction for the potential by solving a set of linear equations

$$\langle S_\alpha \rangle - S_\alpha^{\mathrm{ref}} = A_{\alpha\gamma} \Delta U_\gamma \ , \tag{2.33}$$

where S_α^{ref} is given by the target radial distribution function. The procedure is then repeated until convergence is reached.

A clear advantage of the IMC compared to the IBI method is that the update of the potential is rigorously derived using statistical mechanics and hence the iterative procedure shall converge faster with the IMC update than with the empirical IBI update. Another advantage is that, in the case of multicomponent mixtures, IMC takes into account cross-correlations of the potentials[4], that is updates for U_{AA}, U_{AB}, and U_{BB} are interdependent (A and B denote different particle types). In the IBI method these updates are independent which can lead to convergence problems for multicomponent systems.

The advantages come, of course, at a computational cost. As it is clear from eq. 2.32, one has to calculate cross-correlations of S_α. This requires much longer runs to get statistics that is good enough to calculate the potential update to a similar accuracy as IBI. The errors in the determination of the update functions of IMC and IBI methods are compared in sec. 5.1 for the case of a coarse-grained model of water.

Another issue of the IMC method is the stability of the scheme. Several factors can influence it: the first, and rather technical, point is that $g^{\mathrm{ref}}(r_\alpha)$ has to be calculated using exactly the same convention for the grid as S_α (e.g. the

[4]Changes in the potential of one interaction can affect the distribution functions of the others. This point should not be confused with correlations of observables, such as bond-angle correlations as described in sec. 5.5 for hexane.

function value should be assigned to the middle of the interval), otherwise the scheme becomes unstable. Second, inversion of $A_{\alpha\gamma}$ requires that it shall be well defined. This means that one has to remove the regions which are not sampled, such as those at the beginning of the radial distribution function. The convergence can be significantly improved if a smoothing of the potential update ΔU is used. Note that it is better to do smoothing of the update function, not the potential itself, since the latter has more features which can be lost due to too aggressive smoothing. In addition, since the IMC update assumes a linear expansion in ΔU, overshooting of the update might occur. In this case, convergence can be improved by introducing a multiplicative prefactor for the update function, similar to the stabilizing term in IBI.
Finally, it was also noticed that the systematic error in $\langle S_\alpha S_\gamma \rangle$ is always higher in the vicinity of the cutoff, which leads to a shift in the tail of the interaction potential and, as a result, to a large offset of pressure. The cross-correlation term $\langle S_\alpha S_\gamma \rangle$ is also very sensitive to the box size, and special care must be taken in order to converge the results with respect to the system size. Finite size effects are discussed in detail in sec. 5.3, where liquid methanol is coarse-grained.

2.2.5 Force matching

Force matching (FM) is another approach to evaluate coarse-grained potentials [54, 46, 73]. In contrast to the structure-based approaches, its aim is not to reproduce distribution functions, but instead try to match *forces* on coarse-grained beads as closely as possible. FM is a non-iterative method and hence is less computationally demanding. The formal statistical mechanical framework of force matching applied to a liquid state, or a multiscale coarse-graining method, is given in ref. [62]
The method works as follows. The first assumption is that the coarse-grained force-field (and hence the forces) depends on M parameters $g_1, ..., g_M$. These parameters can be prefactors of analytical functions, tabulated values of the interaction potentials, or coefficients of splines used to describe these potentials.
In order to determine these parameters, the reference forces on coarse-grained beads are calculated by properly re-weighting the forces on the atoms

$$\boldsymbol{f}_I^{\text{ref}} = M_I \sum_i \frac{c_i \boldsymbol{f}_i}{m_i}\,, \tag{2.34}$$

where $M_I = \left(\sum_i c_i^2/m_i\right)^{-1}$ is the mass of the bead I, index i numbers all atoms belonging to this bead, $\boldsymbol{f}_i$ is the force on the atom i, m_i is its mass,

c_i are mapping coefficients used to obtain the position of the coarse-grained bead, $\boldsymbol{R}_I = \sum_i c_i r_i$. If the center of mass is used in the mapping, eq. 2.34 reduces to the sum of the forces.
By calculating the reference forces for L snapshots and by imposing the condition that coarse-grained forces should match the reference forces for each snapshot, $N \times L$ equations can be written

$$\boldsymbol{f}_{Il}^{\mathrm{cg}}(g_1, ..., g_M) = \boldsymbol{f}_{Il}^{\mathrm{ref}},\ I = 1, \ldots, N, l = 1, \ldots, L\ . \tag{2.35}$$

Here $\boldsymbol{f}_{Il}^{\mathrm{ref}}$ is the force on the bead I, $\boldsymbol{f}_{Il}^{\mathrm{cg}}$ is the coarse-grained representation of this force. Index l enumerates snapshots picked for coarse-graining. By running the simulations long enough one can always ensure that $M < N \times L$. In this case the set of equations 2.35 is overdetermined and can be solved in a least-squares sense.
Though the underlying idea of FM is straight forward, implementation wise it is the most complicated method. Going back to the set of equations 2.35 one can see that $\boldsymbol{f}_{il}^{\mathrm{cg}}$ is, in principle, a non-linear function of its parameters $\{g_i\}$. It is, therefore, useful to represent the coarse-grained force-field in such a way that equations 2.35 become linear functions of $\{g_i\}$. This was originally done using cubic splines to describe the functional form of the forces [46]. However, this imposes additional constraints for spline smoothness conditions [74] on the set of linear equations which then has to be solved in a constrained least-squares sense. Step functions can be used instead but this requires a finer grid.
An adequate sampling of the system requires a large number of snapshots L. Hence, the applicability of the method is often constrained by the amount of available memory. To remedy the situation, one can split the trajectory into blocks, find the coarse-grained potential for each block and then perform averaging over the blocks.

2.2.6 Coarse-graining and atomistic morphologies

Since a coarse-grained model only represents a simplified description of the system, a question that might arise is how can such a model help to obtain an equilibrated morphology at an atomistic level.
In fact, coarse-graining reduces the computational costs due to two reasons. First, due to the reduced number of degrees of freedom it takes less computer time for the integration of the equations of motion. Second, using the coarse-grained model results in a speedup of the intrinsic timescales (e.g. the diffusion constant) [42, 63]. Once the system is equilibrated on a coarse-grained level, atomistic details can be reintroduced if the coarse-grained mapping represents the structure of the atomistic molecule and the coarse-grained model

performs correct sampling of correlations. After reintroduction of atomistic details, only a short simulation is required to locally equilibrate the system and compensate for inaccuracies of the coarse-grained model [75]. In total, a speedup of three orders of magnitudes is feasible. In this context, coarse-graining should not be seen as a substitution for atomistic models, but rather as an extension which, when used in combination, extends the accessible range of time- and length-scales.
To summarize, the methods mentioned so far can be used to obtain large scale morphologies at an atomistic level. The theoretical background to link charge carrier mobility to these morphologies is described in the following chapter, where all methods will finally be combined into a global scheme.

Chapter 3

Mobility simulations in organic semiconductors

Charge transport in organic semiconductors differs significantly from transport in metals. In metals, electrons are delocalized and charge carrier mobility is determined from their effective mass and the mean relaxation time of the band states [32]. In organic semiconductors, electrons and holes are localized at ambient conditions: in perfectly ordered crystals due to dynamic disorder (fluctuations of the coupling element are of the same order of magnitude as its value [76]), and in partially disordered systems due to static disorder. In the latter case, charges localize on individual molecules or, generally speaking, on charge transporting units (conjugated segments in the case of polymers, or a part of a molecule [77]). In the limit of strong localization, transport can be described as hopping of charge carriers. In highly ordered materials or crystals, the models for charge transport have to be adjusted. For organic crystals at low temperatures, band transport theories corrected for electron-phonon coupling can be applied [78, 79]. At ambient conditions, semi-classical dynamics can be used [80].

In this chapter, an overview of hopping transport is given, including the theoretical background and practical considerations for computer simulations. Special attention is paid to the link between molecular structure, material morphology and charge mobility. It is assumed that charge transport occurs on timescales much shorter than molecular dynamics which leads to changes in morphology[1]. Therefore, atomistic morphology can be obtained using molecular dynamics simulations and assumed to be frozen during simulation of charge dynamics. For this morphology, a graph for hopping transport is

[1]Local vibrations of molecules, which occur on shorter timescales, are accounted for in the derivation of the Marcus expression for charge transfer.

created by (i) determining hopping sites, e.g. conjugated molecules or their conjugated fragments (ii) calculating the electronic coupling between them, and (iii) using Marcus theory to obtain rates. Then, charge dynamics is simulated using the kinetic Monte Carlo method. This model was previously used to correlate structure and charge mobility in hexabenzocoronene derivatives, which are small conjugated molecules [81, 82, 83, 84, 85, 86, 25].
First, the theory of electron transfer is described and Marcus theory is introduced. Next, methods to estimate the parameters of the Marcus rate equation are given, followed by an outline of the kinetic Monte Carlo algorithm. In sec. 3.7, the aforementioned model is discussed in detail and extended to conjugated polymers.

3.1 Electron transfer

The theory of electron transfer was first developed to study electron transfer in chemical reactions, such as isotopic exchange reactions of iron in aqueous solution [87]

$$\mathrm{Fe}^{2+} + \mathrm{Fe}^{*3+} \rightarrow \mathrm{Fe}^{3+} + \mathrm{Fe}^{*2+} , \tag{3.1}$$

where the asterisk denotes a radioactive isotope. Since hopping transport invokes many transfer "reactions", this formalism can be employed to describe charge dynamics in organic semiconductors.
Electron transfer can be considered as a spontaneous charge redistribution between an initially prepared reactant state, and a well defined product state [88] or, specifically, the transition of a single electron from one localized orbital to another. The two localized orbitals are referred to as the donor (D) and acceptor states (A). The transfer reaction can then be written as

$$D^- A \rightarrow D A^- , \tag{3.2}$$

where D^- denotes the electron being localized on the donor, and A^- the electron being localized on the acceptor.
Since the interactions between molecules are weak in organic semiconductors, the donor $|1\rangle$ and acceptor $|2\rangle$ states can be approximated using (non-interacting) molecular orbitals, or diabatic states. The electronic Hamiltonian of the charge transporting complex can then be written as

$$H_{el} = E_1 |1\rangle \langle 1| + E_2 |2\rangle \langle 2| + J \left(|1\rangle \langle 2| + |2\rangle \langle 1| \right) , \tag{3.3}$$

with the energies E_1 , E_2 of the individual states and the so called *transfer integral* or electronic coupling element J for the two states. The limit of small

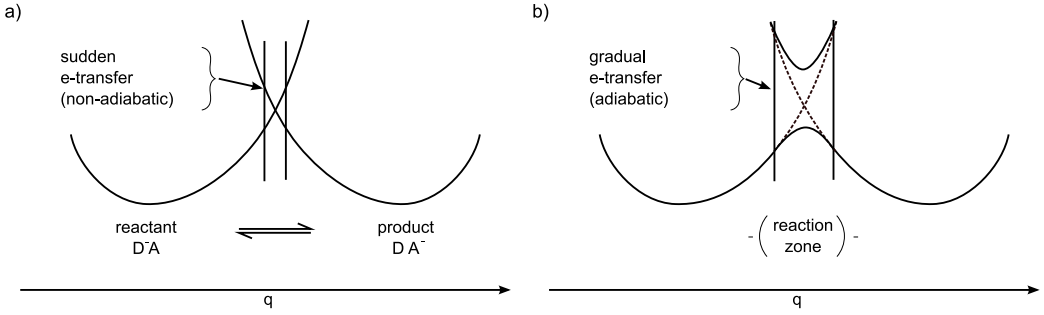


Figure 3.1: **Adiabatic and non-adiabatic limits for charge transfer reactions.** Non-adiabatic (weak coupling) and adiabatic (strong coupling) transfer is depicted. Taken from ref. [89].

and strong coupling are referred to as *non-adiabatic* and *adiabatic* electron transfer, respectively.
To describe the transfer reaction, which is coupled to the nuclear motion, the *reaction coordinate* q, which connects the donor and acceptor states, is introduced. It is related to the coordinates of the nuclei. With the help of q, the two regimes of non-adiabatic and adiabatic transport are schematically represented in figure 3.1. Non-adiabatic transfer is a sudden tunneling process, and follows the Frank-Condon principle [90]. For weakly interacting systems, the diabatic representation provides a good zeroth-order basis set [88]. However, the diabatic crossing becomes increasingly pronounced for stronger coupling, and the transport process cannot be described by two non-interacting states. Therefore, adiabatic transfer is usually described in terms of chemical reaction kinetics using a double well potential [88] (see figure 3.1(b)). The internal energy of the reaction is considered with respect to the reaction coordinate q, and to overcome the barrier between initial and final states, thermal activation is required.

3.2 Marcus theory of charge transfer

The rates for a charge transfer reaction in the non-adiabatic high temperature limit can be calculated using Marcus theory [87]. It is assumed that a transition is initiated by thermal fluctuations of the nuclei, and that the Frank-Condon principle is fulfilled. The classical derivation provides a rate equation which contains an empirical prefactor. A semi-classical derivation was later given by Jortner *et al.* [90], which is based on Fermi's golden rule. What follows is a brief recapitulation of the key points of this derivation in the classical limit, complementing the original formula with an expression for the prefactor.

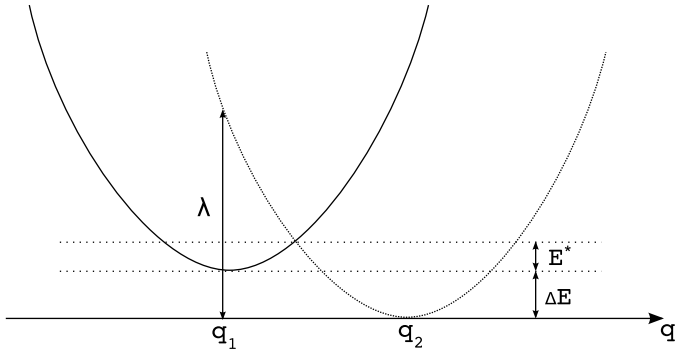


Figure 3.2: **Diabatic energy profiles as a function of the reaction coordinate q for the electron-transfer reaction.** The difference in energy of the initial and final state is given by ΔE. After the electron transfer occurred, the nuclei relax to the new equilibrium. The corresponding change in energy, λ, is called reorganization energy.

Two additional assumptions are needed to obtain the rate equation: first, the transfer integral should not depend on the nuclear coordinates q (vibrations of the molecule). Second, it is assumed that the potential energy surfaces for the initial and final states are harmonic with identical curvatures and nuclear vibrations are treated classically. Within these assumptions, the transfer process is depicted in figure 3.2 and can be formulated in terms of a polaron model with diagonal electron-phonon coupling. The total Hamiltonian of eq. 3.3 is then extended to

$$\begin{aligned} H = &|1\rangle\langle 1| \left\{E_1 + 1/2 f\left(q - q_1\right)^2\right\} \\ &+ |2\rangle\langle 2| \left\{E_2 + 1/2 f\left(q - q_2\right)^2\right\} \\ &+ J\left(|1\rangle\langle 2| + |2\rangle\langle 1|\right) \end{aligned} \tag{3.4}$$

The first and the second terms are the energies represented by the left and right parabolas in figure 3.2. Vibrations are characterized by a single propagating mode with force constant f and equilibrium displacements q_1 and q_2. Since the transfer integral is small in the non-adiabatic limit, it can be treated as a perturbation to the non-interacting donor and acceptor states. With the help of Fermi's golden rule, the total rate within the high temperature limit is given as an integral, i.e. the average over all initial vibrational coordinates, with each term weighted by the Boltzmann probability of being in that state

$$\omega_{ij} = \frac{2\pi}{\hbar} \int dq f(q) \left|J\right|^2 \delta\left[U_1(q) - U_2(q)\right] , \tag{3.5}$$

$$\tag{3.6}$$

where

$$f(q) = \frac{1}{Z} e^{-U_1(q)/k_b T}\,, \tag{3.7}$$

$$U_i = (q) = E_i + 1/2 f\,(q - q_i)^2 \qquad , \qquad i = 1, 2\,. \tag{3.8}$$

The Marcus equation is obtained by solving the integral in eq. 3.5, and is given by

$$\omega_{ij} = \frac{J_{ij}^2}{\hbar} \sqrt{\frac{\pi}{\lambda_{ij} k_\mathrm{B} T}} \exp\left[-\frac{(\Delta G_{ij} - \lambda_{ij})^2}{4\lambda_{ij} k_\mathrm{B} T}\right], \tag{3.9}$$

where k_B is the Boltzmann's constant, $\hbar$ is the Planck constant, λ_{ij} is the reorganization energy, ΔG_{ij} is the difference in free energies of the sites i and j, and J_{ij} is the corresponding transfer integral. The square root in eq. 3.9 is the Franck-Condon factor in the classical limit and the energy in the exponential is the height of the crossing barrier E^*, depicted in figure 3.2. In the following sections, methods to estimate these parameters are outlined.

3.3 Computational methods for evaluating transfer integrals

In the preceding section, the Marcus equation for calculating charge transfer rates was introduced. An important parameter in this equation is the electronic coupling element, or transfer integral J. For equal site energies, $E = E_1 = E_2$, the eigenvalues $E_{+,-}$ of the Hamiltonian H in eq. 3.3 read

$$E_{+,-} = E \pm |J|\,, \tag{3.10}$$

i.e. J is half the orbital energy splitting, ΔE_{orb}, of the charge transporting complex [89]

$$|J| = \frac{1}{2}\,(E_+ - E_-) = \frac{1}{2}\Delta E_{orb}\,. \tag{3.11}$$

where E_+ and E_- denote the energies of the corresponding orbitals.
To calculate the orbital splitting, one quantum chemical calculation has to be performed for each possible transfer process, that is one for each *pair* of molecules. This can be computationally demanding for bigger systems. Also, in the case of different site energies, which can arise due to different types of molecules or polarization effects, additional correction terms are needed [91]. An alternative approach to calculate transfer integrals was proposed by J. Kirkpatrick [92]. Within the frozen orbital approximation, the transporting orbitals $|1\rangle$ and $|2\rangle$ are assumed to be the HOMOs (in the case of hole transport) of each monomer unit, representing the diabatic states of the charge

transporting complex. When the Fock matrix eq. 3.3 is constructed from these monomer states, the transfer integral J is given by its off-diagonal elements, i.e.

$$J = \langle 1| \hat{H} |2\rangle \ . \tag{3.12}$$

In the practical implementation, the transporting orbitals are computed using the semi-empirical INDO method with Zerner's parametrization (ZINDO) [93]. Since this method allows to calculate the transfer integrals based on a weighted overlap of molecular orbitals, it is also referred to as the molecular orbital overlap (MOO) method. The advantage of the MOO method is that the explicit computation of the density matrix for a pair of molecules is not required but can be constructed based on the relative geometry of the two molecules and the transporting orbitals of the isolated molecules. Therefore, only a single ZINDO calculation for each *type* of molecule is needed as input from quantum chemistry. In addition, the overlap integrals for atomic orbitals can be precalculated and stored in tables, which additionally improves the performance of the method. The MOO method is used to calculate transfer integrals throughout this work.

3.4 Reorganization energy

After sudden electron transfer occurs, the system undergoes a structural change to relax its nuclear degrees of freedom to the new energetic minimum. This change in energy is referred to as the *reorganization energy*. Mathematically, the reorganization energy is defined by

$$\lambda = \frac{1}{2} f \left(q_1 - q_2\right)^2 \ , \tag{3.13}$$

where f is the force constant of the harmonic oscillator in eq. 3.4 and q_1, q_2 the corresponding minima of the initial and final states.
The reorganization energy can be directly obtained using quantum chemical methods (e.g. DFT) by calculating the energy difference of the molecule upon relaxation. The reorganization energy is given by the total change in energy of the donor and the acceptor states

$$\lambda = (E_n^c - E_n^n) + (E_c^n - E_c^c) \tag{3.14}$$

where the left (right) term is the contribution of the donor (acceptor) and E_n^c, E_c^n, E_n^n and E_c^c are single point energy calculations. E_n^c is the energy of a charged molecule with a structure optimized in the neutral state, E_c^n the energy of a uncharged molecule in charged conformation and E_n^n (E_c^c) the energy of a neutral (charged) molecule in neutral (charged) conformation.

The reorganization energy which results from the reaction within the charge transporting complex is referred to as the inner sphere reorganization energy. For polarizable systems, or systems with a polarizable solvent, the outer sphere reorganization energy, which accounts for rearrangement of solvent molecules has to be considered as well [94].

3.5 Site energies

The site energy difference ΔG_{ij} is the difference in energy between donor and acceptor states. Several contributions to ΔG_{ij} are conceivable. The two that are considered in this thesis are due to (i) the energies of transporting orbitals being different (HOMO/LUMO level for holes/electrons), (ii) an external electric field.
The first one can arise if different molecule types are involved in the transfer process. In the case of polymers, it can also be due to different lengths of conjugated segments [77]. Practically, it can be obtained using quantum chemical methods by forming the difference between the orbital energies of the respective HOMO/LUMO level for holes/electrons. The second term is needed, if the drift of charge carriers in an external electric field shall be simulated and can be calculated by the projection of the vector which connects the donor and acceptor onto the electric field vector.
Additional contributions, which are due to changes in the electrostatic surrounding, are important if the studied system shows significant electrostatic interactions, e.g. large dipole moments. For polarizable molecules, polarization due to the presence of charge carriers can be important.

3.6 Kinetic Monte Carlo

The Monte Carlo approach refers to a wide range of methods which solve problems in a stochastic way. First applications were Monte Carlo integration, or the Metropolis Monte Carlo algorithm [95], which can be used for canonical sampling of a system. In the 1960s, a new class of Monte Carlo algorithms was developed which also allowed time-evolution of the system [96]. These led to the Kinetic Monte Carlo (KMC) algorithm, which provides a stochastic solution of the master equation [97]. KMC was first applied in the field of organic electronics by Baessler *et al.* [98] to study charge transport in disordered organic semiconductors.
The KMC algorithm describes a Markovian process: a system is represented by discrete states, and the transition from state i to state j is modeled by a

rate constant ω_{ij}. A direct implementation of the KMC algorithm, which is often referred to as continuous-time KMC [99], is depicted in figure 3.3.

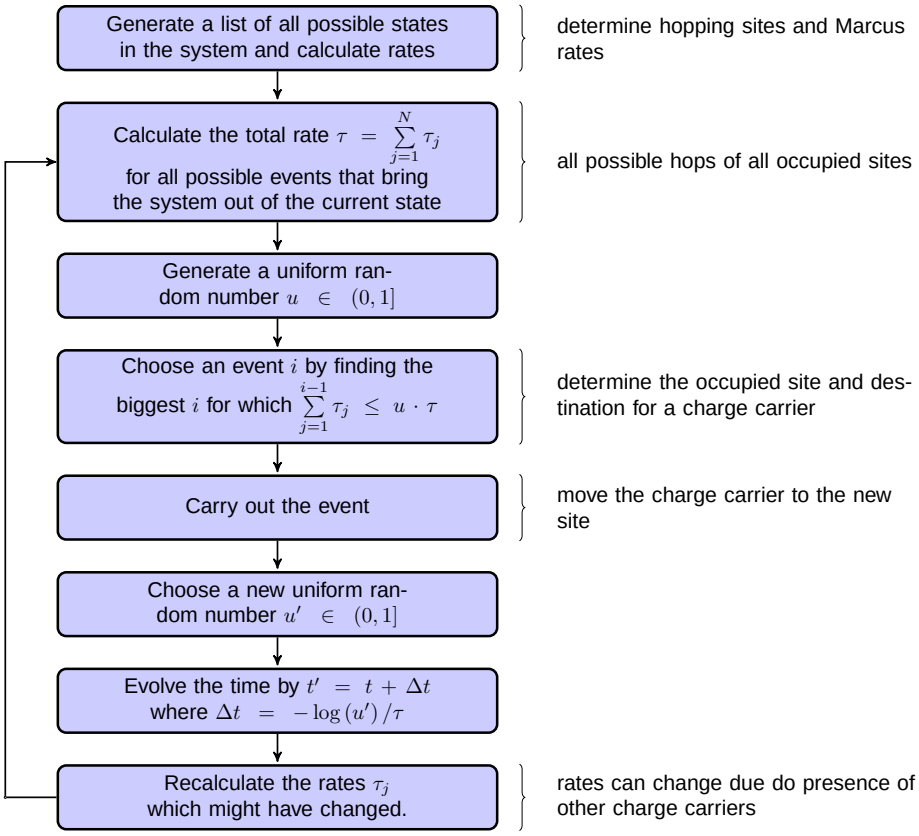


Figure 3.3: **The kinetic Monte Carlo algorithm.** The figure depicts the workflow of the kinetic Monte Carlo algorithm and how it is applied in charge transport simulations.

Hopping transport in organic semiconductors is an ideal process to be studied using the KMC algorithm. The current state of the system is given by certain sites which can be occupied by charge carriers. These sites are connected by transfer processes, with a rate corresponding to the charge transfer rate. Such a graph is depicted in figure 3.4. For the current occupation of sites, the total rate for all possible hops of all charges is calculated. Then the algorithm picks a charge and its next site based on a random number. The time in the system is then evolved by Δt based on the total rate τ and using a second random number u'

$$\Delta t = -\log(u')/\tau\,. \tag{3.15}$$

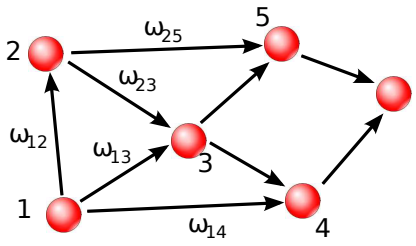


Figure 3.4: **Connectivity graph for the KMC algorithm.** Charge carriers are represented by sites that can be occupied. Links represent charge transfer processes which have certain rates ω_{ij}.

The different occupation of sites also affects the rate constants of other charges that did not move since sites cannot be doubly occupied: the newly occupied site does not accept further charge carriers while the former site is now unoccupied and offers a new destination for hops. In addition, if the Coulomb interaction of charge carriers is taken into account, the changing electrostatic environment and polarization effects might affect hopping rates. For low charge carrier densities, this is avoided by simulating a single charge. As mentioned before, KMC requires a graph consisting of connected sites as an input. Each connection in this graph represents a hopping process, which has a certain rate. In the next section, it is described how such a connectivity graph can be created from atomistic morphologies and how mobilities are calculated.

3.7 Charge transport in realistic morphologies

The Gaussian disorder model is probably the most successful and widely used model to simulate charge transport in disordered organic semiconductors [100, 98, 101, 102]. The underlying assumptions are Miller-Abraham rates for hopping between sites and a Gaussian distribution of site energies ϵ_i. Miller-Abraham rates are written as

$$\nu = \nu_0 \exp\left(-\Gamma_{ij}\frac{r_{ij}}{a}\right)\begin{pmatrix}\exp\left(-\frac{\epsilon_j-\epsilon_i}{k_BT}\right) & \text{for} & \epsilon_j G > \epsilon_i \\ 1 & \text{for} & \epsilon_j < \epsilon_i\end{pmatrix}, \tag{3.16}$$

where Γ_{ij} is the inter-site coupling, ν_0 the attempt-to-escape frequency, r_{ij} the distance between the sites, and a the average lattice distance. This model has, however, significant drawbacks: (i) fitting parameters are involved in order to match experimental data and (ii) regular grids are used, where each node represents a hopping site.

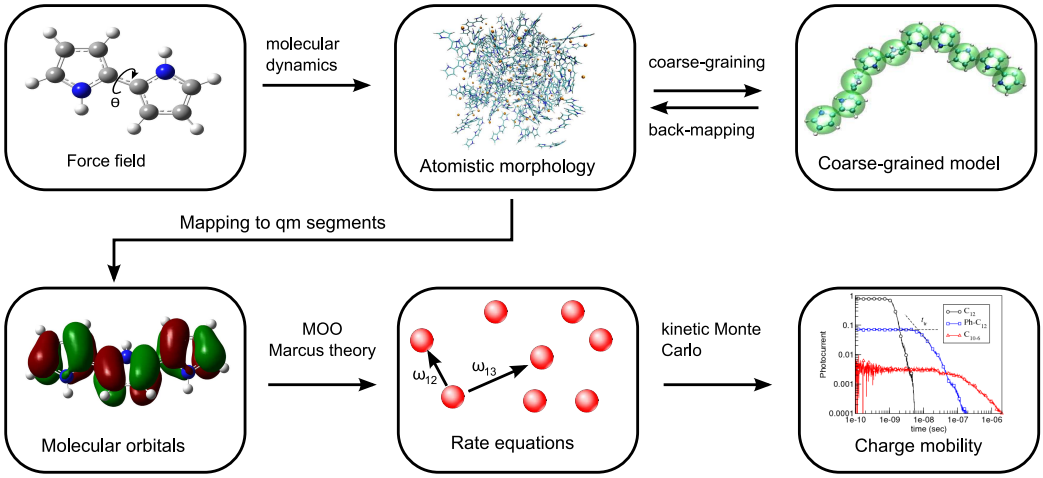


Figure 3.5: **Framework to relate atomistic morphologies to charge carrier mobilities.**

A problem of Miller-Abraham rates is that they only depend on the distance of two molecules but not on their local orientations. Here, Marcus rates can be used where the transfer integral accounts for the local orientation of molecules. Furthermore, the global arrangement of molecules is needed to study the path a charge travels through the sample. This cannot be achieved by regular grids, instead, hopping sites should be determined explicitly in the morphology. The methodology followed in this work is based on previous work on charge transport in discotic liquid crystals [81, 82, 83, 84, 85, 86, 25] and avoids the use of fitting parameters and predefined grids for charge hopping. An outline of the method is provided in figure 3.5.

Morphologies are obtained in atomistic simulations. For novel organic compounds, force field parameters are often not available. They can be parametrized starting from the force field for similar compounds and quantum chemical calculations. Force-fields can be checked by comparing structures obtained in simulations to x-ray scattering and solid-state NMR data, if available. In some cases, time- and lengthscales of atomistic simulations are insufficient, and additional coarse-grained simulations followed by a back-mapping procedure to reintroduce atomistic details [75] can be beneficial as pointed out in chapter 2.

After atomistic simulations are completed, hopping sites which represent diabatic states are determined. For molecules with a well-defined conjugated core, such as HBC (see figure 1.1), this is straight-forward. HBC is a disk-like molecule with side chains that make it soluble in organic solvents. Since the charge transporting orbitals are localized on the core, side chains can be neglected and the entire molecule represents a hopping site. To rein-

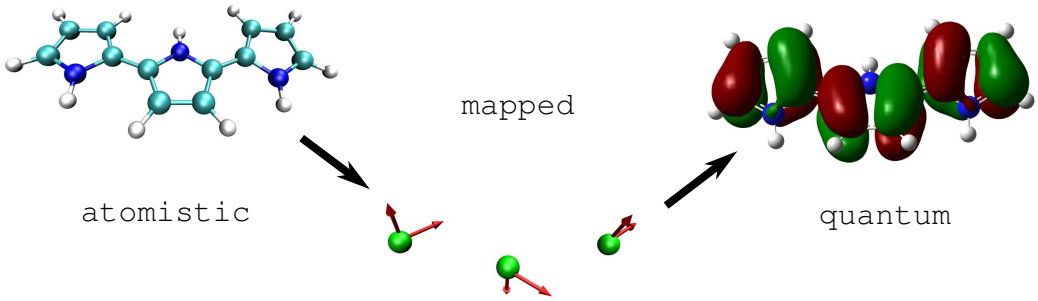


Figure 3.6: **Substitution of an atomistic chain with a QM image.** Each repeat unit is considered to be rigid. In an intermediate step, they are represented by their positions and orientations. The fragments of quantum orbitals are then adjusted to match the positions and orientations of the repeat units.

troduce quantum mechanical orbitals, the positions and orientations of the conjugated cores are calculated and then substituted by rigid copies of structures optimized in quantum chemical calculations (bottom left of figure 3.5). Atomic orbitals, needed to calculate transfer integrals, are provided together with the optimized structures. This step can be seen as back-mapping to quantum chemical details. The benefit of the rigid-copy substitution is that bond length fluctuations are neglected which otherwise violate the diabatic picture of charge transfer.
Transfer integrals are calculated using the previously mentioned MOO method and the corresponding rate equations are then solved using the kinetic Monte Carlo algorithm. One can either mimic a time of flight experiment or calculate the mobility from the average drift velocity of charge carriers in an external electric field (bottom right of figure 3.5). The drift velocity is calculated by averaging the waiting time Δt multiplied by the distance of the hop $\boldsymbol{r}_{ij}$ during the entire KMC run, using periodic boundary conditions.
The aforementioned approach cannot treat conjugated polymers. First, for more complex molecules or polymers, a simple substitution of a whole molecule is not always possible since soft degrees of freedom (e.g. torsions) can lead to big differences in structure. Second, a single molecule (polymer chain) can have several parts where charges can localize (conjugated segments). In this work, the approach is generalized in order to include these two situations.
To compensate for big structural changes, a molecule is split into *rigid fragments*, which are parts of the molecule, e.g. one repeat unit in a polypyrrole chain. When copies from quantum chemical calculations are inserted, each fragment is reoriented to match the conformation of the atomistic simulations. This procedure is demonstrated in figure 3.6. Using atomic orbitals

and assuming that this change in geometry does not affect the atomic orbital coefficients, the MOO method can be used to calculate transfer integrals. Details on the implementation of rigid fragments is given in sec. 4.4.
To take into account that charges can localize on parts of the molecule, they are partitioned into conjugated segments and each segment is treated as a hopping site. Charge transport is then described by hopping between such conjugated segments, either within the molecule, or to a segment on a different chain. An important difference for intra-chain transport is that transfer integrals are big and an expression for adiabatic rates has to be used instead of Marcus rates. This is discussed in detail in chapter 6, where the model is used to study charge transport in polypyrrole oligomers.

Chapter 4

The VOTCA package

The Versatile Object-oriented Toolkit for Coarse-graining Applications [103] (VOTCA)[1] is a software package initially developed to simplify particle-based systematic coarse-graining[2]. Currently, Boltzmann inversion, iterative Boltzmann inversion, inverse Monte Carlo, and force-matching are implemented and tested. At a later stage, additional modules for charge transport calculations, as described in sec. 3.7, and interfaces to the Molecular Orbital Overlap [92] and KMC codes written at Imperial College, London, were added.

The development of such a package is only possible due to the extensive use of modular software design, making VOTCA a flexible platform for the implementation of new methods. The consistent implementation of all methods allows their direct comparison and simplifies the coarse-graining workflow. The coarse-graining tools of the package have been released for public use under an open source license[3], with the status paper published in Journal of Chemical Theory and Computation [103]. The charge transport code is used internally, and its release is planned.

The package is written as a combination of modular C++ code and shell scripts and is easy to extend for new methods and interfaces to other programs. In this chapter, the core design of the package and its main classes are outlined. A brief description of the terminology used, namely the unified modeling language (UML), is given in appendix A.1. The structure of the package is outlined in appendix B.

[1] http://www.votca.org

[2] The VOTCA package is a joint development. The focus of my work was the concept and design of the package, the C++ core development and the implementation of the inverse Monte Carlo method. The scripting framework was mainly developed by Christoph Junghans. Alexander Lukyanov was responsible for the implementation of the force-matching method.

[3] Apache License, Version 2.0

4.1 Coarse-graining engine

The non-iterative coarse-graining methods described in sec. 2.2 derive coarse-grained potentials by analyzing the canonical ensemble of a reference (high resolution) system. If an iterative procedure is used, additional canonical sampling of the coarse-grained system is required. In both cases, sampling can be done using either molecular dynamics (MD), stochastic dynamics (SD), or Monte Carlo (MC) techniques which are implemented in standard simulation packages. Rather than implementing its own MD/SD/MC modules, the toolkit allows swift and flexible integration of existing sampling programs. The analysis needed for coarse-graining and workflow control in the case of the iterative methods is done using the package tools.
The implementation of the analysis kernel follows the rules of modular software design and is written in C++. Common software design patterns are applied to offer the flexibility that is required for an expandable toolkit. The analysis kernel provides structures for topology handling, mapping from atomistic to coarse-grained level, as well as interfaces for reading and writing data of external sampling programs. These structures are similar to those of an MD code. However, since an integration of the equations of motion is not needed, implementation of functional forms for potentials and forces, thermostats, integrators, etc. can be avoided. Another benefit is that most of the analyses are not computationally expensive compared to the sampling of the system, and thus parallelization of the algorithms and data structures is not needed[4].
The coarse-graining procedure is controlled by several Extensible Markup Language (XML) input files, which contain mapping definitions and other options required for the workflow control. The current tools include functions to calculate probability distributions of bonded and non-bonded interactions, correlation and autocorrelation functions, as well as updates for the coarse-grained pair potential. If needed, analysis tools of the package used for sampling can also be integrated into the coarse-graining workflow.

4.2 Core design

An unified modeling language (UML) diagram[5] of the key classes of VOTCA is depicted in figure 4.1. In what follows, classes are emphasized in bold.

[4]The only time-critical part in the analysis is the neighbor search when e.g. calculating radial distribution functions. However, parallelization of the latter is trivial and can be done on script level using time splitting.

[5]A quick introduction to UML diagrams can be found in appendix A.1

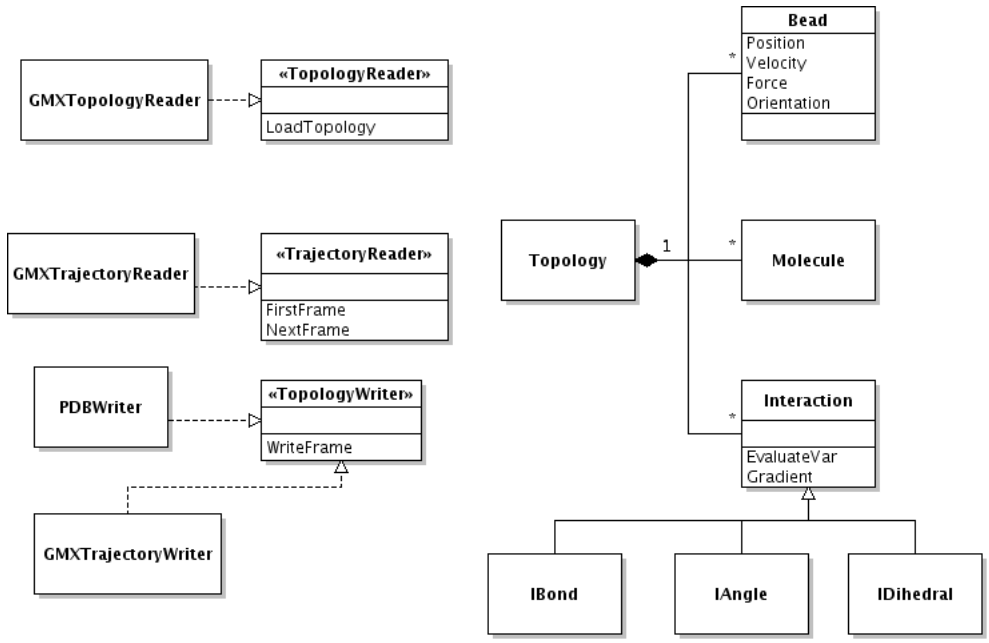


Figure 4.1: **VOTCA classes.** The UML class diagram depicts the core VOTCA classes for topology handling and reading data. The notation of UML class diagrams is outlined in appendix A.1.

The **Topology** class is the central element and serves as a container for information about all beads, molecules and interactions. In addition, it stores properties of the system, e.g. box size or simulation time. The **Bead** class, which can be an atom in the atomistic representation or a coarse-grained bead, contains details on beads. The class contains name, type and position of a bead and can optionally handle orientation, velocity and force, if they are provided in the trajectory. Forces are required for the force-matching method, orientations, which are calculated for non-spherical beads, are used for mapping from atomistic to quantum representation as will be discussed later in sec. 4.4.

The **Interaction** class provides the functionality for bonded interactions. The explicit functional form of the potential is not essential; only the evaluation of the interaction variable, denoted as κ, needs to be performed. κ can be a bond length, angle or dihedral angle. In addition, force matching requires the direction of the force which acts on bead i due to interaction α.

This is given by the gradient of κ with respect to the coordinates of bead i:

$$\begin{aligned} \boldsymbol{F}_i &= -\boldsymbol{\nabla}_i U_\alpha\left(\{r_n\}\right) = -\boldsymbol{\nabla}_i U_\alpha\left(\kappa_\alpha\left(\{r_n\}\right)\right) \\ &= -\frac{\partial U_\alpha}{\partial \kappa}\boldsymbol{\nabla}_i \kappa_\alpha\left(\{r_n\}\right) . \end{aligned} \tag{4.1}$$

The derived classes **IBond**, **IAngle** and **IDihedral** inherit from the **Interaction** class and specialize the evaluation of κ as well as the gradient. For the evaluation of non-bonded interactions, IBI only needs the radial distribution functions and, to perform IMC, the correlations. In case of force-matching, the vector that connects two coarse-grained beads is required. Both can be evaluated using neighborlists, which are either computed via a simple $O(N^2)$ search or a grid search algorithm. The latter scales linearly with the number of particles.

To allow for easy implementation of new file formats and/or different algorithms, for example complex, nonlinear mapping schemes, the object factory pattern [104] is applied in these cases. This guarantees a clean separation of sampling engine dependent components and algorithms used for the coarse-graining process. The implementation of the object factory is based on templates as proposed in ref. [105]. The idea of an object factory is explained in appendix A.2.

The analysis and mapping process is steered by the class **CGEngine** (see figure 4.2). It handles reading of the topology and trajectory, as well as mapping from the atomistic to the coarse-grained resolution, if needed. Here, the mapping functionality is split into several layers. The highest layer is the **TopologyMap**, which contains information on how to map the entire reference topology to the coarse-grained level. It is composed of mapping definitions for each molecule in **MoleculeMap**, which again contains mapping information for each bead in **BeadMap**. Currently two types of bead maps for a linear mapping scheme as given by eq. 2.7 exist. The first one is for spherical beads (**Map_Sphere**) and the second for ellipsoidal beads (**Map_Ellipsoid**), which calculates the orientation of a bead based on the tensor of gyration and positions of selected atoms. Orientations can be used to calculate nematic order parameters and are needed for the mapping process during the charge transport simulations as described in sec. 4.4.

The actual analysis is performed using classes derived from the class **CGObserver**. They are registered at the **CGEngine** and are called whenever new data is available. This reduces the overhead when implementing new functionality and provides a consistent interface for all coarse-graining tools. Boltzmann inversion, evaluation of IMC parameters, and force matching are all implemented as the corresponding observers.

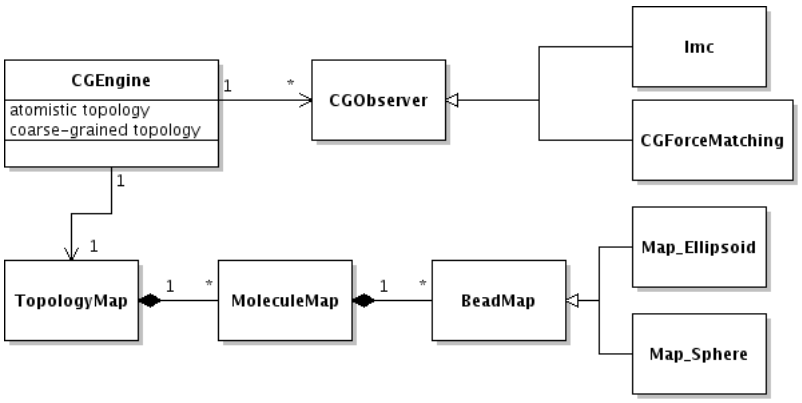


Figure 4.2: **Class diagram for coarse-graining and mapping engine.** The **CGEngine** is the core class which encapsulates the whole process of reading in data, mapping to coarse-grained level and calling analysis classes derived from **CGObserver**.

4.3 Iterative workflow control

The workflow of the iterative procedures is implemented as a set of shell scripts which can, in principle, be run on all available operating systems. The iterative workflow control provides the flexibility to overload existing or call external scripts and programs written in other programming languages. An interface to read values from the steering XML files is provided in C++, Perl and shell. A chart of the workflow is depicted in figure 4.3.
During the global initialization, the initial guess for the coarse-grained potential is calculated from the reference radial distribution function or converted from a given potential guess to the internal format. The actual iterative step starts with an iteration initialization. It searches for possible checkpoints and copies and converts files from the previous step and the base directory to the current step. Then, the simulation run is prepared by converting potentials to the format required by the external sampling program and the sampling run is started. Currently, an interface for GROMACS [37] is implemented. An extension to other packages is straightforward. After sampling the phasespace, the potential update ΔU is calculated. Often the update requires post-processing, such as smoothing, interpolation, extrapolation or fitting to an analytical form. For example a linear pressure correction term, as described in ref. [56], can be seen as post-processing of ΔU, since it solely adds a linear function to ΔU. Finally, the new potential is determined and post-processed. The script either stops after a certain number of iterations or if a convergence criteria is fulfilled. Otherwise, the iterative process continues with the initialization of the next iterative step.

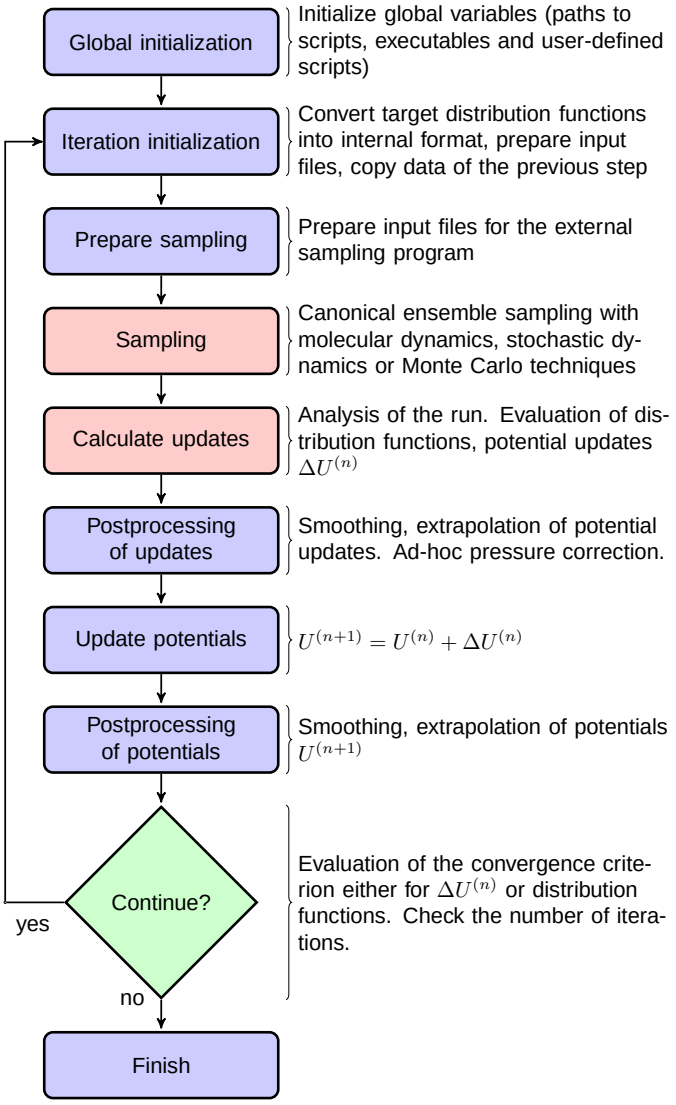


Figure 4.3: Block-scheme of the workflow control for the iterative methods. The most time-consuming parts are marked in red.

4.4 Charge transport modules

The implementation of the methods used for the charge transport studies, as described in chapter 3, is based on the framework of the analysis core. Codes to evaluate transfer integrals using the molecular orbital overlap method and running Kinetic Monte Carlo simulations already exist from previous implementations. However, the interface to connect molecular dynamics topologies with transfer integral calculations, create a set of rates, and finally run a KMC simulation, was redesigned based on the existing VOTCA structures of the coarse-graining engine.

The charge transport calculations require structures to organize atoms in topologies and molecules, similar to those implemented in the coarse-graining core: topology, molecules and atoms have to be defined. However, additional functionality is needed to handle charge transport properties such as site energies, reorganization energies and atomic orbitals from quantum chemical calculations. To avoid duplication of code, the classes **QMTopology**, **QMBead**, and **QMMolecule** are derived from the existing classes **Topology**, **Bead** and **Molecule**, respectively, with additional properties for the charge transport calculations and the interface to the MOO code for the evaluation of transfer integrals. This allows the integration of VOTCA's topology handling, neighbor search algorithms, as well as algorithms for triclinic periodic boundary conditions in the charge transport calculations. The UML diagram of the basic structures is depicted in figure 4.4.

The classes **CrgUnit**, which represents a hopping site, and **JCalculator**, to evaluate the corresponding transfer integrals, are provided in the MOO library and are embedded into existing structures. In the previous implementation, which was used to study charge transport in discotic liquid crystals [81], a charge unit always represented a whole molecule. However, as discussed in sec. 3.7, this approach is not suitable for polymers or molecules which consist of several rigid fragments that have a certain flexibility with respect to each other (see sec. 3.7). The current implementation uses the **QMBead** class to represent the position and orientation of such a rigid fragment. A hopping site can be constructed using several of these rigid fragments (see figure 4.5). In addition, a molecule can have several hopping sites as it is the case for polypyrrole (see chapter 6), where transport is modeled by hopping between conjugated segments.

In order to establish a network for charge hopping, a neighbor list has to be generated. If the shape of hopping sites differs significantly from that of a sphere, e.g. in rod like molecules, using the center of mass of whole units can be problematic: the cut-off threshold of the neighbor search has to be big enough to take into account neighbors in the direction of the long axis of the

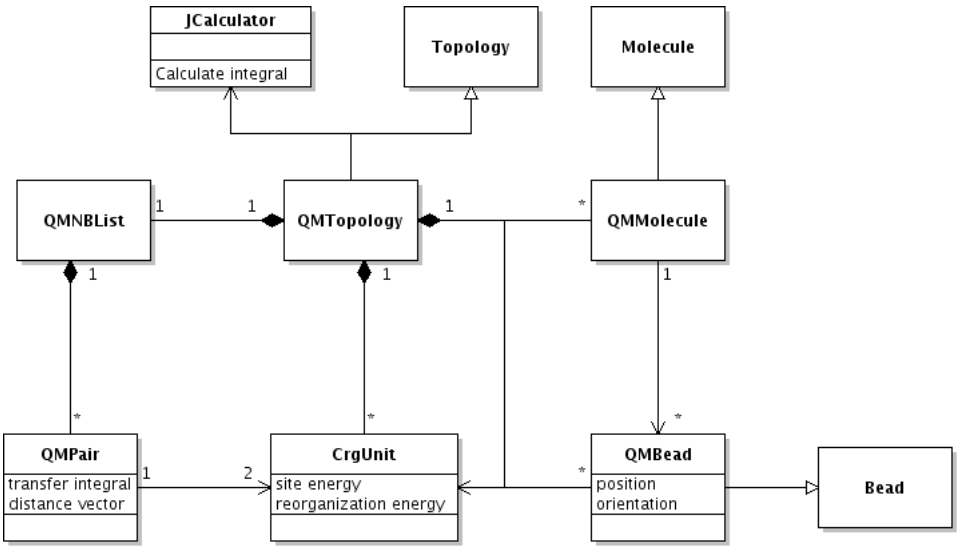


Figure 4.4: **UML class diagram of important MD2QM classes.** A **CrgUnit** represents a hopping site where charges can be localized on. A charge unit is composed of several rigid fragments, each having a certain position and orientation. Fragments are described by the class **QMBead**. A **QMMolecule** can contain several charge units or only a single hopping site.

molecule. This eventually leads to counting of second nearest neighbors in the direction of the short axis. A better approach is to analyze which parts of the molecule are at a distance less than a specific cutoff. This is implemented by creating a neighborlist of rigid fragments (**QMBead**): if two **QMBeads** from different charge units are at a distance less than a certain cutoff, the whole unit is added to the list of possible hops. A second benefit of this procedure is that it allows the use of the same neighbor search algorithms and classes which are implemented in the coarse-graining core.

All steps of the charge transport calculations (mapping from atomistic to rigid copies, calculation of transfer integrals and rates, running kinetic Monte Carlo) are implemented as separate applications. The first application in the workflow maps from an atomistic topology to rigid copies, whose geometries were optimized using a quantum chemical method. The mapping program is implemented as a **CGObserver** and creates a **QMTopology** based on the atomistic input. The **QMTopology** is then written to a file. It contains **QMBead**, **CrgUnit** and the neighbor list information. Subsequently, all further steps are then outlined by (i) reading in the state file, (ii) processing the **QMTopology** data, and (iii), writing out a new state file. To

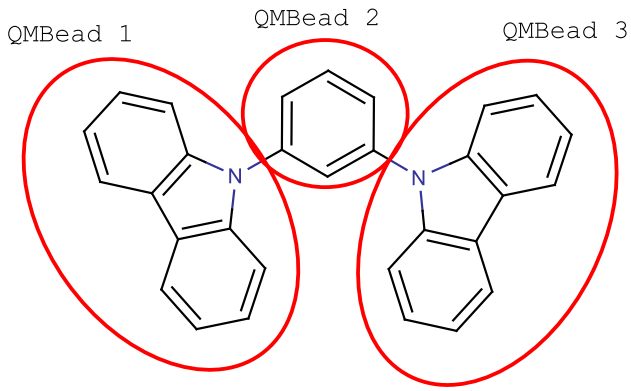


Figure 4.5: **The concept of rigid fragments.** The figure depicts the concept of rigid fragments (**QMBead**) to calculate molecular orbital overlap for the case of a whole mCP molecule considered as a charge unit. Since the side units can rotate around the bond which attaches them to the core of the molecule, the charge unit is constructed based on rigid subsets of the molecule, namely the QMBeads, each of which carries a position and orientation.

avoid duplication of code, the steps (i) and (iii) are encapsulated in the class **QMApplication**. Similar to the **CGObserver**, an application can be inherited from this class and only the functions to perform calculations (e.g. evaluate transfer integrals, calculate rates, run KMC) need to be implemented.

The approach described above requires a separate program for each calculation to allow for customization in the workflow (e.g. use a different method to calculate rates). This can be a cumbersome task, since many intermediate steps are involved. For this purpose, the **QMCalculator** was designed which defines an interface for classes which act on the **QMTopology** as depicted in figure 4.6. The processing can then be divided into three categories: It can act on the whole topology (e.g. electrostatic calculations), on pairs of charge units (**PairCalculator**, e.g. transfer integrals, rate calculations), or on single hopping sites (**SiteCalculator**). In combination with an object factory, this concept allows to bundle several calculations into a single program while still maintaining the flexibility needed to replace algorithms based on input files.

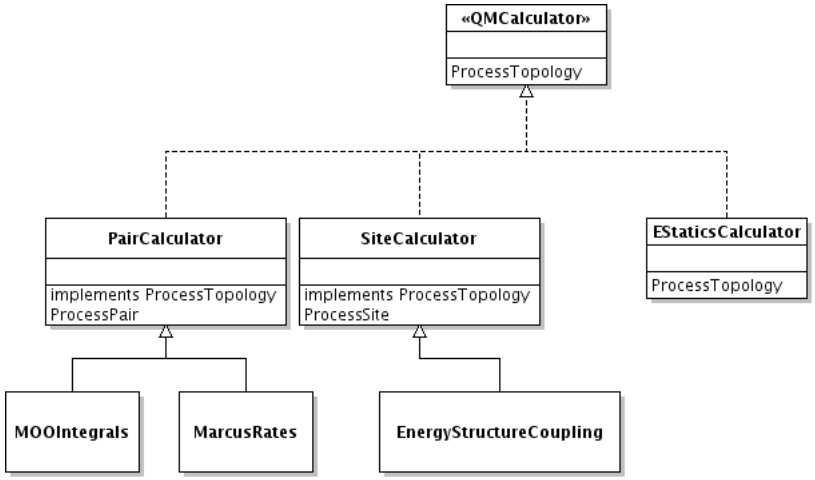


Figure 4.6: **QMTopology processing.** Classes that manipulate or analyze the **QMTopology** are all implementations of the **QMCalculator** interface which allows the uniform treatment of algorithms in an object factory.

4.5 Summary

A package for systematic coarse-graining was developed and released under an open source license. The package provides a platform for the consistent implementation of coarse-graining methods and thereby allow for their direct comparison. Currently, Boltzmann inversion, IBI, IMC, and force-matching are available, and an implementation of a simplex algorithm [106] is in progress. Due to the use of modular software design patterns, the package is easy to extend. For example the charge transport simulations performed in chapter 6 are implemented on the basis of this framework.
In the next chapter, the package functionality will be demonstrated by coarse-graining a set of reference systems, the different methods will be compared, and problems that can occur are pointed out.

Chapter 5

Comparison of coarse-graining methods

Four different methods to derive coarse-grained potentials were described in sec. 2.2: Boltzmann inversion, iterative Boltzmann inversion (IBI), inverse Monte Carlo (IMC) and the force-matching (FM) method. To illustrate the functionality of the VOTCA package and to point out strengths and weaknesses of each method, coarse-graining of four reference systems is performed: SPC/E water, liquid methanol, liquid propane and a single hexane molecule. The systems are chosen in such a way that the corresponding coarse-grained potentials have already been obtained using one or more techniques, providing a good reference point for comparison. Finally, a coarse-grained model of polypyrrole, a conjugated polymer, is developed in sec. 5.6, which can be used to generate large scale morphologies for charge transport simulations.

5.1 SPC/E water

Water is one of the most studied liquids, both from the point of view of all-atom representations as well as coarse-grained models [107, 108]. A detailed comparative study on coarse-graining of different water models using IBI is given in ref. [64]. Here, the focus is on one of the all-atom models of water, the SPC/E water model [109, 110]. The corresponding parameters of this 3-site model are given in the caption of figure 5.1. Note that this is a rigid model, i. e. the distances between two hydrogens as well as oxygen and hydrogens are constrained during the molecular dynamics runs. For the coarse-grained representation a one-site representation with a pair potential $U(R_{ij})$, where R_{ij} connects the centers of mass of water molecules i and j, is applied.
The reference simulation was performed in an all-atom molecular dynamics

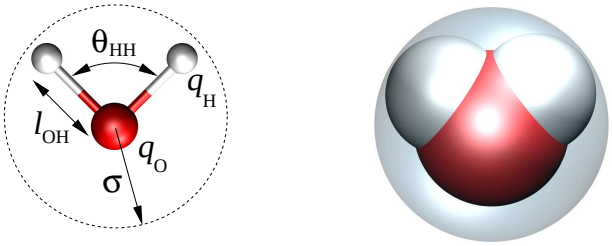


Figure 5.1: **SPC/E water model.** Van der Waals excluded volume and coarse-grained representations of a single SPC/E water molecule. Parameters: $\sigma = 3.166\,\text{Å}$, $\epsilon = 0.650\,\text{kJ mol}^{-1}$, $l_{\text{OH}} = 1.0000\,\text{Å}$, $q_{\text{H}} = +0.4238\,e$, $q_{\text{O}} = -0.8476\,e$, $\theta_{\text{HH}} = 109.47°$. The coarse-grained model has one bead per water molecule. The center of mass of the molecule is used as position of the bead.

(MD) simulations consisting of 2180 water molecules. The system was first equilibrated in the NPT ensemble at 300 K and 1 bar for 100 ns using the Berendsen thermostat and barostat [111]. The last 80 ns were used to determine the equilibrium box size of 4.031 nm, which was then fixed during the 45 ns production run in the NVT ensemble using a stochastic dynamics algorithm [112]. For all further analysis, only the last 40 ns were used. The radial distribution function was calculated using a 0.01 nm grid spacing. Snapshots were output every 0.4 ps.

Force matching potentials were calculated using blocks of 6 snapshots each. A spline grid spacing of 0.02 nm was used in the interval from 0.24 to 1 nm. For the iterative procedures, the potential of mean force was taken as an initial guess for the interaction potential. The coarse-grained box had the same system size as in the atomistic simulations. Simulations of the coarse-grained liquid were done using a stochastic dynamics algorithm [112]. 300 iterations of 100 ps each were performed when using IBI. For IMC 10 iterations of 500 ps each were simulated. Additionally, two iterations of triangular smoothing were applied to the IMC potential update, ΔU. The cut-off was chosen at 0.9 nm with a grid spacing of 0.01 nm.

The reference radial distribution function, $g^{\text{ref}}(r)$, coarse-grained potentials and corresponding radial distribution functions are shown in figure 5.2. IBI and IMC give practically the same interaction potential. Although the force-matched potential has a very similar structure with two minima, the corresponding radial distribution function is very different from the target one. Possible reasons for these discrepancies are discussed in refs. [73, 113, 68] and stem from the fact that FM aims at reproducing the many-body potential of mean force, which does not necessarily guarantee perfect pairwise distribu-

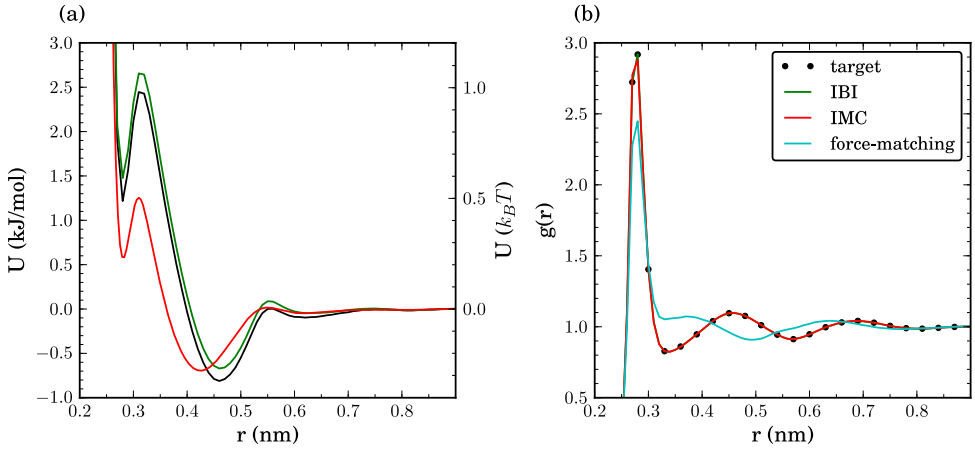


Figure 5.2: **Coarse-graining of water.** (a) Coarse-grained potentials for SPC/E water obtained using different coarse-graining techniques. (b) Corresponding radial distribution functions. Curves for IBI and IMC are on top of each other.

tion functions, considering the fact that the basis set in the coarse-grained force-field is limited. This indicates, that using just a two-body interaction for water might not be the best choice and can lead to complications [114]. The problem of a limited basis set when using force-matching is addressed in sec. 5.5.

Note that all three methods lead to a different pressure of the coarse-grained system: 8000 bar (IBI), 9300 bar (IMC), and 6500 bar (FM). The different pressures for the iterative methods are due to different accuracy of the potential update. Indeed, changes of pressure can significantly affect the potential, especially its long tail [56, 115]. However, they hardly change the radial distribution function due to small compressibility of water. One can improve the agreement between the iterative methods by using pressure correction terms for the update [56, 64].

5.2 Performance of the iterative methods

The performance of the iterative methods depends on two factors: (i) the average (over all bins) error of the potential update $\epsilon_{\Delta U}$ with respect to simulation time and (ii) the number of iterations required for convergence.

For the following analysis, the average error of the update is defined as

$$\epsilon_{\Delta U} = \frac{1}{N} \sum_{i=0}^{N} \epsilon(\Delta U(r_i)) \ , \tag{5.1}$$

where N is the number of bins and $\epsilon(\Delta U(r_i))$ is the error of the update function at a separation r_i. $\epsilon(\Delta U(r_i))$ was calculated using a Jackknife analysis [116].
The average error of the potential update as a function of the run length is shown in figure 5.3(a). One can see that, for both methods, the error decreases as $1/\sqrt{L}$, where L is the number of snapshots used for averaging. However, the prefactor for the IBI update error, which is based on the radial distribution function, is at least ten times smaller than that for the IMC update error, which makes use of cross-correlations of S_α. This observation implies that, in order to have the same accuracy of the update function, IMC needs significantly longer sampling.

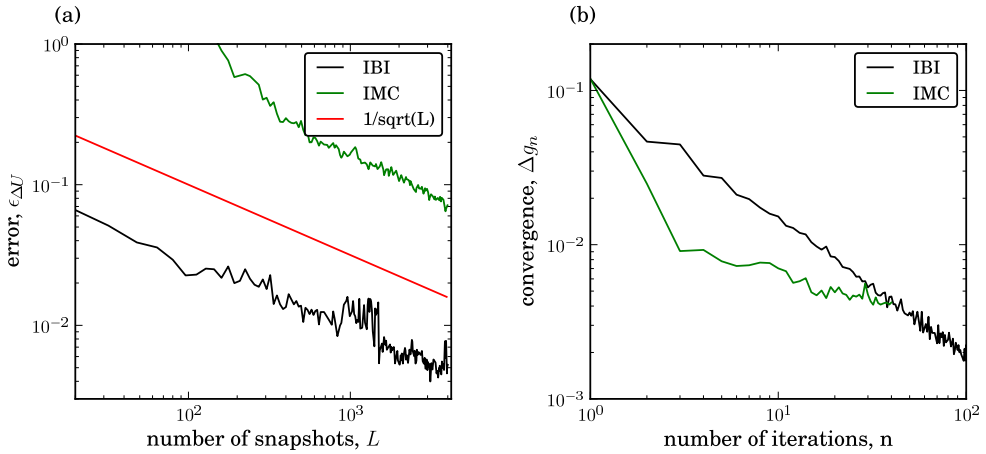


Figure 5.3: **Performance of the iterative methods for water** (a) Average error of the potential update function versus number of snapshots used for calculating the update function. (b) Root mean square deviation of the reference and current radial distribution function versus iteration step. One can see that IMC converges faster than IBI.

This disadvantage can be compensated by the efficiency of the update function, which is assessed by computing the root mean square deviation, Δg_n, of the current and target radial distribution functions

$$\Delta g_n^2 = \int \left[g^{\mathrm{ref}}(r) - g^{(n)}(r) \right]^2 \mathrm{d}r \ . \tag{5.2}$$

Δg_n is plotted as a function of the number of iterations, n, in figure 5.3(b). It is clear that IMC converges much faster than IBI, though the root mean square deviation saturates after some number of iterations. The saturation of IMC might be due to the smoothing of the potential update or due to finite-size effects as pointed out in the next section.
For IBI, each iteration was long enough (800 frames) to provide a smooth potential update that could be applied without any further refinement (except for extrapolation). Contrary to this, although each iteration for IMC was longer (4000 frames), statistics was still not good enough to provide a smooth update function. Smoothing always had to be applied to the potential update. The smoothing was applied to the update and not the potential itself, since the potential has stronger variations which might disappear after too aggressive smoothing. The potential updates for IBI, IMC before and after smoothing are shown in figure 5.4.

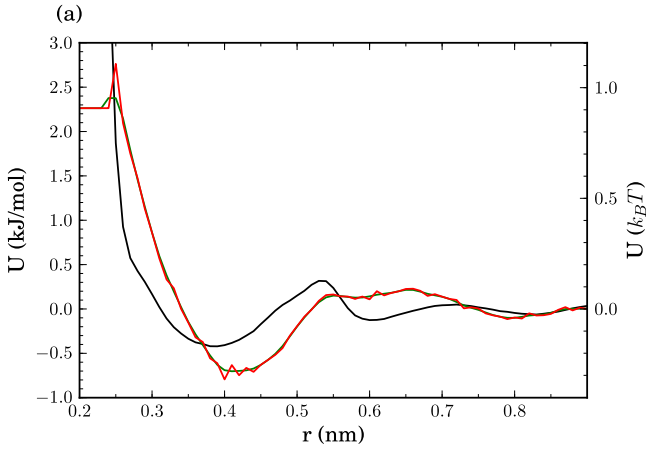


Figure 5.4: **Potential update for water.** The figure depicts the potential update for the first iteration step using IBI and IMC. Since IMC needs significant longer runs to calculate the update, smoothing was always applied to the potential update.

With the given parameters, the total computation time needed for the 20 IMC iterations is comparable to 300 IBI iterations. However, if smoothing is applied to the IBI update as well, the length of the sampling runs can be reduced significantly and the real computation time is below one hour (on a single core of a Core2Duo desktop machine).
It is obvious that, for monocomponent liquids, IBI performs significantly better than IMC. However, it is difficult to compare these methods for general case, for example in multicomponent mixtures, where cross correlations are important. Nevertheless, our experience shows that IMC is very sensitive to

choice of the correct set of parameters and therefore needs significant tuning. In the case of propane (see sec. 5.4), which was taken as an example for a "multicomponent liquid" (different bead types at chain ends), no working parameterset to run IMC without manual interaction was found. In contrast, IBI did not show such problems.

5.3 Finite size effects in methanol

Liquid methanol is the second example of coarse-graining of *non-bonded* interactions that is discussed here. In fact, FM has already been used to coarse-grain this system [115] and, contrary to water, the liquid structure (radial distribution function) is well reproduced by the FM coarse-grained potential. In addition, the excluded volume of methanol is larger than that of water and the undulations of the radial distribution function extend up to 1.5 nm. As will be shown, this leads to pronounced finite size effects for IMC, since it has a non-local potential update. FM and IBI do not have this problem, since the IBI potential energy update is local and FM is based on pair forces. The range of the latter is much shorter than the correlation length of structural properties (such as undulations of the radial distribution function), which may propagate over the boundaries for small boxes.
Simulation parameters were taken from ref. [115] where the OPLS [38, 117] all-atom force-field was used. Atomistic simulations were performed with 1000 methanol molecules in a cubic box (4.09 nm box size) at 300 K using the Nose-Hoover thermostat [118, 119]. The system was equilibrated for 2 ns followed by a production run of 18 ns. The reference radial distribution function was calculated using snapshots every 0.5 ps and is shown in figure 5.5(b).
The FM potential was calculated using blocks of 6 frames each and a spline grid of 0.02 nm. With this potential, coarse-grained simulations were performed using a stochastic dynamics integrator, 1000 beads and with the same box size and the same temperature as in the atomistic simulations. The system was equilibrated for 40 ps followed by a production run of 160 ps. Snapshots were stored every 5 ps and used to calculate the radial distribution function.
For the iterative procedures, the potential of mean force was taken as an initial guess. The cutoff was chosen at 1.54 nm with a grid spacing of 0.01 nm. For IBI, 300 iterations were performed using stochastic dynamics with the same parameters as in the FM-based procedure. The IMC iterations were performed with 8000 molecules and a box size of 8.18 nm. The total length of the run was 1 ns and snapshots were stored every 0.2 ps. Two smoothing

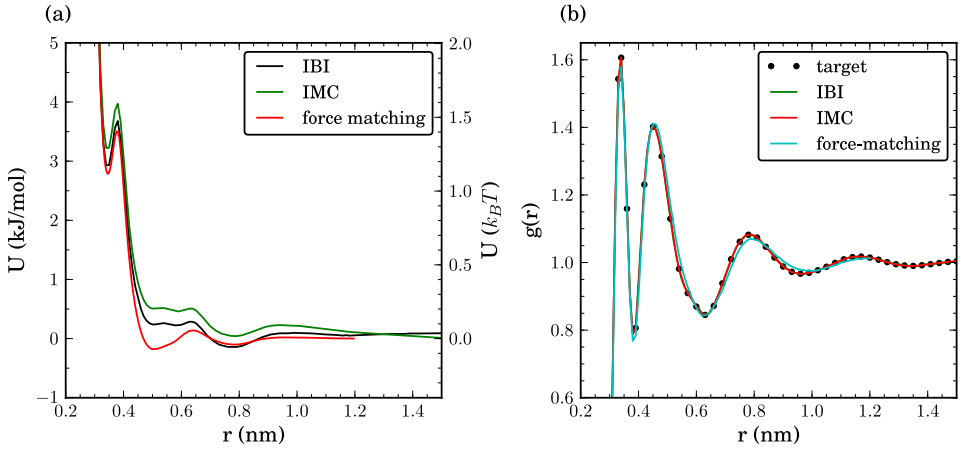


Figure 5.5: **Coarse-graining of methanol** (a) Coarse-grained potentials. (b) Corresponding radial distribution functions. Curves for IBI and IMC are on top of each other.

steps were used at each iteration for the potential update, ΔU.

The coarse-grained potentials for all methods are shown in figure 5.5(a). In spite of small differences between the coarse-grained potentials, the agreement between the reference and coarse-grained radial distribution functions is excellent, as can be seen from figure 5.5(a).

It is important to mention that the IMC method, which has a non-local update, is prone to systematic errors due to finite size effects and hence requires much larger simulation boxes in order to calculate the potential update. This is due to artificial cross-correlations of S_α at large distances, which lead to a linearly shifted potential, and, as a consequence, to a much higher pressure of the coarse-grained system. In contrast, IBI and FM work well with system sizes of the order of two radial distribution function cutoff lengths.

To illustrate this point, simulation boxes of three different sizes, with 1000, 2000 and 8000 methanol molecules (box size of 4.09 nm, 5.15308 nm and 8.18 nm, simulation times of 3 ns, 2 ns and 1 ns respectively), were prepared. The IMC iterative procedure was repeated until the potentials converged, and these are shown in figure 5.6(a). One can see that the potentials significantly differ from each other. These differences lead to small deviations in the tail of the radial distribution function. However, the deviations vanish in a systematic way for bigger boxes, as illustrated in figure 5.6(b), where the integrated squared difference of the reference and current distribution functions is plotted.

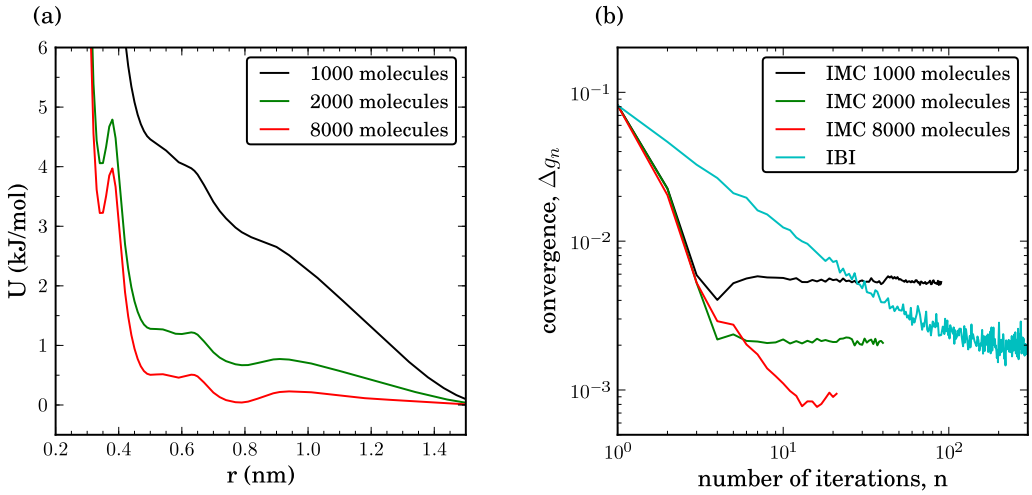


Figure 5.6: **Box size dependence in methanol** (a) coarse-grained potentials using 10 IMC iterations for simulation boxes with 1000, 2000 and 8000 methanol molecules (box size 4.09 nm, 5.15308 nm, and 8.18 nm) equilibrated at the same density. (b) Root mean square deviation of the reference and the current radial distribution function versus number of iterations. Similar to liquid water, IMC converges faster than IBI. The convergence saturates and the saturation error strongly depends on the system size.

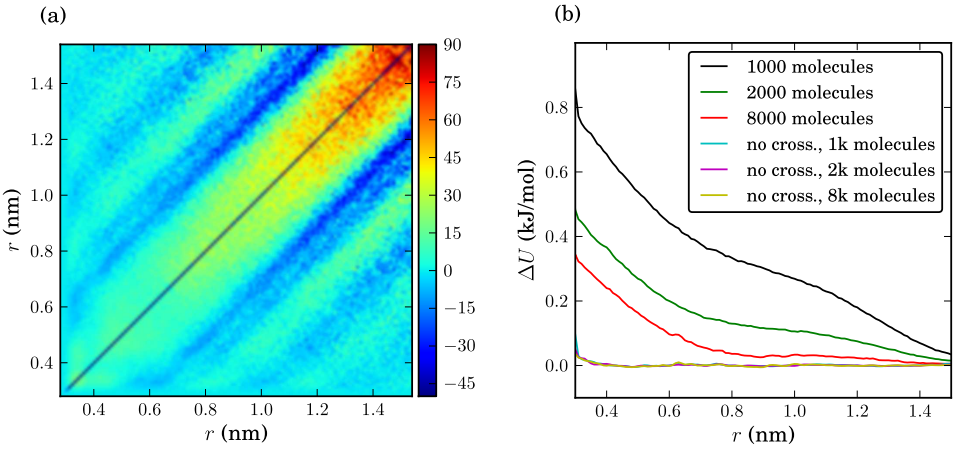


Figure 5.7: **Dependence of the potential update on system size when using IMC.** (a) Update matrix of a system consisting of 1000 molecules (b) Potential update for different number of molecules with and without cross-correlations taken into account.

More detailed analysis has shown that, for small boxes, an additional linear term in the potential update appears (see figure 5.7). To determine the origin of this term, ΔU was calculated using the full matrix $A_{\alpha\beta}$ as well as only its diagonal elements. The potential after 50 IBI iterations was taken as an initial guess. Without the off-diagonal elements, ΔU was small once the reference and coarse-grained radial distribution functions were matching each other. Inclusion of the off-diagonals elements always resulted in an additional, practically linear, term in the potential update which became smaller for large boxes. This observation leads to the conclusion that the off-diagonal elements of the matrix $A_{\alpha\beta}$ systematically change with the box size. However, the physical reasons are not yet fully understood.
To summarize, IMC should be used with care for small systems. The potential update (or the coarse-grained potential) must be converged with respect to the simulation box size. In the case of methanol coarse-graining, a box of size three times the radial distribution function cut-off was not enough to achieve the converged potential for IMC, even though this is sufficient for the IBI and FM methods.

5.4 Liquid propane: from an all- to an united-atom description

So far, coarse-graining of *non-bonded* degrees of freedom using liquid water and methanol were demonstrated. In this section, it is shown how *bonded* interactions can be coarse-grained by deriving a united atom model (i.e. hydrogens embedded into heavier atoms) from an all-atom model of liquid propane[1]. The mapping scheme, as well as the bonded coarse-grained variables (two bonds, b, and one angle, θ) are shown in figure 5.10(a). This coarse-graining scheme has two different bead types: an inner bead, of type B, with two hydrogens, and two outer beads, of type A, with three hydrogens. As a result, three types of non-bonded interactions, U_{AA}, U_{BB}, and U_{AB} must be determined.
As before, atomistic simulations were performed using the OPLS all atom force-field [38, 117]. A box of liquid propane was first equilibrated at 200 K

[1]The united atom model used here shall not be confused with the united atom models commonly used in the atomistic force-field community, for example OPLS-UA force-field [38, 117]. The latter models map the potentials, which are analytical functions of bonds, angles, dihedral angles and non-bonded interactions, onto thermodynamic properties of the corresponding substances. In our case coarse-grained potentials are tabulated functions of coarse-grained variables and only the mapping (hydrogens embedded into heavier atoms) is similar to that of the united atom force-fields.

and 1 bar in the NPT ensemble for 10 ns using the Berendsen thermostat and barostat [111]. The equilibrated box of the size $4.96337 \times 5.13917 \times 4.52386\,\mathrm{nm}^3$ was then simulated for 10 ns in the NVT ensemble at 200 K using velocity rescaling with a stochastic term [120]. No bond constraints were used during the simulations and hence the integration timestep was 1 fs. Snapshots were written every 1 ps.

In the case of iterative methods, the bonded potentials (bond and angle) were calculated by Boltzmann-inverting the corresponding distribution functions of a single molecule in vacuum, according to eq. 2.23. The propane molecule in vacuum was simulated in an stochastic dynamics run [112] for 100 ns with snapshots stored every 2 ps. Non-bonded potentials were iteratively refined by using IBI with a grid spacing of 0.01 nm and a cutoff of 1.36 nm (1.38 nm) for A-A, A-B (B-B) interaction types, respectively. The run length for each iteration was 50 ps with snapshots written every 0.5 ps. At every iteration step only one interaction type was corrected. Coarse-graining using the IMC method was not in an automatized way because of problems with finite size effects (see methanol) as well as jumps in the potential update, which occur at short particle separations due to bad sampling of these areas[2]. When using the FM method, both bonded and non-bonded potentials were obtained at the same time, since FM does not require the explicit separation of bonded and non-bonded interactions.

The obtained potentials are shown in figure 5.8 and figure 5.9. FM and Boltzmann inversion-derived bond and angle potentials (figure 5.9) perfectly agree with each other. The non-bonded potentials, shown in figure 5.8(a), are not identical, but have similar shapes and barrier heights. This results in a good reproducibility of the propane liquid structure by the FM-based coarse-grained potentials, as can bee seen from the radial distribution functions shown in figure figure 5.8(b). Again, as expected, IBI reproduces the reference radial distribution functions exactly.

To summarize, the united atom model of liquid propane is an ideal example of coarse-graining where the structure- and force-based methods result in similar bonded and non-bonded interaction potentials. As will be seen later, this is due to (i) the completeness of the basis set used to construct the coarse-grained force-field; and (ii) independence of bond and angular degrees of freedom. The latter can be understood with the help of a histogram showing the correlation of b and θ, depicted in figure 5.10. The symmetric shape

[2]The non-bonded potential updates rapidly change signs at short particle separations. This originates from bad sampling of these areas as well as from inverting the full IMC matrix including all cross-correlations of different interactions. One could correct these errors by hand. However, an automatic procedure, which is the focus of this work, is cumbersome.

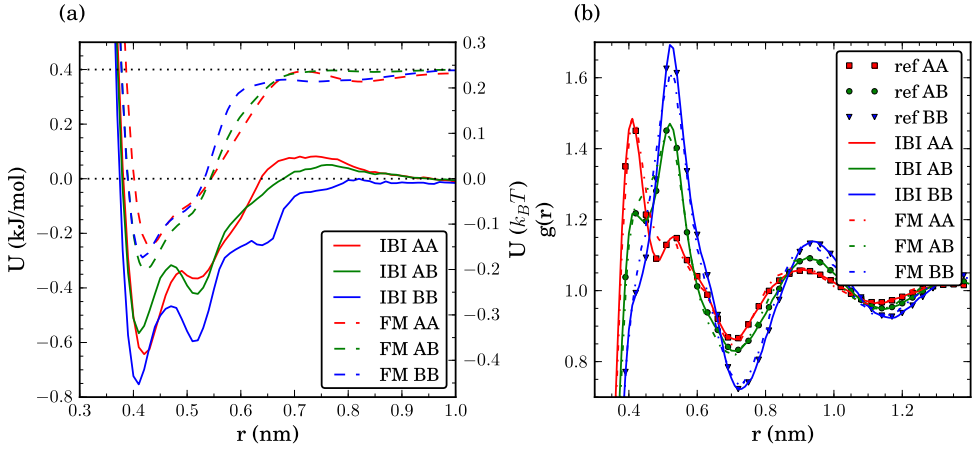


Figure 5.8: **Non-bonded interactions for propane.** (a) Non-bonded interaction potentials U_{AA}, U_{BB}, and U_{AB} obtained with IBI and FM methods. For clarity, FM potentials are offset along the y axis. (b) Corresponding radial distribution functions, plotted together with the atomistic radial distribution function.

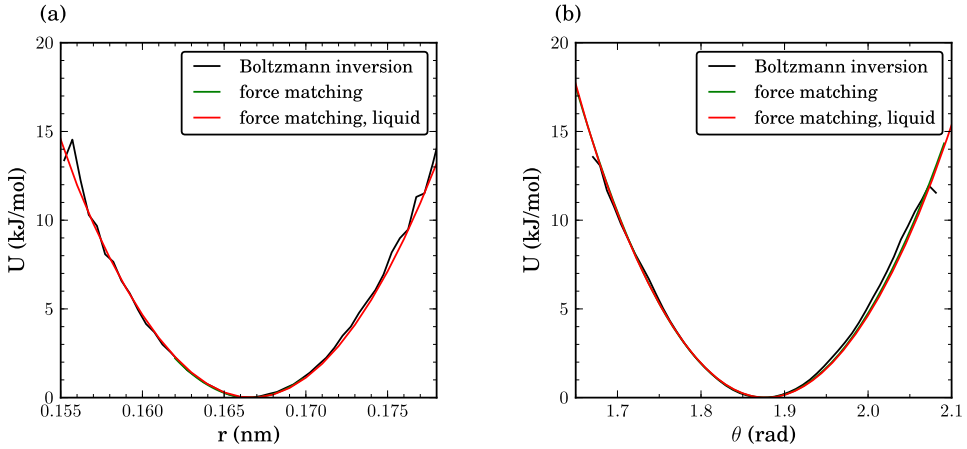


Figure 5.9: **Bonded interactions for propane.** (a) Bond potential obtained for a single molecule in vacuum by Boltzmann-inverting the corresponding distribution function, using FM for a single propane molecule in vacuum and force matching for liquid propane. (b) Angular coarse-grained potentials. FM-based distributions for a single molecule and the liquid are on top of each other.

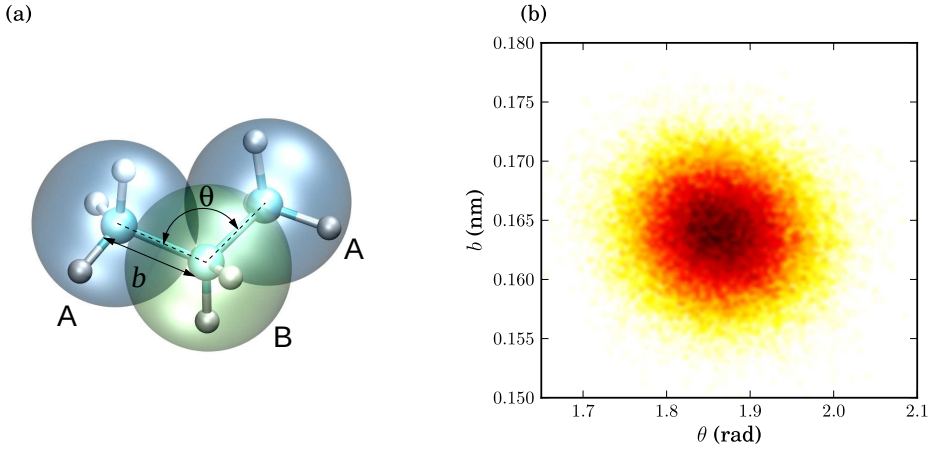


Figure 5.10: **Propane, model and correlations.** (a) The all-atom and coarse-grained representations of a propane molecule, bead types, and coarse-grained bonded degrees of freedom (bond b and angle θ) (b) The correlations of b and θ.

of the distribution plotted clearly indicates, that b and θ are uncorrelated. In the next section, coarse-graining of a single molecule of hexane will be discussed, for which this is not the case.

5.5 Angular potential of a hexane molecule

The last example discussed here is the angular potential of a hexane molecule coarse-grained into a three-bead chain, with two carbon atoms per bead (see figure 5.11(a)). Atomistic simulations of a single hexane molecule in vacuum were performed using the all-atom OPLS force-field and a stochastic dynamics integrator [112]. The run length was 1000 ns and snapshots were stored every 2 ps.

The coarse-grained angular potential was again obtained by Boltzmann-inverting the angular distribution function or by using the FM method (blocks of $5 \cdot 10^4$ frames each and a spline grid of 0.05 nm, sampling in the $\theta \in [1.6, 3.14]$ interval was used). Both coarse-grained potentials are shown in figure 5.12. The corresponding distribution functions, together with the reference function obtained from the atomistic simulations, are shown in figure 5.11(b).

The distribution which corresponds to simple Boltzmann inversion is practically identical to the reference distribution, while the FM-based distribution

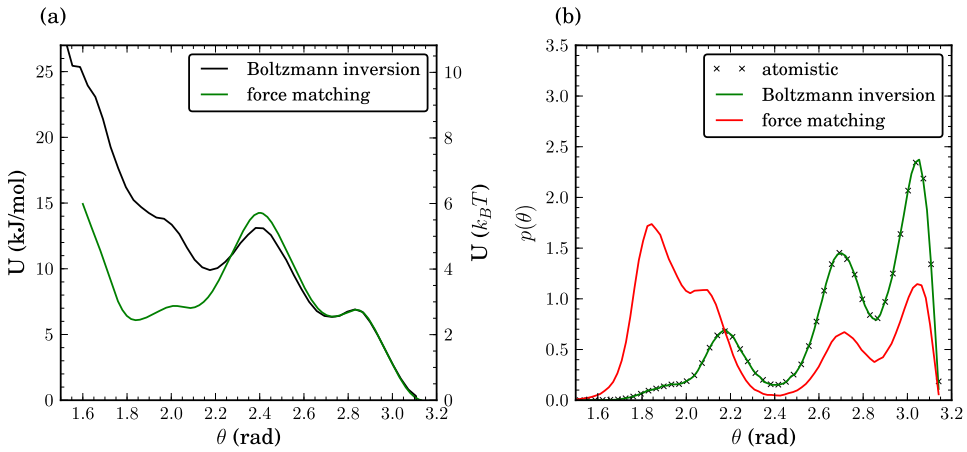


Figure 5.11: **Angular potential and distribution for hexane** (a) Coarse-grained angular potentials obtained using Boltzmann inversion (no iterations) and FM for a single hexane molecule in vacuum. (b) Probability density (probability distribution normalized by the interval) obtained from the atomistic run as well as from the runs using coarse-grained angular potentials.

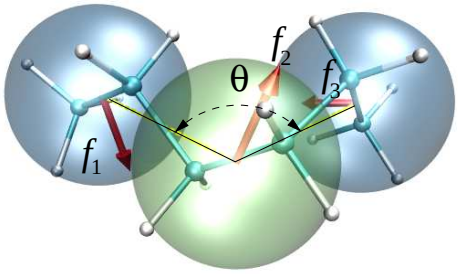


Figure 5.12: **Forces in hexane.** Hexane molecule and its coarse-grained representation. Arrows indicate the directions of the forces on three beads for a specific snapshot. Force-matching fails since forces on beads have an *out of plane* component which cannot be reproduced by the coarse-grained model.

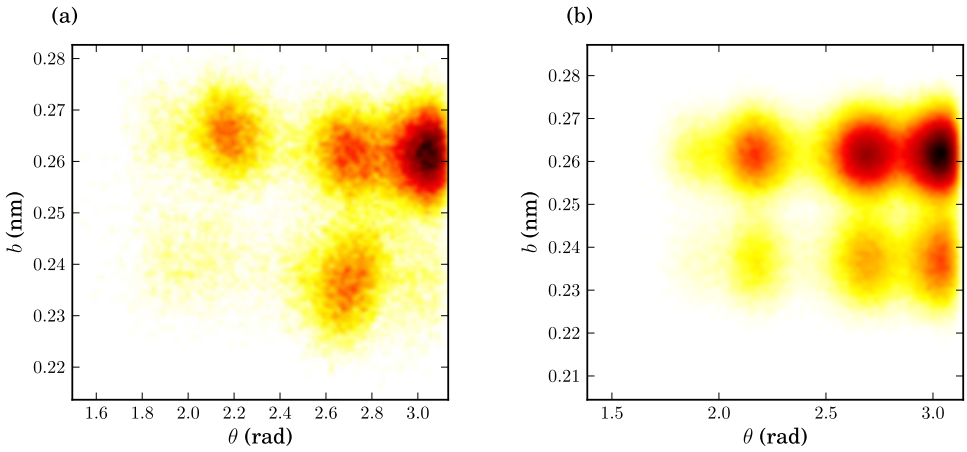


Figure 5.13: **Correlations in hexane.** (a) Bond length b and angle θ are correlated in the mapped atomistic trajectory. (b) Since the coarse-grained model cannot reproduce these correlations, sampling of the configurational space is different. This will lead to complications when back-mapping to atomistic details.

samples small angles more often. This is a direct consequence of a very deep local minimum in the angular potential at these angles. It is easy to understand why FM fails to predict the relative height of this minimum. On a coarse-grained level the change of the angle from large to small values corresponds to the reorientation of the dihedral angles at the atomistic level. This reorientation results in instantaneous forces, $\boldsymbol{f}_1$, $\boldsymbol{f}_2$, $\boldsymbol{f}_3$, on the beads which have an *out of plane* component, where the plane is defined by the centers of the beads (see figure 5.12). The coarse-grained potential, however, has only an angular term, U_θ, and hence can only capture forces which lie *in the plane* in which the angle θ is defined. Hence, only the *projections* of the forces on this plane are used in FM, and this clearly leads to underestimation of the position of the second minimum, since the work conducted by the out-of-plane forces is completely ignored.
Additionally, this mapping scheme does not have independent variables, e. g. bond and angle degrees of freedom are coupled, as can be seen from the two dimensional histogram shown in figure 5.13. This means that, even though Boltzmann inversion reproduces correct distributions, sampling of the configurational space is incorrect because of the lack of cross-correlation terms in the coarse-grained potential. For this simple model, the latter issue can only be solved by an additional coupling term for the bond-angle potentials.

5.6 A coarse-grained model for polypyrrole

All examples discussed so far are small molecules. This section demonstrates how coarse-graining of a polymer can be performed using a combination of Boltzmann inversion and its iterative counterpart. Bonded interactions were derived for a single chain in vacuum using Boltzmann inversion. The trajectory was prepared by excluding non-bonded interactions of atoms, which imposes decoupling of the bonded and non-bonded interactions in the coarse-grained representation. The non-bonded potential was calculated independently matching the radial distribution function for a melt of monomers using IBI. The development of the atomistic model is described in detail in sec. 6.1. A 1:1 mapping scheme, in which one chemical repeat unit is mapped onto one bead was used, and the center of mass of a repeat unit is taken as position of the coarse-grained bead. Beads are connected by bond, angle and dihedral potentials.

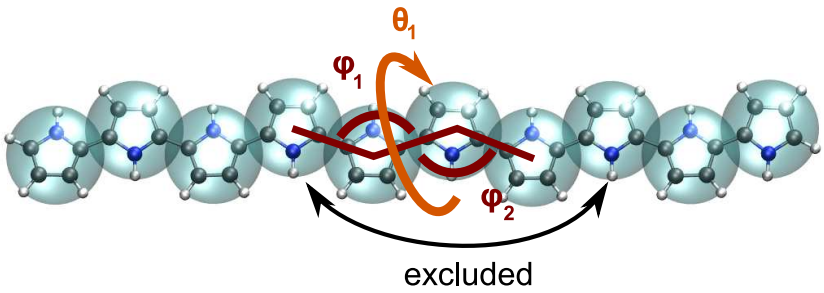


Figure 5.14: **Coarse grained model for polypyrrole.** Each repeat unit was mapped to one bead. Beads are connected by bond, angle and dihedral potentials. To separate bonded and non-bonded interactions, all interactions in the atomistic model which would contribute to the non-bonded interactions in the coarse-grained model have to be excluded in the sampling run for deriving the bonded interactions.

Bonded potentials of the coarse-grained model were derived using the Boltzmann inversion method [42]. Phase space was sampled by a stochastic dynamics simulation of an isolated atomistic chain in vacuum for 40 ns taking snapshots every 2 ps at 200 K, 300 K, and 400 K. To separate bonded and non-bonded interactions, all interactions in the atomistic model which would contribute to the non-bonded interaction in the coarse-grained model have to be excluded. This was achieved by excluding all non-bonded interactions in the atomistic model that correspond to 1-5 or further interactions in the coarse-grained scheme (see figure 5.14). The rest of the non-bonded interactions were accounted for since contribute to the bonded potentials in the

coarse-grained model. The distributions and Boltzmann inverted potentials of the dihedral angle are plotted for different temperatures in figure 5.15.

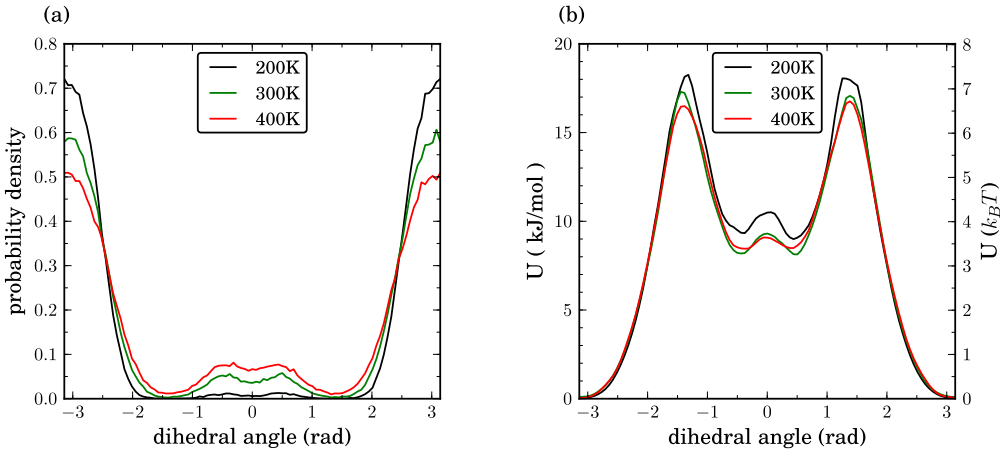


Figure 5.15: **Coarse-grained dihedral in polypyrrole.** The probability distribution (a) and potential (b) for the dihedral potential that connects the beads in the coarse-grained model is plotted at 200K (black), 300K (blue) and 400K (red). The potential was calculated by Boltzmann inverting the probability distribution followed by a smoothing step to filter noise. With increasing temperature, the probability distribution broadens but the corresponding potentials do not change much. This fact is not compulsory since entropic contributions can, in principle, change the potential and are temperature dependent.

The distribution functions show that, in the coarse-grained model, bonds are stiff. To a good approximation one can use constrained bonds of length 0.377 nm. The angle potential was fitted with a harmonic potential ($\varphi_0 = 140\,\text{deg}$, $k = 600\,\text{kJ/mol/rad}^2$) while a tabulated potential is used for the coarse-grained dihedral potential. The bond and angle potentials only very weakly depend on temperature, therefore, all values were derived at 300 K. To calculate the non-bonded potentials, the IBI method [56, 58] was used to match the radial distribution function (rdf) of liquid pyrrole. Radial distribution functions were calculated based on trajectories of atomistic MD simulations. A box of 512 pyrrole monomers was first equilibrated in a NPT run (Berendsen thermo- and barostat) at 1 bar and 200 K, 300 K and 400 K for 1 ns, followed by a production run in the NVT ensemble for 2 ns saving snapshots every 0.2 ps. Monomers were mapped to coarse-grained beads and the rdfs were calculated. The rdfs, which are shown in figure 5.16, reflect

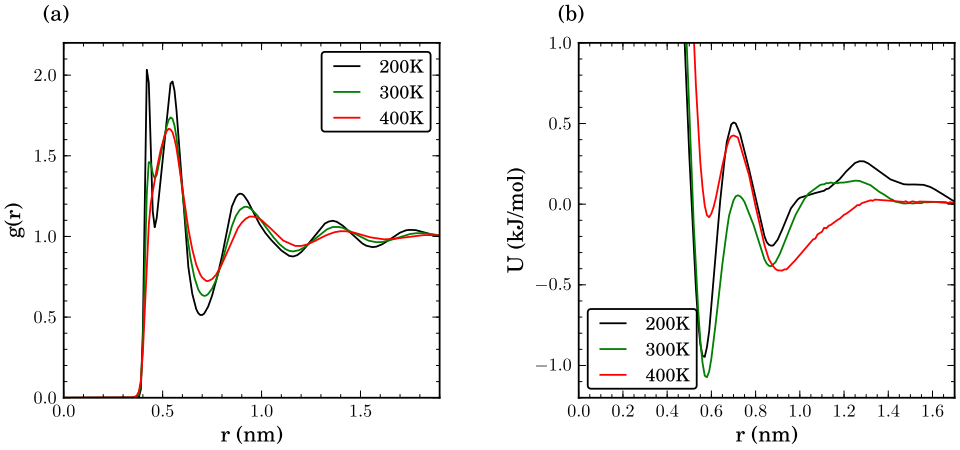


Figure 5.16: **Non-bonded potentials for coarse-grained pyrrole.** (a) Radial distribution functions of an atomistic melt of monomers mapped onto coarse-grained beads for 200 K, 300 K and 400 K. Coarse-grained RDFs match perfectly and are not shown. (b) Coarse-grained potentials obtained using the IBI method.

the ellipsoid-like shape of the pyrrole monomers: at 400 K, the first peak is slightly deformed. Due to different relative orientations of neighboring monomers, this peak splits up into two smaller peaks with decreasing temperature, which corresponds to the face to face and T-shaped arrangement of neighboring molecules.
300 IBI iterations, 1 ns each, with snapshots saved every 1 ps, were performed to obtain the non-bonded potentials. Box size and initial coordinates were taken from the mapped atomistic configuration. For the systems at 200 K and 300 K, the potential update was scaled by a factor 0.5 to stabilize the scheme.
In principle, one has to study a set of oligomers to account for connectivity effects (repeat units which are embedded in a chain are sampled differently than free monomers). However, the rdf of an equilibrated melt would be needed as a reference which cannot be equilibrated at ambient conditions, since the polymer is in a glassy state.

5.7 Conclusions

To summarize, four methods, namely Boltzmann inversion, iterative Boltzmann inversion, inverse Monte Carlo, and force matching, have been used to coarse-grain a set of reference systems: liquid water, methanol, liquid propane, and a single molecule of hexane.
All implemented methods have advantages as well as shortcomings. Inverse Monte Carlo has an update function which is more efficient than that of the iterative Boltzmann inversion method. However, inverse Monte Carlo is very sensitive to the system size and statistical averaging. In general, more attention has to be paid for the preparation of the iterative runs when using inverse Monte Carlo, and IBI turns out to be a more robust method. Force matching, which is a not an iterative method, has problems is the basis set used to represent the coarse-grained potential energy surface is incomplete. An incomplete basis set leads to potentials that might show a completely non-physical behavior. It should always be kept in mind that the coarse-grained systems are physically different to the reference systems, and that the coarse-graining methods cannot be used as a black box and require thorough cross-checking.
Due to the intrinsic approximations of coarse-graining, the generated morphologies will describe correctly only the global morphology but not the local packing. However, atomistic details can be reintroduced and only a short run at atomistic level is then required for local equilibration and compensation of inaccuracies of the coarse-grained model [75, 121].

Chapter 6

Charge transport in Polypyrrole

In this chapter, a computational study of the relationship between the arrangement of polymer chains and charge mobility is presented for neutral and oxidized oligomers of polypyrrole (PPy) (see figure 6.1). Historically, polypyrrole is probably the very first polymer reported to have good conducting properties when oxidized [122]. It is easy to synthesize and handle; conductivities up to $300\,\mathrm{S}\cdot\mathrm{cm}^{-1}$ were reported in the oxidized state. The structure and therefore the conductivity of polypyrrole heavily depends on the processing. Similar to the other conjugated polymers, the extreme insolubility of polypyrroles in organic solvents hinders detailed analysis. Therefore, the physical properties and structural characteristics of PPy are not well understood. An experimental overview of the electrochemistry of conducting polypyrrole films is given in ref. [123]. Here only the main and rather sparse theoretical contributions to the field are mentioned.

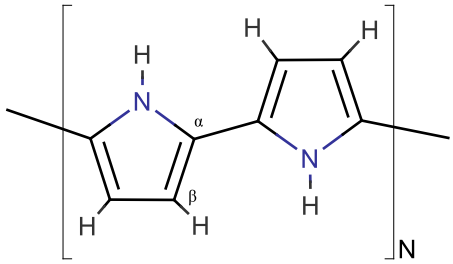


Figure 6.1: **Chemical structure of polypyrrole.** The figure depicts the chemical structure of neutral polypyrrole. In the oxidized case, one electron per 3 repeat units is removed (on average) and the chain is positively charged.

The effect of doping on the geometric and electronic structure has been studied in pioneering works of Brédas et. al. [124, 125]. It was shown that for high

doping levels bipolaron bands are formed, confirming experimental observations of the spinless nature of charge carriers in the highly doped state. Similar analysis was performed using density functional approaches [126, 127]. The hopping transport picture and underlying parameters were analyzed for pyrrole oligomers by Hutchison et. al. [128]. The stability of oligomeric PPy structures bonded through α and β carbons was studied by Yutsever. It was shown that PPy is able to form-branch like structures [129, 130].
At the atomistic level of description, OPLS-AA force field parameters were derived for liquid pyrrole by Jorgensen et. al. [131]. Molecular dynamics simulations of solvated reduced and oxidized polypyrrole were performed by Cascales et. al. [132, 133]. In these simulations, the standard GROMOS torsion potential connecting repeat units was used for the neutral as well as the doped polymer. Apart from the fact that this potential contradicts recent quantum-chemical calculations [134], it differs significantly for doped and oxidized states as shown in sec. 6.1. Hence, the first important task on the way of modeling PPy morphologies is the development of a reliable atomistic force field, with appropriate parameters for the torsion potential and correct partial charge distributions in reduced and oxidized states. This is performed in sec. 6.1.
The aim of this study is to investigate the effect of global arrangement of chains on charge carrier mobility, i.e. whether optimizing the alignment of chains can help to improve conductive properties. The model used is based on the framework described in sec. 3.7 and was previously used to correlate structure and charge mobility in discotic liquid crystals [81, 82, 83, 84, 85, 86, 25, 135]. First, the morphology of an assembly of chains using molecular dynamics (MD) is simulated. In order to tweak the morphology two methods are used: the initial arrangement of polymer backbones for the MD simulation is changed and the simulation is performed on oxidized polymer chains in the presence of counterions. In particular three types of chain alignments are used: all the chains aligned in one direction, the chains lying in one plane, and an isotropic arrangements of chains. The morphology is then used to compute charge transfer rates and hence charge mobility. The main change in the modeling of charge transport compared to the case of discotic liquid crystals is that charge transfer units are defined in terms of subsets of the polymer chain (conjugated segments), the use of different rate equations for intra- and intermolecular transport, and the computation of the nearest neighbor list for the three dimensional network of chains.
In discotic liquid crystals it was argued that the value of the charge mobility is determined by the distribution of the intermolecular transfer integrals, which is directly linked to the azimuthal register and separation between the molecules [81]. The charge transport in these systems has one-dimensional

character and is limited by the tail of small transfer integrals (defects) in the system. This can be easily rationalized because in one dimension the smallest transfer integrals are the bottlenecks for charge transfer. The assembly of PPy chains is of course not one dimensional and therefore it is harder to predict what the bottleneck of charge transport is. It will be shown that, in this case, mobility is almost entirely determined by the magnitude of the transfer integral which allows the whole simulation box to be connected together. Most intriguingly, the size of this transfer integral is not strongly affected by the changes in morphology. The main conclusion is therefore that, for the case of such amorphous polymers, charge mobility is mostly determined by local packing and is almost independent on the alignment of chains.

6.1 Atomistic Model

The first step is to set up a reliable atomistic model of PPy. As a starting point, the OPLS-AA force field parameters are taken, which were parametrized based on pyrrole monomers [131]. The parameters of the dihedral and improper potentials connecting adjacent repeat units are of course not available in the standard force field. For the charge transport studies in this work, the dihedral potential of the torsion connecting two adjacent repeat units is of special interest, since it influences the global conformation of the chain the most. In addition, conjugated segments are determined based on fluctuations of this torsion angle. The missing parameters were obtained by fitting the potential energy surface to first principles calculations. In what follows, the procedure for the dihedral potential is described. Bonds and angles would need tuning if vibrational spectra should be studied, but their exact force constants are less important for morphology. The same holds for improper dihedrals, which are mainly used to keep conjugated rings planar. For neutral polypyrrole, the torsion angle θ of the N-C-C-N dihedral in 2,2'-bipyrrole (see figure 6.2) was scanned and the torsion potential was evaluated using B3LYP hybrid density functional and 6-311G(d,p), 6-311G++(d,p) and 6-311++G(3df,3pd) basis set. An additional MP2 calculation was performed using the 6-31G(d) basis set. All calculations were performed using the GAUSSIAN package [36]. At each scanning step, the dihedral angle θ was fixed at the value of interest while the rest of the structure was optimized. The dependence of the potential on the torsion angle is depicted in figure 6.2. [1]

[1] One should always comment on the accuracy of DFT: Even though DFT does not show perfect agreement with higher correlated quantum chemical methods when applied

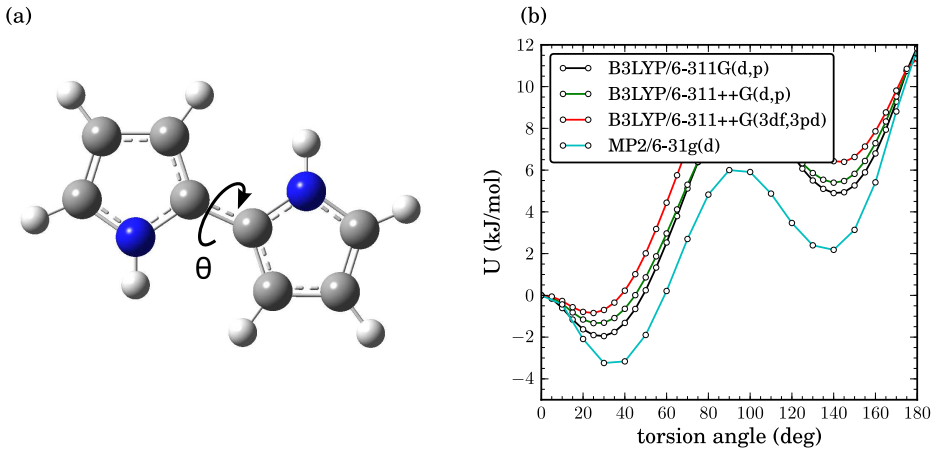


Figure 6.2: **Neutral polypyrrole force-field scans.** The potential energy curve for a 2,2'-bipyrrole as a function of the dihedral angle θ with B3LYP/MP2 and different basis sets.

To obtain the parameters of the dihedral potential from the *ab-initio* potential energy, the optimized geometries were further relaxed in MD, again constraining the dihedral angle and with its parameters set to zero. The force-field based potential energy of these relaxed contributions was then subtracted from the potential energy curve provided by *ab-initio* B3LYP/6-311++G(3df,3pd) calculation.

Note that direct fitting, without subtraction, is a common mistake which leads to a double counting of those energy contributions which were already taken care of in the existing atomistic model, such as bond, angle, improper dihedral, Coulomb and Lennard-Jones interaction potentials. Another, more subtle problem is that the minimum energy conformation in the atomistic model has small deviations compared to the quantum chemically optimized structures. Without relaxation of internal stresses in MD (mainly due to improper dihedrals, angles and bonds which hardly contribute to the global conformation), this would lead to different barriers. The relaxation in MD leads to potential barriers of equal heights for torsions in the chain, both, in quantum chemical and atomistic calculations.

to judging rotational barriers in conjugated materials [136], its accuracy is still sufficient for the studies on polypyrrole in this work [134], given the other errors introduced in atomistic simulations.

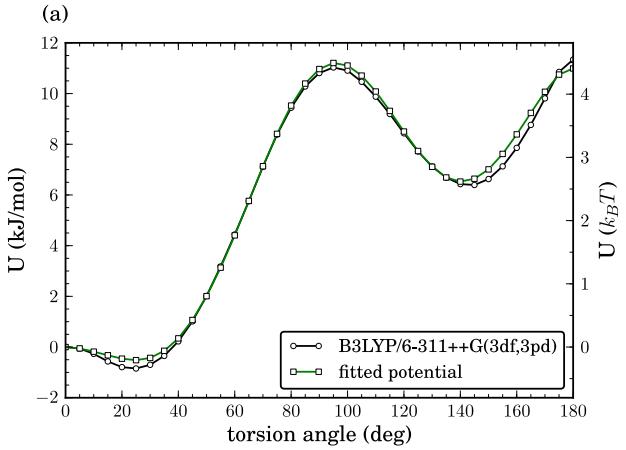


Figure 6.3: **Neutral polypyrrole force-field fitting.** The potential energy curve for a 2,2'-bipyrrole as a function of the dihedral angle θ and the according calculation the fitted MD potential.

A Ryckaerd-Belleman type potential

$$V_{rb} = \sum_{n=0}^{5} C_n \left(\cos\theta\right)^n \ . \tag{6.1}$$

was used to fit the potential energy difference. The resulting potential is depicted in figure 6.3. Atomic charges were calculated using B3LYP/6-311G(d,p) and CHELPG [137] fitting procedure in a chain of 8 repeat units. All force-field parameters are given appendix C.
To study the effect of doping on the force field parameters, the torsion angle of a charged chain was studied. Experimentally, the best conductivities are obtained for a doping rate of one charge per three repeat units [138]. Here, a charged tetramer was used and the central dihedral was constrained to the value of interest while the other degrees of freedom were free to relax (similar to the dimer scan in the neutral case). A charged tetramer is a good compromise, since it reproduces the total charge of experiments, as well as allows for the computational treatment of the scan on a quantum chemical level. In addition, the scan is carried out for the inner dihedral since this configuration is symmetric and is representative for the bonding situation in a polymer.
Figure 6.4(a) shows the potential energy of the neutral and the charged tetramer. It is clear that the barrier is increased and the planar configuration is favored when the polymer is doped. This high barrier cannot be reproduced by the atomistic models for the neutral case, even if the charges are adjusted

to the doped case. Moreover, for a doped polymer, partial charges of the atoms do not increase homogeneously: the whole distribution along the chain changes. Hence, both the dihedral potential and partial charges should be adjusted in order to correctly describe doped chains. All parameters are given in appendix C.

The chemical reason for the increase of planarity in PPy upon doping is the quinoid structure of oxidized PPy, i. e. the double bond character of the bond connecting successive monomers. When an oxidized oligomer is forced to acquire a twisted conformation the molecule will find itself in a diradical state which cannot be properly described by ground state theories such as DFT. Having said this, it is also clear that small deviation from planarity (where diradical formation can be assumed not to play an important role) results in much larger potential energies in the oxidized than in the reduced states. DFT cannot give correct quantitative estimates of the rotational barrier in charged PPy but can certainly describe the general trends: a reduction in bond length of the bond which connects repeat units and a increase in stiffness. To take this into account at the force-field level, changes in bond lengths were ignored and only the dihedral potential and partial charges were adjusted.

When fitting the atomistic potential, the unknown torsion potential occurs at three positions in the molecule. To calculate energies in MD, the DFT optimized structures were taken and all three dihedrals were restrained to their values while the remaining degrees of freedom were relaxed. The functional form for the atomistic potential was determined by a multidimensional fit to the Ryckaerd-Belleman parameters. Since a Ryckaerd-Belleman potential linearly depends on its parameters, they can be obtained by solving a set of linear equations.

In the doped state, the additional charge is delocalized along the chain. A priori it is not clear, whether this leads to a coupling of neighboring dihedral potentials. To test this, additional DFT scans, shown in figure 6.4(b), were performed around one of the outer dihedral of the tetramer, while the middle dihedral was fixed to 0, 10, 20 degrees. These calculations show, that the shape of the potential does not change and only the position of the minimum is affected, which is due to non-bonded interactions.

Furthermore, the potential derived from the tetramer was tested for a hexamer by comparing the potential barrier from MD and DFT, and no difference was found. These points indicate that the same torsional potential can be used for all repeat-unit-linking dihedral angles of the polymer.

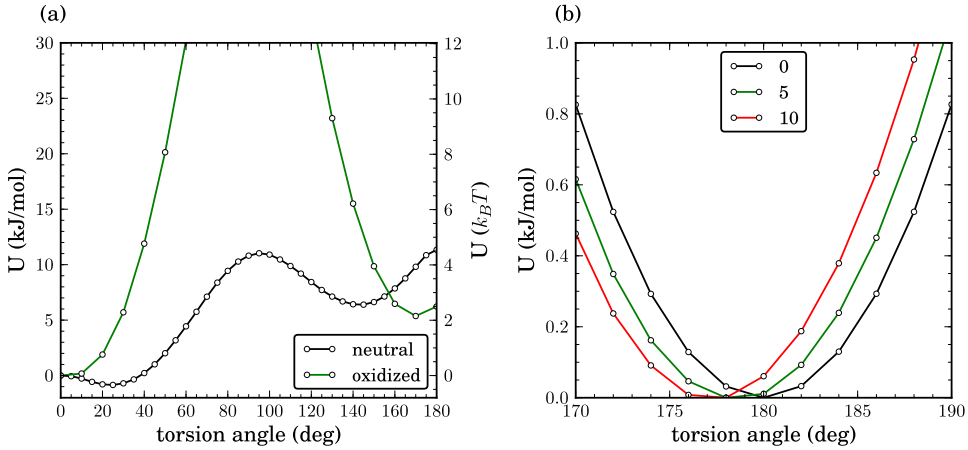


Figure 6.4: **Oxidized polypyrrole force-field.** (a) the torsion barrier of the neutral and the oxidized polymer. While in the neutral case the equilibrium is approximately 30° and the height of the barrier only a few $k_{\mathrm{B}}T$, the oxidized state has a high barrier and flat geometry. (b) The figure shows the energies, when one dihedral angle is fixed at a certain value while another one is rotated. The calculations shows no coupling between the torsion potentials. The curvature does not change, and the small shift can be explained due to non-bonded effects.

6.2 Morphologies

The main aim of this study is to investigate the effect of chain alignment on charge carrier mobility. Therefore, three kinds of molecular morphologies were prepared with different types of molecular ordering, which will be referred to as a *fiber*, *slice* and *isotropic* (figure 6.5).
In case of the fiber, polymer chains (originally set up on an ideal lattice with experimental density for liquid pyrrole) were aligned along the x axis, and shuffled randomly. The slice morphology was generated by randomly rotating chains in the xz plane, as well as along the chain axis and again shuffled randomly[2]. Finally, isotropic configuration was achieved by assigning random orientations to all the chains and randomly translating them in space. Using liquid crystalline terminology, fiber corresponds to a nematic mesophase with the order parameter $S = 1$, slice with $S = -1/2$, and isotropic morphology to the isotropic mesophase with $S = 0$. All systems consisted of oligomers of length 10 repeat units, 512 (256 for the fiber) molecules in total.

[2]Note that Euler angles are not suitable for creating uniform random rotations, a method as described in ref. [139] should be used.

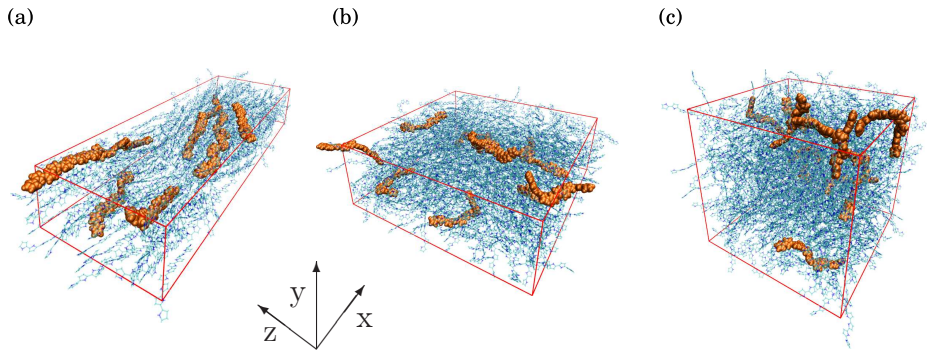


Figure 6.5: **Polypyrrole morphologies.** Representative snapshots of three different molecular arrangements. (a) Fiber, with all molecules aligned along the z axis. (b) Slice, with all molecules aligned in the xy plane. (c) Isotropic, with randomly oriented molecules. To emphasize molecular alignment, several randomly chosen molecules are shown using a different representation.

The morphologies can, in principle, be equilibrated using the coarse-grained model developed in sec. 5.6. However, this would destroy the alignment of chains which was imposed during preparation. Instead, atomistic MD simulations were used to locally equilibrate the samples while conserving the global order. Therefore, these are model systems which do not necessarily correspond to experimental structures.
It is important to start the MD run with the approximate density of the equilibrated sample. The chains are strongly attractive, starting with a low density would immediately lead to clustering of chains and equilibration, even with strongly heating up the sample, is hardly possible. However, starting with the approximate density leads to overlap between chains in the initial configuration. To remove this overlap, short runs were performed in which non-bonded interactions were gradually switched on. To locally equilibrate the systems, the conformations were then heated up to 500 K for 5 ns. Finally the systems were cooled down to 300 K during 1 ns followed by a final equilibration at room temperature for 2 ns. All runs were performed using the GROMACS package [37]. For the further charge transport calculations, 50 snapshots were saved every 2 ps. Representative snapshots of all three morphologies are shown in figure 6.5.
In the oxidized case, three charges per chain were assumed [138], while adding Cl^- as counterions (see figure 6.6(a)). An equilibrated conformation from the neutral case was taken as a starting point. Charges and the dihedral parameter of the force field were changed to the oxidized ones and counterions

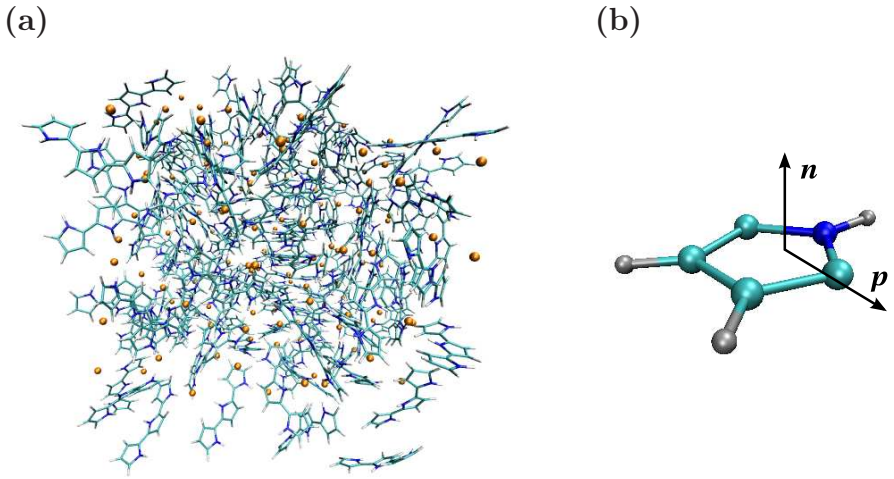


Figure 6.6: **Oxidized polypyrrole morphology.** (a) Isotropic morphology of oxidized polypyrrole using Cloride as counter ions. (b) The normal vector $\boldsymbol{n}$ and vector $\boldsymbol{p}$ connecting the two carbon atoms involved in bonding are used to characterize polypyrrole morphologies.

added at random positions. The non-bonded interactions of the counter ions were slowly switched on during a stochastic dynamics run.
To characterize the generated morphologies, the radial distribution function of the centers of mass of the repeat units was calculated. In order to confirm that the three starting morphologies and the two different force-fields corresponding to a flexible reduced polymer or a stiff oxidized one have indeed resulted in assemblies with different degree of orientational ordering of chains, three nematic order parameters were computed for each morphology. The nematic order parameters are based on the vector $\boldsymbol{n}$, normal to the pyrrole repeat unit plane, vector $\boldsymbol{p}$ along atoms where adjacent units are attached to, and the end-end vector of the chain. The physical meaning of vectors $\boldsymbol{n}$, $\boldsymbol{p}$ is shown in figure 6.6(b). The order parameter is defined as the largest eigenvalue of the nematic order tensor

$$S_{\alpha\beta} = \frac{1}{2} \left\langle 3u_\alpha u_\beta - \delta_{\alpha\beta} \right\rangle , \tag{6.2}$$

where $\alpha, \beta = x, y, z$, $\boldsymbol{u}$ is a unit vector of interest, $< \cdots >$ denotes time and ensemble average. The nematic order parameters are referred to as $S_{\boldsymbol{n}}$, $S_{\boldsymbol{p}}$, and $S_{\text{end-end}}$.
The results are summarized in table 6.1. The fiber has rather large nematic order parameter $S_{\text{end-end}}$, which indicates good alignment of chains along the x axis. The slice has a negative value of the order parameter, which is in agreement with the (practically isotropic) molecular ordering in the xz

plane. Finally, the isotropic morphology has a rather small value of the order parameter, which is again in good agreement with the isotropic distribution of molecular orientations in the box. The same trend can be observed for the oxidized polypyrrole morphologies, where the stiffness of chains does not really affect the orientational ordering of oligomers.

6.3 Charge transport parameters

The main issue in order to simulate charge dynamics is to determine the charge transport unit. Many mechanism can lead to charge localization on a polymer: polaron self trapping [140, 141], breaks in conjugation due to chemical defects [142] or due to the torsion angle between neighboring monomers [143] and finally rapidly changing intramolecular transfer integrals leading to the dynamic disorder which is responsible for charge localization in molecular crystals [76, 80]. Three limiting cases for charge localization were considered: (i) a totally localized model where each repeat unit is a charge transporting unit, (ii) a totally delocalized model where each oligomer is a charge transporting unit, and (iii) a model where the polymer is partitioned into conjugated segments. In the latter case, the places where the oligomer is artificially cut are chosen according to the value of the torsion angle, that is when the torsion angle between two repeat units is greater than 45° (see figure 6.7). This approximation can only capture the static, conformational mechanism for charge localization. Other effects should be integrated in future extensions to the model.
When computing mobilities, the model (iii) in which a distribution in conjugation lengths is considered was applied. However, the totally localized model (i) is used to investigate the effect of changing intramolecular rates and the totally delocalized model (ii) was used in the percolation analysis for intermolecular rates whilst entirely excluding intramolecular transfer in-

morphology	neutral			oxidized		
	$S_{\boldsymbol{n}}$	$S_{\boldsymbol{p}}$	$S_{\text{end-end}}$	$S_{\boldsymbol{n}}$	$S_{\boldsymbol{p}}$	$S_{\text{end-end}}$
fiber	-0.38	0.70	0.85	-0.38	0.64	0.84
slice	0.36	-0.27	-0.33	0.34	-0.24	-0.30
isotropic	-0.14	0.13	0.16	0.10	0.10	0.14

Table 6.1: **Nematic order parameters in polypyrrole morphologies.** Order parameters were calculated for $\boldsymbol{n}$, $\boldsymbol{p}$, and end-to-end vectors which characterize average orientational ordering of repeat units in oligomers ($S_{\boldsymbol{n}}$, $S_{\boldsymbol{p}}$) and chains in a simulation box ($S_{\text{end-end}}$).

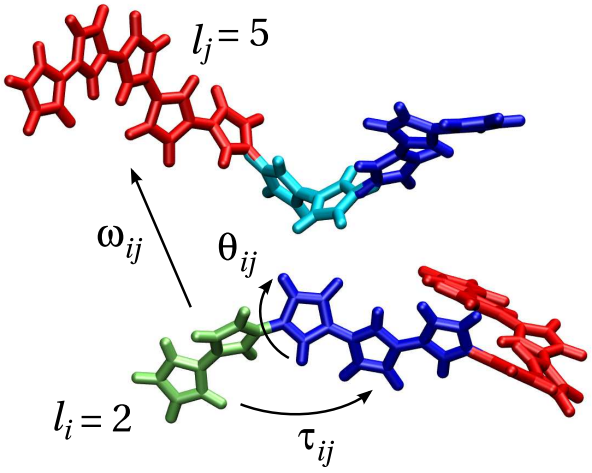


Figure 6.7: **Model for hopping on conjugated segments.** A segment is assumed to be conjugated if the torsion angle θ_{ij} is less than 45°. The length of the conjugated segment is denoted as l, the rates for intermolecular transport is given by ω_{ij} (Marcus rates) while intrachain transport is described via τ_{ij} (adiabatic rate).

tegrals.

Having defined the charge transport units, it remains to chose an expression for the charge hopping rates as well as to evaluate the necessary parameters. Here one must distinguish between the cases of transport along chains and between chains. As will be shown later, the typical internal reorganization alone is approximately 0.4 eV (depending on conjugation length) and the peak in intermolecular transfer integrals is at approximately 10 meV. Therefore it is reasonable to treat intermolecular transport in the non-adiabatic regime and use the Marcus expression to calculate hopping rates [90, 87]. On the other hand, transfer integrals along the chain can be as large as $J_0 \cos(\pi/4) \sim 0.7$ eV. Such transfer integrals put transport firmly in the adiabatic regime, possibly even in the delocalized regime, and therefore must be described by a different formula.

To calculate the necessary parameters for conjugated segments of different length, the ends of the conjugated segments were saturated with hydrogens. The chains were first optimized using the B3LYP functional and 6-311+G(d,p) basis set. Since only transport of positive charge carriers are considered, the site energy difference ΔE_{ij} is given by the difference of the HOMO energies and a contribution due to the external electric field. The HOMO values as a function of segment length are depicted in figure 6.9(a).

The reorganization energy was calculated as $\lambda_{ij} = \frac{1}{2}(\lambda_i + \lambda_j)$, where λ_i is the reorganization energy of an oligomer of length i and can be computed using the method described in sec. 3.4.
In principle it would also be possible to compute the contribution to site energy differences from electrostatic energies. However, there is a significant problem in doing this in the case when the dielectric constant is not only due to electronic polarization, as it is in the case of the oxidized polymer. In fact, for a heavily doped polyelectrolyte the optical dielectric constant can be very large - up to 20 according to some measurements [144]. Under such circumstances, decoupling the molecular dynamics from the charge dynamics becomes rather contrived as rearrangement of ions would make the largest contribution to the dielectric constant. Therefore, only the effect of conjugated segment length on site energy is considered here. In this sense, the main difference between oxidized and neutral PPy is due to chain stiffening, which will affect the morphology.

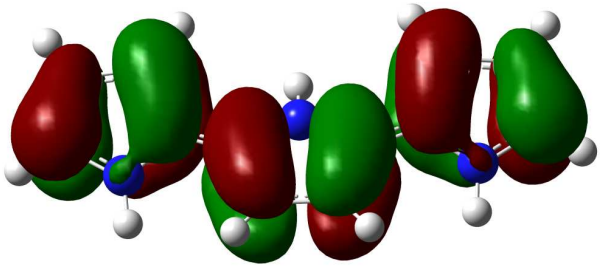


Figure 6.8: **HOMO of a polypyrrole trimer.**

Intermolecular transfer integrals J_{ij} are computed using the molecular orbital overlap method as described in sec. 3.3. The transporting orbital for positive charge carriers is the HOMO, which is shown in figure 6.8 for a polypyrrole segment. For the adiabatic intrachain transport a simple expression derived from transition state theory [90, 94] is used instead of the Marcus equation

$$\tau_{ij} = \nu \exp\left[-\frac{(\Delta E_{ij} - \lambda_{ij})^2}{4\lambda_{ij} k_\mathrm{B} T} - J_0 \cos\theta_{ij}\right]. \tag{6.3}$$

In this equation ν is a prefactor related to the frequency of the promoting mode and the relevant Franck Condon factor [90]. Simulations where charges are allowed to hop only between *monomers* have shown that a value of $10^{15}\,\mathrm{s}^{-1}$ is sufficient to stop the global mobility depending on the choice of this parameter. By making sure that this rate is sufficiently fast it is therefore ensured that the rate limiting step in charge transport simulations

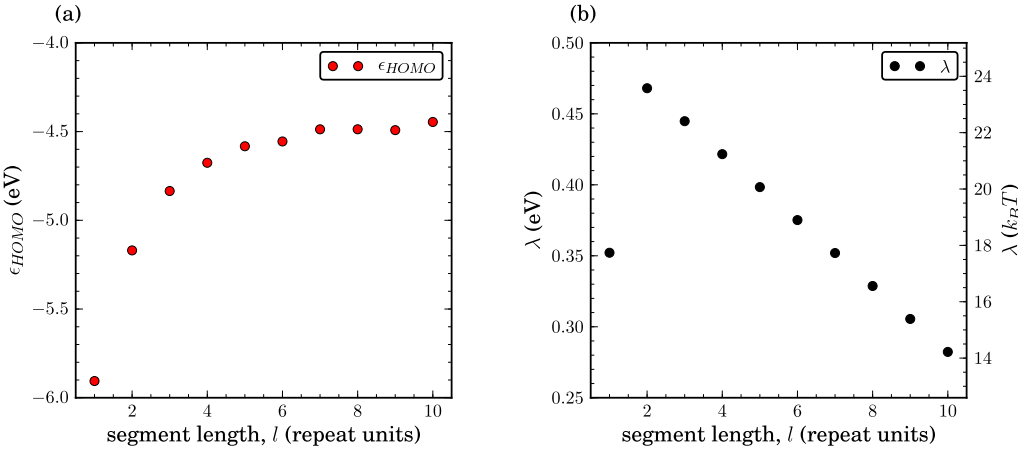


Figure 6.9: **Length dependence of site energies and reorganization energies.** a) Dependence of the HOMO energy of the oligomer on its length. HOMO orbital of a polypyrrole trimer is depicted in the inset. b) Dependence of the reorganization energy λ_i on oligomer length i.

is always an intermolecular process. The intramolecular transfer integral is modeled by the following form: $J_{\text{intra}} = J_0 \cos\theta_{ij}$, where θ_{ij} is the torsional angle between neighboring charge transport units i and j, and J_0 is chosen to have a value of 1 eV. All other symbols have the same meaning as in eq. 3.9. First, the effect of segment length on *interchain* rates ω_{ij} is analyzed. As it is shown in figure 6.9(a), the site energy ϵ_{HOMO} increases with the oligomer length, which indicates that positive charges (holes) prefer to occupy larger charge transport units. A slight complication to this is the fact that the reorganization energy λ also depends on the length of oligomer considered, as shown in figure 6.9(b). With the exception of oligomers of length 1 repeat unit (ru) the reorganization energy decreases with length of a conjugated segment. The net effect on the rate therefore is that jumps from shorter to longer segments are in general favored, but if the length difference is too large, the rates become slow due to being in the inverted regime of Marcus theory[3]. These observations are summarized in figure 6.10, which shows the hopping rate between two oligomers as a function of their size. It is clear that the rate prefactor (excluding the transfer integral) is, relatively speaking, rather insensitive to oligomer length in a large region. Essentially, energetic disorder becomes a minor effect if oligomer lengths become greater than 3 ru.
Most of the disorder introduced in the simulation is therefore a result of the

[3]The reason for the inverted regime is that an additional vibrational excitation is needed to induce the electron transfer.

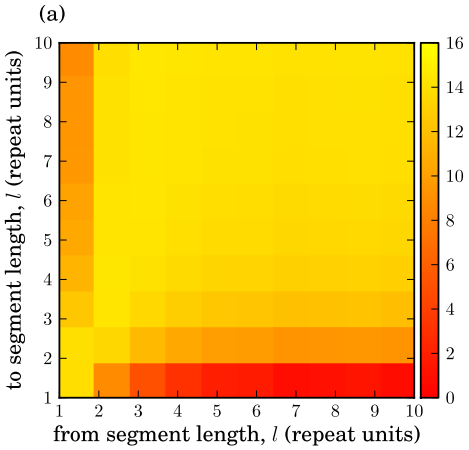


Figure 6.10: **Rates for hopping to conjugated segments of different length.** Dependence of the rate (assuming a transfer integral of 0.1 eV, in units of s^{-1}) on the length of the oligomer from which the charge is hopping (x axis) and on the length of the oligomer to which it is hopping (y-axis).

distribution of transfer integrals, shown in figure 6.11(b). Showing the distribution of transfer integrals next to the radial distribution function, figure 6.11(a), reveals how closely these two quantities are related. Note that this relation is evident only if the *logarithm* of the transfer integrals is compared to the radial distribution function, since the transfer integral decreases exponentially with increasing distance. The radial distribution function shows the density at a shell at a certain distance from the center of a monomer. The first two sharp peaks (both labeled with a letter A) correspond to the nearest intrachain neighbors (there are two peaks because the two possible orientations of an intrachain neighbor result in slightly different distances). The next peak (B) corresponds to the nearest interchain neighbor. This structure is exactly repeated in the distribution of logarithm of the transfer integrals.

One can also see, that even though the different morphologies have very different order parameters and seem very different to the naked eye, the distributions of transfer integrals are in fact rather similar. This is because transfer integrals are essentially a local property which depends only very weakly on the global ordering. The only morphology which does display a small difference in transfer integrals is the fiber morphology in the reduced form for which the peak of the transfer integrals is slightly greater. Notice that the height of the peaks in the distribution of transfer integrals is somewhat deceiving, since all distributions are normalized by the total number

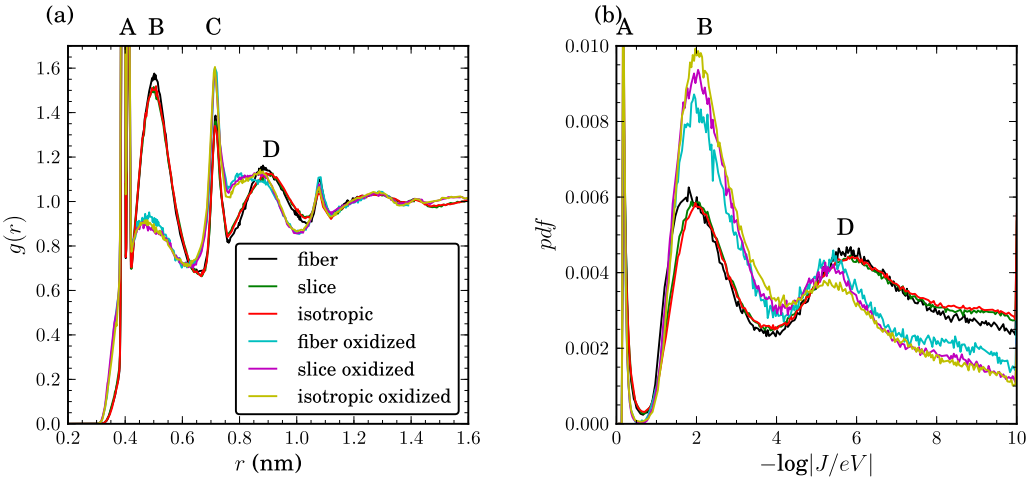


Figure 6.11: **Relation of radial distribution function and transfer integrals.** Radial distribution function for monomers (a) and distribution of logarithm of transfer integrals (b) for three different morphologies and two oxidation states. The distribution of transfer integrals is normalized to the *total* number of transfer integrals in the morphology. The peak A is due to intramolecular charge transfer between neighboring repeat units, peak B is due to the intermolecular transfer for the first shell of neighbors, peak C is due to next nearest neighbors along the chain (and therefore is absent in the distribution of the transfer intergals). Peak D corresponds to the second shell of neighbors participating in the intermolecular charge transfer.

of neighbors. A higher first peak is likely to represent fewer second nearest neighbors within the cutoff distance rather than more nearest neighbors. This is why the oxidized geometries, with their longer average lengths, tend to have a higher first peak.

Oxidation is believed to have two main effects on charge mobility: it makes the polymer backbones significantly stiffer and it also changes the mediators of charge transport from polarons to bipolarons [124, 145, 125, 146, 126, 127]. If one considers two charge transporting units that can support either polarons or bipolarons, one can imagine two ways that charges move: either two charges move at once or one moves at a time. It can be distinguished which of these cases is more likely by considering the transfer integral related to each case. Using Slater rules and within the frozen orbital approximation it is deduced that in the latter case the transfer integral will correspond to the exchange integral of those two orbitals, whereas in the former case the transfer integral will correspond again to the expectation value of the Fock matrix. Since the expectation value of the Fock matrix is always greater than

the exchange integral it is much faster for a single charge to be exchanged between molecules. The mobility that is computed in the case of oxidized polypyrrole is the mobility of a single charge moving on a background of overall charged oligomers. The charge of each oligomer is deduced from its length; oligomers of length one to three will vary from having one charge to having two charges, lengths four to six vary from having two charges to three charges and longer oligomers vary from three to four charges. The transfer integrals are computed using the relevant molecular orbitals.

6.4 Charge transport simulation

Once the position of atoms is determined from MD, the oligomers are segmented into conjugated units based on the torsion angle between subsequent monomers. If the absolute value of this torsion angle is greater than 45°, the monomers belong to different conjugated segments. Then segments of the MD run are substituted with structure optimized in quantum chemical calculations (see sec. 3.7 for details). Each oligomer has its position and orbitals adjusted to the position and orientation of the monomer to ensure a close match to the atomistic morphology. This special substitution solves two issues: first, deviations in bond stretching which occur due to errors in the force field are compensated, and second, the structure for the further evaluation corresponds to the QM optimized structure for which the orbitals were calculated. Without this substitution, the diabatic picture of charge transfer would be violated.
Transfer integrals, site energies and rates of charge transfer are then computed, using the methodologies described chapter 3. The charge dynamics is simulated using kinetic Monte Carlo algorithm [147]. The mean of the time-average drift velocity of charges in an electric field over various MD snapshots is calculated to ensure good statistics. All simulations were performed using the VOTCA package [103].
Table 6.2 shows the mobility for the different morphologies averaged over 100 MD snapshots. Only mobilities greater than $10^{-6}\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$ were taken into account and 7.4% of all the simulations were neglected. This cutoff must be chosen because if a KMC simulation is started with a charge on an conjugated unit of length 1 ru or 2 ru it is very unlikely to escape. As argued in the sec. 6.6, this is justified because torsional angles fluctuate on a time scale significantly faster than that of the very slow hops. In other words, the charge localized on a short transport unit is practically instantaneously delocalized by vibrations of the backbone.
With the mobilities at hand, one can estimate the conductivities of PPy

morphology	mobility x	mobility y	mobility z
fiber	2.3 ± 1.6	1.8 ± 1.0	1.7 ± 0.9
slice	1.3 ± 0.4	0.7 ± 0.3	1.2 ± 0.4
iso	0.8 ± 0.3	0.7 ± 0.2	0.9 ± 0.3
fiber ox	1.9 ± 1.8	1.5 ± 1.1	1.1 ± 1.0
slice ox	2.6 ± 2.1	2.4 ± 1.5	3.0 ± 2.1
iso ox	3.5 ± 1.7	3.1 ± 1.5	3.4 ± 1.5

Table 6.2: **Charge mobilities.** Mobilities in different directions for the various morphologies, in $10^{-2}\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$. Mobility was computed from 100 frames, with an electric field of $10^5\,\mathrm{V\,m^{-1}}$, considering a distribution of conjugated fragments and averaging only those values greater than $10^{-6}\ \mathrm{cm^2\,V^{-1}\,s^{-1}}$.

which have been measured. Assuming that the charge density is of three electrons per oligomer of length 10, a charge carrier concentration of the order of $2.5 \times 10^{21}\,\mathrm{cm^{-3}}$ should be expected. Using this concentration the simulated mobilities can be converted to a conductivity in the order of $12\,\mathrm{S\ cm^{-1}}$, a value which is in the range measured for this material [123].

6.5 Cluster analysis

In previous studies on one dimensional transport in discotic liquid crystals [81, 85, 25], relating the transfer integral distribution to the mobility was relatively easy: since the transport is one dimensional, the mobility is essentially set by the tail of low transfer integrals. In three dimensions it is much harder to state a priori what will be the rate limiting step when a charge traverses such a box.

One way of doing this is to look at the point when transfer integrals "percolate" the box. Percolation is defined in the following way: one first partitions the system on clusters of molecules which are connected by transfer integrals of at least a certain value J_{min}. Then the size of the largest cluster in the box is plotted as a function of J_{min}. Percolation occurs when J_{min} is large enough for the biggest cluster to span the entire box. The percolation threshold is obtained by looking at the relative size of the largest cluster which is defined as a ratio of number of molecules in the largest cluster to the total number of molecules in the simulation box. This calculation was carried out for the transfer integral between oligomers of length 10. If a distribution of conjugation lengths with different length would be considered, the interpretation would be complicated by the fact that oligomers of different length

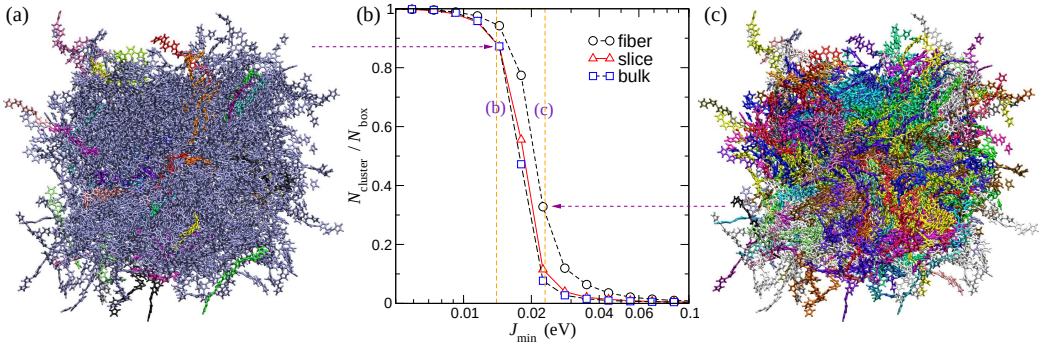


Figure 6.12: **Percolation analysis.** (a) Simulation snapshot after percolation ($J_{\min} = 23\,\text{meV}$). Different clusters are depicted by different colors. The largest cluster spans the whole box. (b) Relative to the simulation box size of the largest cluster given a certain value $J_{\min}$. (c) Simulation snapshot before percolation ($J_{\min} = 14\,\text{meV}$). Clusters are not well connected and are small compared to the box size.

have different aspect ratios and hence different percolation thresholds. The dependence of the relative cluster size on the transfer integral is shown in figure 6.12.

From figure 6.12 it is clear that the value of the transfer integral when the largest interconnected cluster starts spanning the whole box is approximately 18 meV for the slice and isotropic morphologies and slightly over 20 meV for the fiber morphology. This value also corresponds to the first peak in the distribution of intermolecular transfer integrals, peak B as shown in figure 6.11. In order to convert this limiting transfer integral to a rate for the slowest step, the expression from Marcus theory in eq. 3.9 can be used. In the case when ΔE_{ij} is small and due only to the electric field F, i. e. $\Delta E_{ij} = Fd$, the rate equation for the forward, ω_{ij}^{f}, and backward, ω_{ij}^{b}, rates takes the following form

$$\omega_{ij}^{\text{f,b}} = \frac{J_{ij}^2}{\hbar}\sqrt{\frac{\pi}{k_{\text{B}}T\lambda_{ij}}}\exp\left[-\frac{\lambda_{ij}}{4k_{\text{B}}T}\right]\left(1 \pm \frac{Fd}{2k_{\text{B}}T}\right), \tag{6.4}$$

where d is the distance between neighboring sites (at the cluster boundary) and all other symbols have the same meaning as in eq. 3.9. The mobility can then calculated as

$$\mu = \frac{d(\omega^{\text{f}} - \omega^{\text{b}})}{F} = \frac{J_{ij}^2 d^2}{\hbar k_{\text{B}}T}\sqrt{\frac{\pi}{k_{\text{B}}T\lambda_{ij}}}\exp\left[-\frac{\lambda_{ij}}{4k_{\text{B}}T}\right]. \tag{6.5}$$

In other words, one obtains an expression for mobility which is independent of the electric field. Note also that it was assumed that there is no energetic

disorder and that all hops are identical. Using $d = 5 \times 10^{-8}\,\mathrm{cm}$ (the first maximum of the radial distribution function), $J_{ij} = 0.018\,\mathrm{eV}$ (value which corresponds to the percolation threshold), $k_\mathrm{B}T = 0.0259\,\mathrm{eV}$, $\lambda_{ij} = 0.35\,\mathrm{eV}$ (average value of the reorganization energy), and $\hbar = 6.57 \times 10^{-16}\,\mathrm{eV}$, an estimate of the mobility of $3 \times 10^{-2}\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$ is obtained. Comparing this value to the mobilities simulated using the KMC method in table 6.2, it is evident that this estimate of the mobility is indeed a good one.
In fact, all the morphologies have rather similar mobilities. In other words, it seems that, even though the morphologies are very different on a global scale, the mobilities are limited by J_min which is similar in all morphologies.

6.6 Validity of the model and outlook

An important shortcoming of this model is that it neglects the fact, that charge transport units are not fixed in time due to fluctuations (dynamics) of the torsion angle between repeat units. These fluctuations occur on timescales of hundred femtoseconds, as can be seen from the distribution of monomer lifetimes shown in figure 6.13(a).

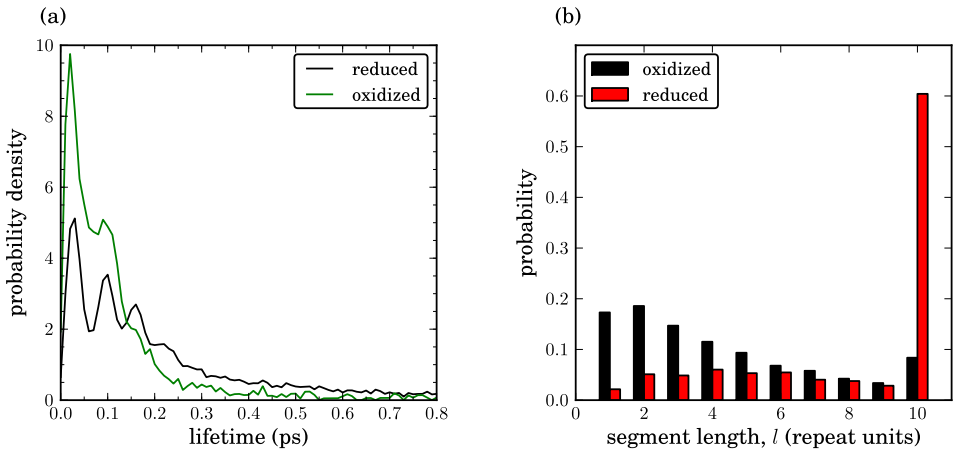


Figure 6.13: **Conjugated segments distributions.** All results were obtained at 300K. (a) Probability distribution of the lifetime of conjugated segments of length 1. The lifetime of conjugated segments is in the order of couple tenths of picoseconds. (b) Distribution of lengths of conjugated segments. In the reduced case, dimers are the most frequently appearing conjugated segments. In the oxidized case, most of the chains are flat and most frequent length of a conjugated segment is the total length of an oligomer (10 ru).

The lifetime analysis was performed by doing a short MD run of 40 ps and saving snapshots at each MD step (2 fs) at 300K. The lifetime was determined by book-keeping the creation/annihilation of charge units over time due to the variation of torsion angles. The normalized histogram was calculated using a 10 fs grid spacing. In the reduced case, only 1.5% of monomers live longer than 5 ps. In the oxidized case, the lifetime of segments of all lengths (except 10-mers) is shorter than 5 ps.
An important difference of the reduced and oxidized morphologies is the distribution of conjugated segment lengths. As depicted in figure 6.13b, the oxidized morphologies consist of mainly conjugated segments of 10 repeat units while in the reduced case short segments are favored.
Because of the short lifetimes, special care has to be taken if charges are trapped on a short living segment. Figure 6.10 shows that, if a charge resides on an conjugated segment of length one and is surrounded by segments of length greater than 3, the average hopping time will be much longer than the persistence time of this conjugated segment. Therefore, it is important to disregard those simulations where events occur on much slower timescales. Furthermore, a residing charge would lead to chain flattening, similar to chain stiffening upon oxidation. This will additionally shorten the lifetime of a conjugated segment, helping the charge to escape.

6.7 Conclusions and outlook

To summarize, a combination of quantum and classical levels of description is used to investigate the link between morphology and charge transport in six different disordered polymer assemblies. The most important structural parameter is shown to be the local density, as described by the radial distribution function. A method to analyze the three dimensional assembly of polymers is discussed and the concept of a transfer integral threshold is introduced.
The main observation is that for non-crystalline conjugated polymers charge carrier mobility is rather insensitive to global orientational molecular ordering. In fact the key morphology parameter as far as mobility is concerned is not the regularity of the arrangements of chains, but their local packing. This is an important observation and rises doubts on an often held opinion that good alignment and orientation of chains is important in order to improve the mobility of a conjugated polymer, at least in the limit of relatively high disorder, with no crystallization.
It is also important to emphasize that the proposed model (as any model) makes several approximations. The main ones are: (i) the introduction of

conjugated segments based uniquely on the cutoff of the dihedral angle and (ii) static treatment of chains during charge transport. The latter point is partially addressed by estimating an average lifetime of a conjugated segment of length n.

Since a charge carrier residing on a conjugated segment, similar to oxidation, flattens the chain, a possible extension to the model would be to couple MD simulations to the charge carrier dynamics. However, since an explicit treatment of the electronic structure is computationally expensive, one can try to adjust the force-field parameters "on the fly", while charge carriers travel through the sample, and simultaneously propagate atoms using MD and charge carriers using the kinetic Monte Carlo method.

The second shortcoming of the model it that structure optimization and orbital coefficient calculations are needed for each possible length of a conjugated segment. This was possible for the oligomers (10 repeat units maximum) used in this study. In case of polymers with big conjugation lenghts, this is computationally not feasible. In order to resolve this issue, the charge patching method [148] can be used, where conjugated segments are constructed based on short building blocks, e.g. monomers.

Conclusion and discussion

In this work, a model that combines quantum and classical methods (i.e. molecular dynamics, coarse-graining, transfer integral calculation, Marcus theory and the kinetic Monte Carlo algorithm) was developed and implemented in a software package. In particular, its aim is to link charge carrier mobility to realistic morphologies in conjugated polymers. To this end, morphologies at an atomistic level have to be obtained first, and then charge dynamics can be simulated using this morphology.

Atomistic morphologies can, in principle, be generated using atomistic molecular dynamics simulations. However, the accessible range of time- and length-scales is limited in this case. Systematic coarse-graining can be used to overcome these limitations. Boltzmann inversion, iterative Boltzmann inversion, the inverse Monte Carlo method and the force-matching method were implemented in the Versatile Object-oriented Toolkit for Coarse-graining Applications (VOTCA). The package follows the rules of modular software design and provides a flexible platform for the consistent implementation of coarse-graining techniques to allow for their direct comparison. VOTCA is available for public use under an open source software license (www.votca.org).

To illustrate the package functionality and to compare the different methods, coarse-graining of a set of model systems, namely SPC/E water, methanol, liquid propane and a single chain of hexane, was performed. It was shown that all implemented methods have advantages as well as shortcomings. For example, inverse Monte Carlo has an update function which is more efficient than that of the iterative Boltzmann inversion method. However, the inverse Monte Carlo update is very sensitive to the system size and statistical averaging. Using force-matching, which is not an iterative method, can lead to unphysical potentials if the basis set used to represent the coarse-grained potential energy surface is incomplete.

In the context of future development of the VOTCA package, an implementation of the simplex algorithm which allows the parametrization of the coarse-grained model for thermodynamic properties, for example free energies, is planned. In addition, an iterative method which combines the sta-

bility of the local update of the iterative Boltzmann inversion method and the efficiency of the inverse Monte Carlo update could be beneficial for the development of coarse-grained models.
Another planned feature of the VOTCA package is back-mapping to atomistic details. A coarse-grained model only represents a simplified description of the system. In order to generate equilibrated morphologies at an atomistic level, atomistic details have to be reintroduced after equilibration at a coarse-grained level. Then only a short equilibration run at atomistic level is required to locally equilibrate the sample.
In the second part of this thesis, the software package was used to develop a model to simulate charge mobility in atomistic morphologies of conjugated polymers. The link between morphology and charge mobility in six different assemblies of doped and neutral states of polypyrrole, a conjugated polymer, was investigated. Charge transport was modeled as hopping between conjugated segments, where a segment represents a diabatic state for charge transfer. Inter-chain transport was treated within the high temperature non-adiabatic limit of Marcus theory while an adiabatic rate expression was used for intra-chain transport. Charge dynamics was then simulated using the kinetic Monte Carlo algorithm.
The main observation was that for non-crystalline conjugated polymers, charge carrier mobility is insensitive to the global orientational molecular ordering. In fact, the key morphology parameter, as far as mobility is concerned, is not the regularity of the arrangements of chains, but the local packing. The obtained mobilities, when converted to conductivities, are of the order same of magnitude as the experimentally measured values.
A natural extension of this work will be to apply the developed framework to study problems related to organic bulk heterojunction solar cells. In the latter, a mixture of two different materials is used in order to create a donor-acceptor interface which triggers charge separation. One of the challenges here is to properly describe the structure of the interface and the non-equilibrium morphology as a function of processing conditions.

Appendix A

Software design

A.1 UML class diagrams

The Unified Modeling Language (UML) is a graphic notation, that helps to describe the design of a software package. Here, only a subset of the full notation is used to sketch the functionality of the VOTCA package. The diagrams in this work should be seen as an outline of the software design rather than its accurate specification. This section outlines the notation for class diagrams providing enough background to understand the graphs described in chapter 4. A more detailed essay of UML can e.g. be found in [149].

A UML *class diagram* describes the types of objects in the program as well as the relationship among them. The basic notation for a class diagram is outlined in figure A.1, where boxes represent classes in the system and lines/arrows their relationship. Four different relationships are used here to describe the package. The composition is a way to express that an object is composed of different elements, which means that the object serves as a container for the element. This can be seen in a similar way as a car is composed of an engine and wheels. The second relationship depicted in figure A.1 expresses an association with an object. The difference between composition and association can be confusing, it has do do with ownership and what happens if one object is destroyed. It can be best explained by the following example: a topology is composed of several atoms. If the topology is deleted, all atoms will be deleted as well. In this respect, a composition is much stronger than an association. An example for association would be a pair of atoms in the neighbor list. The pair references two atoms, but if the pair will be destroyed, it will not influence the atoms.

The last two arrows denote inheritance. In C++, there is no clear differ-

ence between the notations for *inherits from* and *implements*. The distinction should be seen in the following way: an interface definition contains no logic but just provides the interface to access the functionality which is implemented in the child classes (only abstract functions). On the opposite, if *inherits from* is used, the parent class itself might provide functionality which is then extended in the child classes.

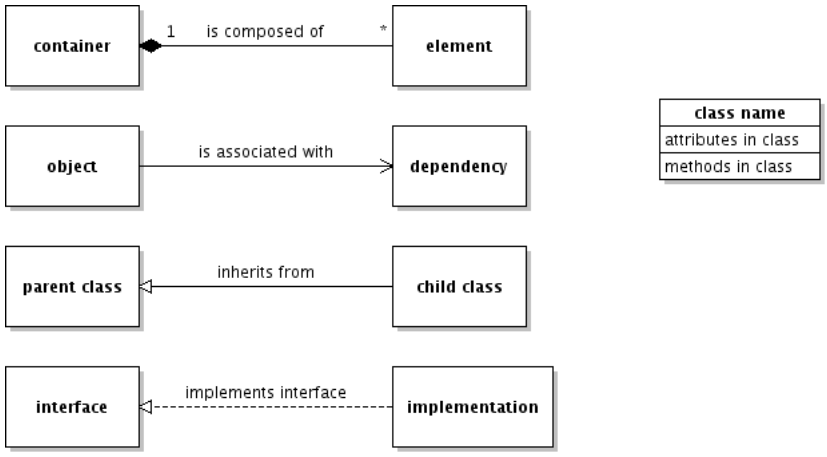


Figure A.1: **Symbols in UML class diagrams.** Boxes denote classes and arrows their relationship. Numeric values and asterisks on arrows denote the multiplicity, which is an indication of how many objects may fill the property.

A.2 The factory design pattern

Two of the main concepts to allow for abstraction and modularity in object-oriented programs is *inheritance* and *polymorphism*: a child objects can implement and extend the functionality of a parent object (*inheritance*). Using virtual functions, the child class has the same signature (methods) as the parent object. Hence, it can be accessed in an identical way, while the methods provide specialized actions based on the type of object (*polymorphism*).
The aforementioned concepts allow to dynamically decide during runtime, which object is created, and therefore, which functionality is used. A typical example where, polymorphism can useful is a file reader object for e.g. reading molecular dynamics (MD) trajectories. The interface to access the content is defined first, e.g. to read coordinates of atoms. However, these are stored in a different way, depending on the type of file to be read. Then, reading from the file is implemented in a child class, where one class is derived for each

specific format. Reading of MD data can then be performed for all formats in a uniform way defined by the interface.
Depending on the file format to be read, a different reader object has to be created. However, directly creating the object by specifying its class is a static construction. Here, the factory pattern [104] can be used to encapsulate the creation. In other words, a factory is an object to create other objects sharing the same interface without explicitly specifying a class [105]. All objects that can be created are stored in a pool and are referred to by identifiers, e.g. keywords. For the specific example mentioned above, this would be the file type.
Another advantage of the factory pattern is that plugins can be easily developed without recompiling a program. The implementation of the derived object is done in a dynamically linked library. This allows for extending a program by for example placing such a dynamically linked library in a folder and it is then loaded during execution of the main program.

Appendix B

Structure of the VOTCA code

The VOTCA project is divided into several modules, which are organized as separate repositories of a version control system: *tools*, *CSG*, *MOO*, *KMC* and *MD2QM*. Each module groups the functionality for a specific field of application.
The repository for *tools* contains generic helper functions and classes which are accessed from every module in the project. Tools is independent of other VOTCA modules. The coarse-graining applications and topology handling are implemented in *CSG*. *MOO* and *KMC* provide the functionality to evaluate transfer integrals and run kinetic Monte Carlo simulations, respectively. *CSG*, *MOO* and *KMC* are individual modules and they are independent. The last module, *MD2QM*, combines *tools*, *CSG*, *MOO* and *KMC* to implement the charge transport framework described in sec. 3.7.
Depending on the purpose, only a subset of modules is needed. The coarse-graining applications, which are already available for public use, combine *tools* and *CSG*. In the following, a brief outline of each module is given. However, since the project is rapidly evolving, this description is only a snapshot of the current status. For up to date and complete information, the reader is referred to the project page (www.votca.org).

B.1 Tools

The module *tools* provides a library of helper functions and classes which are used throughout the whole package. The implementation of the object factory, which is described in sec. A.2, is also a part of the *tools* library.
The most frequently used classes are the vector (**vec**) and matrix (**matrix**) classes, which implement vector and matrix operations in three dimensional space. These classes overload the standard operators which allows for their

simple usage and improves the readability of the implemented calculations.

Another important class is the **Property** object. It provides an interface for storing, passing and accessing arbitrary options read from an xml-file. Using this class, options which are not known a priori can be read and forwarded to the corresponding sub-parts of an application.

The remaining components are generic helper functions. Examples are reading, parsing and organization of data (such as tables) and calculating statistics (such as histograms or correlations). Many classes are interface wrappers for existing libraries. These wrappers allow to substitute libraries by alternative implementations without major rewriting of code.

B.2 CSG

The *CSG* repository contains the coarse-graining applications. Tools and *CSG* have been released as the VOTCA coarse-graining package [103]. *CSG* provides a library, in which topology handling, mapping and the analysis engine are implemented. Its basic layout and most of the classes are described in chapter 4. In general, *CSG* should be seen as a library for trajectory analysis. The library is used to implement several coarse-graining applications:

program	**description**
csg_map	Map a trajectory from atomistic to coarse-grained level.
csg_boltzmann	Analyze single chain trajectories, create histograms and Boltzmann inverted potentials for bonded interactions.
csg_fmatch	Application to run force-matching.
csg_stat	Evaluate bonded- and non-bonded statistics, calculates radial distribution functions and bonded probability distributions, evaluates cross-correlations for IMC. This application is mainly used internally by the scripting framework but can also be used manually.
csg_dump	Debugging tool to check topology data which is read.
csg_gmxtopol	Convert an atomistic topology and mapping file to a Gromacs topology to help setting up input files for the coarse-grained run.
csg_resample	Resample grid spacing of a table, e.g. tabulated potential. Mainly used by the scripting framework.

In addition, the scripting framework provides the following programs:

program	description
csg_inverse	Iterative framework to perform IBI and IMC.
csg_call	Manually run an internal script of the scripting framework.

Details on the usage of the applications can be found in the VOTCA manual or on www.votca.org.

B.3 MOO and KMC

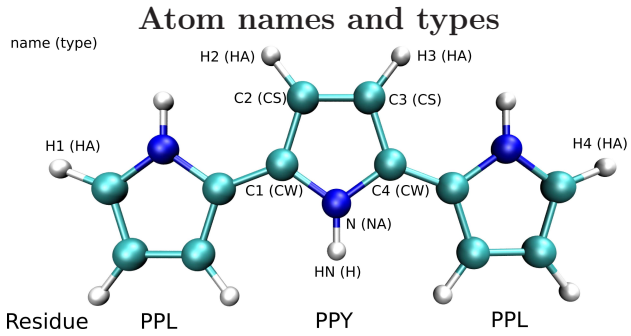
The modules *MOO* to calculate transfer integrals and *KMC* to run a kinetic Monte Carlo simulation were developed at Imperial College, London. To simplify access to the *MOO* library, an interface to create charge units and calculate transfer integrals between them was added. Therefore, interfacing is done in two classes, the **JCalc** which controls the creation of charge units and the **CrgUnit**, which represents a hopping site. **JCalc** also wraps the calculation of transfer integral between charge units.
In addition to the library in *MOO*, an application to calculate transfer integrals for a pair of molecules having a certain relative orientation is provided. For evaluation of whole topologies which consist of several molecules, *MD2QM* was developed.

B.4 MD2QM

The charge transport simulations described in sec. 3.7 require all modules mentioned: *CSG* for topology handling, *MOO* to evaluate transfer integrals and *KMC* to simulate charge dynamics. As pointed out in sec. 4.4, *MD2QM* inherits the topology handling classes of *CSG* and integrates *MOO* and *KMC*. The design is described in chapter 4. However, a complete list of application names is not yet available.

Appendix C

Force-field parameters for polypyrrole



Bonds

bond	r_0	k
CW-CW	0.14423	428441.6
CS-CS	0.14240	392459.2
CW-CS	0.13670	456892.8
CS-HA	0.10800	307105.6
CW-HA	0.10800	307105.6
NA-H	0.10100	363171.2
CW-NA	0.13810	357313.6

Angles

angle	ϕ_0	k
CS-CW-NA	106.500	585.760
CW-NA-CW	109.800	585.760
CW-CS-CS	107.300	585.760
CW-CS-HA	125.700	292.880
HA-CS-CS	127.500	292.880
CW-NA-H	125.100	292.880
NA-CW-CW	121.900	527.000
CS-CW-CW	131.600	527.000
NA-CW-HA	121.600	292.880
CS-CW-HA	132.100	292.880

Atomic charges

PPY end group

Atom	neutral	oxidized
N	-0.219	-0.426
HN	0.277	0.449
C1	-0.165	0.022
H1	0.157	0.185
C2	-0.172	-0.169
H2	0.124	0.171
C3	-0.195	-0.173
H3	0.127	0.173
C4	0.066	0.068

PPY

Atom	neutral	oxidized
N	-0.3	-0.501
HN	0.3	0.383
C1	0.1	0.235
C2	-0.24	-0.185
H2	0.14	0.159
C3	-0.24	-0.185
H3	0.14	0.159
C4	0.1	0.235

Ryckeard-Belleman-Dihedral NA-CW-CW-NA

	$\mathbf{C_0}$	$\mathbf{C_1}$	$\mathbf{C_2}$	$\mathbf{C_3}$	$\mathbf{C_4}$	$\mathbf{C_5}$
neutral	16.0392	-2.48322	-25.8586	3.30161	11.034	-2.18395
oxidized	243.6699	-61.1535	-147.8553	154.0901	75.7420	-102.4387

Bibliography

[1] P. Calvert. Inkjet Printing for Materials and Devices. *Chemistry of Materials*, **13**(10):3299–3305, 2001.

[2] 2005. http://www.smh.com.au/news/breaking/electronicpaper-that-bends/2005/07/15/1120934404860.html.

[3] 2004. http://www.livescience.com/technology/041224-_solar_panels.html.

[4] 2007. http://www.msnbc.msn.com/id/18867499/.

[5] A. Murphy and J. Frechet. Organic semiconducting oligomers for use in thin film transistors. *Chem. Rev.*, **107**(4):1066–1096, 2007.

[6] J. Roncali. Synthetic principles for bandgap control in linear pi-conjugated systems. *Chem. Rev.*, **97**(1):173–205, 1997.

[7] J. Wu, W. Pisula, and K. Müllen. Graphenes as potential material for electronics. *Chem. Rev.*, **107**(3):718–747, 2007.

[8] P. Leclere, M. Surin, P. Viville, R. Lazzaroni, A. Kilbinger, O. Henze, W. Feast, M. Cavallini, F. Biscarini, A. Schenning, and E. Meijer. About oligothiophene self-assembly: From aggregation in solution to solid-state nanostructures. *Chem. Mat.*, **16**(23):4452–4466, 2004.

[9] P. Leclere, M. Surin, P. Jonkheijm, O. Henze, A. Schenning, F. Biscarini, A. Grimsdale, W. Feast, E. Meijer, K. Müllen, J. Bredas, and R. Lazzaroni. Organic semi-conducting architectures for supramolecular electronics. *Eur. Polym. J.*, **40**(5):885–892, 2004.

[10] P. Blom and M. Vissenberg. Charge transport in poly(p-phenylene vinylene) light-emitting diodes. *Mat. Sci. Eng. Rep.*, **27**:53–94, 2000.

[11] R. Kroon, M. Lenes, J. Hummelen, P. Blom, and B. De Boer. Small bandgap polymers for organic solar cells (polymer material development in the last 5 years). *Polymer Reviews*, **48**(3):531–582, 2008.

[12] G. Wegner. Polymers with metal-like conductivity - a review of their synthesis, structure and properties. *Angew. Chem.*, **20**(4):361–381, 1981.

[13] D. Kumar and R. Sharma. Advances in conductive polymers. *Europ. Polym. J.*, **34**(8):1053–1060, 1998.

[14] A. MacDiarmid. Nobel Lecture: "Synthetic metals": A novel role for organic polymers. *Rev. Mod. Phys.*, **73**(3):701–712, 2001.

[15] J. Kalinowski. *Organic Light-Emitting Diodes: Principles, Characteristics & Processes*. Marcel Dekker Inc, collectors ed/. edition, 2004.

[16] I. Kymissis. *Organic Field Effect Transistors: Theory, Fabrication and Characterization*. Springer-Verlag Gmbh, 1 edition, 2009.

[17] C. Brabec, U. Scherf, and V. Dyakonov. *Organic Photovoltaics: Materials, Device Physics, and Manufacturing Technologies*. Wiley-VCH Verlag GmbH & Co. KGaA, 1. auflage edition, 2008.

[18] I. McCulloch, M. Heeney, M. Chabinyc, D. DeLongchamp, R. Kline, M. Coelle, W. Duffy, D. Fischer, D. Gundlach, B. Hamadani, R. Hamilton, L. Richter, A. Salleo, M. Shkunov, D. Sporrowe, S. Tierney, and W. Zhong. Semiconducting Thienothiophene Copolymers: Design, Synthesis, Morphology, and Performance in Thin-Film Organic Transistors. *Adv. Mat.*, **21**:1091–1109, 2009.

[19] A. van de Craats, J. Warman, A. Fechtenkotter, J. Brand, M. Harbison, and K. Müllen. Record charge carrier mobility in a room-temperature discotic liquid-crystalline derivative of hexabenzocoronene. *Adv. Mat.*, **11**(17):1469–1472, 1999.

[20] H. Sirringhaus, P. Brown, R. Friend, M. Nielsen, K. Bechgaard, B. Langeveld-Voss, A. Spiering, R. Janssen, E. Meijer, P. Herwig, and D. de Leeuw. Two-dimensional charge transport in self-organized, high-mobility conjugated polymers. *NATURE*, **401**(6754):685–688, 1999.

[21] Y. Kim, S. Cook, S. Tuladhar, S. Choulis, J. Nelson, J. Durrant, D. Bradley, M. Giles, I. Mcculloch, C. Ha, and M. Ree. A strong regioregularity effect in self-organizing conjugated polymer films

and high-efficiency polythiophene: fullerene solar cells. *Nat. Mat.*, **5**(3):197–203, 2006.

[22] W. Brütting. *Physics of Organic Semiconductors*. Wiley-VCH, 1. auflage edition, 2005.

[23] X. Yang and J. Loos. Toward High-Performance Polymer Solar Cells: The Importance of Morphology Control. *Macromolecules*, **40**(5):1353–1362, 2007.

[24] G. Yu, J. Gao, J. C. Hummelen, F. Wudl, and A. J. Heeger. Polymer Photovoltaic Cells: Enhanced Efficiencies via a Network of Internal Donor-Acceptor Heterojunctions. *Science*, **270**(5243):1789–1791, 1995.

[25] X. Feng, V. Marcon, W. Pisula, M. Hansen, J. Kirkpatrick, F. Grozema, D. Andrienko, K. Kremer, and K. Müllen. Towards high charge-carrier mobilities by rational design of the shape and periphery of discotics. *Nat. Mat.*, **8**(5):421–426, 2009.

[26] S. Tiwari and N. Greenham. Charge mobility measurement techniques in organic semiconductors. *Optical and Quantum Electronics*, **41**(2):69–89, 2009.

[27] H. Scher and E. W. Montroll. Anomalous transit-time dispersion in amorphous solids. *Phys. Rev. B*, **12**:2455, 1975.

[28] P. Borsenberger, E. Magin, M. van der Auweraer, and F. de Schryver. The Role of Disorder on Charge Transport in Molecularly Doped Polymers and Related Materials. *phys. stat. sol.(a)*, **140**:9, 1993.

[29] M. Grünewald, B. Pohlmann, B. Movaghar, and D. Würtz. Theory of non-equilibrium diffusive transport in disordered materials. *Philos. Mag. B.*, **49**:341–356, 1984.

[30] B. Movaghar, M. Grünewald, B. Ries, H. Bässler, and D. Würtz. Diffusion and relaxation of energy in disordered organic and inorganic materials. *Phys. Rev. B.*, **33**:5545, 1986.

[31] G. Horowitz. Organic Field-Effect Transistors. *Advanced Materials*, **10**(5):365–377, 1998.

[32] V. Coropceanu, J. Cornil, D. A. da Silva Filho, Y. Olivier, R. Silbey, and J. Brédas. Charge Transport in Organic Semiconductors. *Chemical Reviews*, **107**(4):926–952, 2007.

[33] C. Tanase, P. W. M. Blom, D. M. de Leeuw, and E. J. Meijer. Charge carrier density dependence of the hole mobility in poly(p-phenylene vinylene). *physica status solidi (a)*, **201**(6):1236–1245, 2004.

[34] D. Sholl and J. A. Steckel. *Density Functional Theory: A Practical Introduction*. John Wiley & Sons, 1. auflage edition, 2009.

[35] D. C. Rapaport. *The Art of Molecular Dynamics Simulation*. Cambridge University Press, 2 edition, 2004.

[36] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision C.02, 2004. Gaussian, Inc., Wallingford, CT, 2004.

[37] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theo. Comp.*, **4**(3):435–447, 2008.

[38] W. Jorgensen and J. Tirado-Rives. Development of the OPLS-AA force field for organic and biomolecular systems. *J. Chem. Soc., Abstr.*, **216**:U696–U696, 1998.

[39] P. Cieplak, C. I. Bayly, I. R. Gould, J. K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, and W. D. Cornell. A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *Journal of the American Chemical Society*, **117**:5179–5197, 1995.

[40] N. Allinger, Y. Yuh, and J. Lii. Molecular mechanics. The MM3 force field for hydrocarbons. 1. *Journal of the American Chemical Society*, **111**(23):8551–8566, 1989.

[41] D. York, L. Pedersen, and T. Darden. Particle Mesh Ewald - an N.log(N) method for Ewald sums in large systems. *The Journal of chemical physics*, **98**(12):10089–10092, 1993.

[42] W. Tschöp, K. Kremer, J. Batoulis, T. Burger, and O. Hahn. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polymerica*, **49**:61–74, 1998.

[43] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, and M. L. Klein. A coarse grain model for phospholipid simulations. *J. Phys. Chem. B*, **105**(19):4464–4470, 2001.

[44] C. F. Abrams and K. Kremer. Combined coarse-grained and atomistic simulation of liquid bisphenol A-polycarbonate: Liquid packing and intramolecular structure. *Macromolecules*, **36**(1):260–267, 2003.

[45] T. Murtola, E. Falck, M. Patra, M. Karttunen, and I. Vattulainen. Coarse-grained model for phospholipid/cholesterol bilayer. *J. Chem. Phys.*, **121**(18):9156–9165, 2004.

[46] S. Izvekov and G. A. Voth. Multiscale coarse graining of liquid-state systems. *J. Chem. Phys.*, **123**(13):134105, 2005.

[47] Q. Sun and R. Faller. Systematic coarse-graining of a polymer blend: Polyisoprene and polystyrene. *J.Chem. Theo. Comp.*, **2**(3):607–615, 2006.

[48] V. A. Harmandaris, N. P. Adhikari, N. van der Vegt, and K. Kremer. Hierarchical modeling of polystyrene: From atomistic to coarse-grained simulations. *Macromolecules*, **39**(19):6708–6719, 2006.

[49] L. Yelash, M. Müller, P. Wolfgang, and K. Binder. How Well Can Coarse-Grained Models of Real Polymers Describe Their Structure? The Case of Polybutadiene. *J. Chem. Theor. Comp.*, **2**(3):588–597, 2006.

[50] A. Y. Shih, A. Arkhipov, P. L. Freddolino, and K. Schulten. Coarse grained protein-lipid model with application to lipoprotein particles. *J. Phys. Chem. B*, **110**(8):3674–3684, 2006.

[51] A. P. Lyubartsev. Multiscale modeling of lipids and lipid bilayers. *Eur. Biophys. J.*, **35**(1):53–61, 2005.

[52] J. Zhou, I. F. Thorpe, S. Izvekov, and G. A. Voth. Coarse-grained peptide modeling using a systematic multiscale approach. *Biophys. J.*, **92**(12):4289–4303, 2007.

[53] A. Villa, N. van der Vegt, and C. Peter. Self-assembling dipeptides: including solvent degrees of freedom in a coarse-grained model. *Phys. Chem. Chem. Phys.*, **11**(12):2068–2076, 2009.

[54] F. Ercolessi and J. B. Adams. Interatomic Potentials from 1st-Principles Calculations - the Force-Matching Method. *Europhys. Lett.*, **26**(8):583–588, 1994.

[55] B. Hess, C. Holm, and N. van der Vegt. Modeling Multibody Effects in Ionic Solutions with a Concentration Dependent Dielectric Permittivity. *Phys. Rev. Lett.*, **96**:147801, 2006.

[56] D. Reith, M. Pütz, and F. Müller-Plathe. Deriving effective mesoscale potentials from atomistic simulations. *J. Comp. Chem.*, **24**(13):1624–1636, 2003.

[57] A. Lyubartsev and A. Laaksonen. Calculation Of Effective Interaction Potentials From Radial-Distribution Functions - A Reverse Monte-Carlo Approach. *Phys. Rev. E*, **52**(4):3730–3737, 1995.

[58] A. K. Soper. Empirical potential Monte Carlo simulation of fluid structure. *Chem. Phys.*, **202**:295–306, 1996.

[59] S. Izvekov, M. Parrinello, C. J. Burnham, and G. A. Voth. Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: A new method for force-matching. *J. Chem. Phys.*, **120**(23):10896–10913, 2004.

[60] G. Toth. Effective potentials from complex simulations: a potential-matching algorithm and remarks on coarse-grained potentials. *J. Phys. Cond. Mat.*, **19**(33):335222, 2007.

[61] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *The Journal of Physical Chemistry B*, **111**(27):7812–7824, 2007.

[62] W. G. Noid, J. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen. The multiscale coarse graining method. 1. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.*, **128**:244114, 2008.

[63] D. Fritz, V. A. Harmandaris, K. Kremer, and N. F. A. van der Vegt. Coarse-Grained Polymer Melts Based on Isolated Atomistic Chains: Simulation of Polystyrene of Different Tacticities. *Macromolecules*, **42**(19):7579–7588, 2009.

[64] H. Wang, C. Junghans, and K. Kremer. Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining? *Eur. Phys. J. E*, **28**(2):221–229, 2009.

[65] V. A. Harmandaris, D. Reith, N. Van der Vegt, and K. Kremer. Comparison between coarse-graining models for polymer systems: Two mapping schemes for polystyrene. *Macromol. Chem. Phys.*, **208**(19-20):2109–2120, 2007.

[66] A. Villa, C. Peter, and N. van der Vegt. Self-assembling dipeptides: conformational sampling in solvent-free coarse-grained simulation. *Phys. Chem. Chem. Phys.*, **11**(12):2077–2086, 2009.

[67] R. L. Henderson. Uniqueness theorem for fluid pair correlation-functions. *Phys. Lett. A*, **49**(3):197–198, 1974.

[68] T. Murtola, A. Bunker, I. Vattulainen, M. Deserno, and M. Karttunen. Multiscale modeling of emergent materials: biological and soft matter. *Phys. Chem. Chem. Phys.*, **11**(12):1869–1892, 2009.

[69] J. T. Chayes and L. Chayes. On the validity of the inverse conjecture in classical density functional theory. *J. Stat. Phys.*, **36**:471–488, 1984.

[70] J. T. Chayes, L. Chayes, and E. H. Lieb. The inverse problem in classical statistical-mechanics. *Comm. Math. Phys.*, **93**(1):57–121, 1984.

[71] S. Leon, N. van der Vegt, L. Delle Site, and K. Kremer. Bisphenol A polycarbonate: Entanglement analysis from coarse-grained MD simulations. *Macromolecules*, **38**(19):8078–8092, 2005.

[72] C. Junghans, M. Praprotnik, and K. Kremer. Transport properties controlled by a thermostat: An extended dissipative particle dynamics thermostat. *Soft Matter*, **4**(1):156–161, 2008.

[73] W. G. Noid, J. W. Chu, G. S. Ayton, and G. A. Voth. Multiscale coarse-graining and structural correlations: Connections to liquid-state theory. *J. Phys. Chem. B*, **111**(16):4116–4127, 2007.

[74] C. DeBoor. *A Practical Guide to Splines: 027*. Springer, Berlin, illustrated edition edition, 1978.

[75] W. Tschop, K. Kremer, O. Hahn, J. Batoulis, and T. Burger. Simulation of polymer melts. II. From coarse-grained models back to atomistic description. *Acta Polymerica*, **49**:75–79, 1998.

[76] A. Troisi and G. Orlandi. Charge-transport regime of crystalline organic semiconductors: Diffusion limited by thermal off-diagonal electronic disorder. *Phys. Rev. Lett.*, **96**(8):086601, 2006.

[77] V. Rühle, K. J., and D. Andrienko. A multiscale description of charge transport in conjugated oligomers. *The Journal of Chemical Physics*, **132**(13):134103, 2010.

[78] Y. C. Cheng, R. J. Silbey, D. A. da Silva Filho, J. P. Calbert, J. Cornil, and J. L. Brédas. Three-dimensional band structure and bandlike mobility in oligoacene single crystals: A theoretical investigation. *The Journal of Chemical Physics*, **118**(8):3764–3774, 2003.

[79] A. Troisi. Prediction of the Absolute Charge Mobility of Molecular Semiconductors: the Case of Rubrene. *Advanced Materials*, **19**(15):2000–2004, 2007.

[80] A. Troisi, D. Cheung, and D. Andrienko. Charge Transport in Semiconductors with Multiscale Conformational Dynamics. *Phys. Rev. Lett.*, **102**(11):116602, 2009.

[81] J. Kirkpatrick, V. Marcon, J. Nelson, K. Kremer, and D. Andrienko. Charge mobility of discotic mesophases: a multiscale quantum/classical study. *Phys. Rev. Lett.*, **98**:227402, 2007.

[82] D. Andrienko, J. Kirkpatrick, V. Marcon, J. Nelson, and K. Kremer. Structure-Charge Mobility Relation for Hexabenzocoronene Derivatives. *Phys. Stat. Sol. B*, **245**:830, 2008.

[83] V. Marcon, J. Kirkpatrick, W. Pisula, and D. Andrienko. Supramolecular Structure of Perylene Tetracarboxdiimides. *Phys. Stat. Sol. B*, **245**:820, 2008.

[84] V. Marcon, T. Vehoff, J. Kirkpatrick, C. Jeong, D. Y. Yoon, K. Kremer, and D. Andrienko. Columnar mesophases of hexabenzocoronene derivatives. I. Phase transitions. *J. Chem. Phys.*, **129**:094505, 2008.

[85] J. Kirkpatrick, V. Marcon, K. Kremer, J. Nelson, and D. Andrienko. Columnar mesophases of hexabenzocoronene derivatives. II. Charge carrier mobility. *J. Chem. Phys.*, **129**(9):094506, 2008.

[86] V. Marcon, D. Breiby, W. Pisula, J. Dahl, J. Kirkpatrick, S. Patwardhan, F. Grozema, and D. Andrienko. Understanding Structure-Mobility Relations for Perylene Tetracarboxydiimide Derivatives. *J. Am. Chem. Soc.*, **131**(32):11426–11432, 2009.

[87] R. A. Marcus. Electron transfer reactions in chemistry. Theory and experiment. *Rev. Mod. Phys.*, **65**(3):599–610, 1993.

[88] V. May and O. Kuhn. *Charge and Energy Transfer Dynamics in Molecular Systems. A Theoretical Introduction*. Wiley-Vch, 2000.

[89] M. D. Newton. Quantum chemical probes of electron-transfer kinetics: the nature of donor-acceptor interactions. *Chemical Reviews*, **91**(5):767–792, 1991.

[90] K. F. Freed and J. Jortner. Multiphonon Processes in the Nonradiative Decay of Large Molecules. *J. Chem. Phys.*, **52**(12):6272–6291, 1970.

[91] E. F. Valeev, V. Coropceanu, D. A. da Silva Filho, S. Salman, and J. Bredas. Effect of Electronic Polarization on Charge-Transport Parameters in Molecular Organic Semiconductors. *Journal of the American Chemical Society*, **128**(30):9882–9886, 2006.

[92] J. Kirkpatrick. An approximate method for calculating transfer integrals based on the ZINDO Hamiltonian. *Int. J. Quant. Chem.*, **108**:51–56, 2007.

[93] J. Ridley and M. Zerner. An intermediate neglect of differential overlap technique for spectroscopy: Pyrrole and the azines. *Theoretica Chimica Acta*, **32**(2):111–134, 1973.

[94] P. Barbara, T. Meyer, and M. Ratner. Contemporary issues in electron transfer research. *J. Phys. Chem.*, **100**(31):13148–13168, 1996.

[95] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, **21**(6):1087–1092, 1953.

[96] J. R. Beeler. Displacement Spikes in Cubic Metals. I. α-Iron, Copper, and Tungsten. *Phys. Rev.*, **150**(2):470–487, 1966.

[97] A. Voter. Introduction to the Kinetic Monte Carlo method. In *Radiation Effects in Solids*, pages 23, 1. 2007.

[98] H. Bässler. Charge Transport in Disordered Organic Photoconductors - a Monte Carlo Simulation Study. *Physica Status Solidi B*, **175**(1):15–56, 1993.

[99] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz. A new algorithm for Monte Carlo simulation of Ising spin systems. *Journal of Computational Physics*, **17**(1):10–18, 1975.

[100] P. M. Borsenberger, L. Pautmeier, and H. Bassler. Charge transport in disordered solids. *J. Chem. Phys.*, **94**:5447–5454, 1991.

[101] R. Coehoorn, W. F. Pasveer, P. A. Bobbert, and M. A. J. Michels. Charge-carrier concentration dependence of the hopping mobility in organic materials with Gaussian disorder. *Phys. Rev. B*, **72**(15):155206, 2005.

[102] J. A. Freire. Field dependence of the mobility in organic insulators with a generic Gaussian correlation between the site energies. *Phys. Rev. B*, **72**(12):125112, 2005.

[103] V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko. Versatile Object-oriented Toolkit for Coarse-graining Applications. *J. Chem. Theor. Comp.*, **5**:3211–3223, 2009.

[104] E. Gamma, R. Helm, and R. E. Johnson. *Design Patterns. Elements of Reusable Object-Oriented Software.* Addison-Wesley Longman, Amsterdam, 1st ed., reprint. edition, 1995.

[105] A. Alexandrescu. *Modern C++ Design, Generic Programming and Design Patterns Applied*. Addison-Wesley Longman, Amsterdam, 2001.

[106] R. Faller, H. Schmitz, O. Biermann, and F. Müller-Plathe. Automatic parameterization of force fields for liquids by simplex optimization. *Journal of Computational Chemistry*, **20**(10):1009–1017, 1999.

[107] I. Nezbeda and J. Slovak. A family of primitive models of water: Three-, four and five-site models. *Mol. Phys.*, **90**(3):353–372, 1997.

[108] A. Wallqvist and R. D. Mountain. Molecular Models of Water: Derivation and Description. *Rev. Comp. Chem.*, **13**:183–247, 2007.

[109] P. G. Kusalik and I. M. Svishchev. The spatial structure in liquid water. *Science*, **65**:1219–1221, 1994.

[110] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, **91**:6269–6271, 1987.

[111] H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, and J. R. Haak. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**:3684–3690, 1984.

[112] W. F. Van Gunsteren and H. J. C. Berendsen. A Leap-frog Algorithm for Stochastic Dynamics. *Mol. Sim.*, **1**(3):173, 1988.

[113] M. E. Johnson, T. Head-Gordon, and A. A. Louis. Representability problems for coarse-grained water potentials. *J. Chem. Phys.*, **126**(14):144509, 2007.

[114] V. Molinero and E. B. Moore. Water Modeled As an Intermediate Element between Carbon and Silicon†. *The Journal of Physical Chemistry B*, **113**(13):4008–4016, 2009.

[115] W. G. Noid, P. Liu, Y. Wang, J. Chu, G. Ayton, S. Izvekov, H. C. Andersen, and G. Voth. The multiscale coarse-graining method. 2. Numerical implementation for coarse-grained molecular models. *J. Chem. Phys.*, **128**:244115, 2008.

[116] W. Janke. *Statistical Analysis of Simulations: Data Correlations and Error Estimation*, volume 10 of *NIC Series, From Theory to Algorithms, Lecture Notes*. 2002.

[117] W. L. Jorgensen and J. Tirado-Rives. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci.*, **102**(19):6665–6670, 2005.

[118] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, **52**:255–268, 1984.

[119] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, **31**(3):1695, 1985.

[120] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, **126**:014101, 2007.

[121] C. Peter and K. Kremer. Multiscale simulation of soft matter systems?from the atomistic to the coarse grained level and back. *Soft Matter*, , 2009. accepted.

[122] B. Bolto and D. Weiss. Electronic Conduction In Polymers .2. Electrochemical Reduction Of Polypyrrole At Controlled Potential. *Aust. J. Chem.*, **16**(6):1076, 1963.

[123] A. Diaz, J. Castillo, J. Logan, and W. Lee. Electrochemistry Of Conducting Polypyrrole Films. *J. Elec. Chem.*, **129**:115–132, 1981.

[124] J. L. Bredas, B. Themans, and J. M. Andre. Bipolarons in polypyrrole chains. *Phys. Rev. B*, **27**(12):7827–7830, 1983.

[125] J. L. Bredas, B. Themans, J. G. Fripiat, J. M. Andre, and R. R. Chance. Highly conducting polyparaphenylene, polypyrrole, and polythiophene chains: An ab initio study of the geometry and electronic-structure modifications upon doping. *Phys. Rev. B*, **29**(12):6761–6773, 1984.

[126] R. Colle and A. Curioni. Density-Functional Theory Study of Electronic and Structural Properties of Doped Polypyrroles. *J. Am. Chem. Soc.*, **120**(19):4832–4839, 1998.

[127] R. Colle and A. Curioni. Density-functional theory and Car-Parinello study of electronic, structural, and dynamical properties of the hexapyrrole molecule. *J. Phys. Chem. A*, **104**(37):8546–8550, 2000.

[128] G. R. Hutchison, M. A. Ratner, and T. J. Marks. Hopping Transport in Conductive Heterocyclic Oligomers: Reorganization Energies and Substituent Effects. *Journal of the American Chemical Society*, **127**(7):2339–2350, 2005.

[129] M. Yurtsever and E. Yurtsever. A theoretical study of structural defects in conjugated polymers. *Synthetic metals*, **101**:335–336, 1999.

[130] E. Yurtsever, O. Esentürk, H. Ö. Pamuk, and M. Yurtsever. Structural studies of polypyrroles - I. An ab-initio evaluation of bonding through alpha and beta carbons. *Synthetic metals*, **98**(3):221–227, 1999.

[131] N. McDonald and W. Jorgensen. Development of an all-atom force field for heterocycles. Properties of liquid pyrrole, furan, diazoles, and oxazoles. *J. Phys. Chem. B*, **102**(41):8049–8059, 1998.

[132] J. J. Lopez Cascales and T. F. Otero. Molecular dynamic simulation of the hydration and diffusion of chloride ions from bulk water to polypyrrole matrix. *J. Chem. Phys.*, **120**(4):1951–1957, 2004.

[133] J. J. Lopez Cascales and T. F. Otero. Molecular Dynamics Simulations of the Orientation and Reorientational Dynamics of Water and Polypyrrole Rings as a Function of the Oxidation State of the Polymer. *Macromolecular Theory and Simulations*, **14**(1):40–48, 2005.

[134] J. Sancho-Garci and A. Karpfen. The torsional potential in 2,2 '-bipyrrole revisited: High-level ab initio and DFT results. *Chemical Physics Letters*, **411**:321–326, 2005.

[135] A. Lukyanov, C. Lennartz, and D. Andrienko. Amorphous films of tris-8(hydroxyquinoline aluminium): force-field, morphology, and charge transport. *Phys. Stat. Sol. A*, **206**:2737–2742, 2009.

[136] B. Mannfors, J. Koskinen, L.-O. Pietilä, and L. Ahjopalo. Density functional studies of conformational properties of conjugated systems containing heteroatoms. *Journal of molecular structure. Theochem*, **393**:39–58, 1997.

[137] C. B. Breneman and K. E. Wiberg. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformationanal analysis. *J. Comp. Chem.*, **11**(3):361–373, 1990.

[138] W. Goedel, G. Holz, G. Wegner, J. Rosenmund, and G. Zotti. Electrochemical Investigations Of A Substituted Oxidation Stable Polypyrrole. *Polymer*, **34**(20):4341–4346, 1993.

[139] J. Arvo. Fast Random Rotation Matrices. In *Graphics Gems III*. Elsevier LTD, 1992.

[140] T. Holstein. Studies of polaron motion .2. the small polaron. *Ann. Phys.*, **8**(3):343–389, 1959.

[141] A. Romero, D. Brown, and K. Lindenberg. Electron-phonon correlations, polaron size, and the nature of the self-trapping transition. *Phys. Lett. A*, **266**:414–420, 2000.

[142] L. Zoppi, A. Calzolari, A. Ruini, A. Ferretti, and M. Caldas. Defect-induced effects on carrier migration through one-dimensional poly(para-phenylenevinylene) chains. *Phys. Rev. B*, **78**(16):165204, 2008.

[143] D. Cheung, D. McMahon, and A. Troisi. Computational Study of the Structure and Charge-Transfer Parameters in Low-Molecular-Mass P3HT. *J. Phys. Chem. B*, **113**(28):9393–9401, 2009.

[144] R. Singh and A. Narule. Correlation between dielectric relaxation and d.c. electricla conduction in polypyrrole family of polymers. *Synthetic Metals*, **82**:245–149, 1996.

[145] J. L. Bredas, J. C. Scott, K. Yakushi, and G. B. Street. Polarons and bipolarons in polypyrrole: Evolution of the band structure and optical spectrum upon doping. *Phys. Rev. B*, **30**(2):1023–1025, 1984.

[146] R. R. Chance, J. L. Bredas, and R. Silbey. Bipolaron transport in doped conjugated polymers. *Phys. Rev. B*, **29**(8):4491–4495, 1984.

[147] J. Nelson and R. Chandler. Random walk models of charge transfer and transport in dye sensitized systems. *Coord. Chem. Rev.*, **248**:1181–1194, 2004.

[148] L. Wang. Charge-Density Patching Method for Unconventional Semiconductor Binary Systems. *Physical Review Letters*, **88**(25):256402, 2002.

[149] M. Fowler. *UML Distilled: A Brief Guide to the Standard Object Modeling Languange*. Addison-Wesley Longman, Amsterdam, 3. a. edition, 2003.

Acknowledgments

Victor Rühle
Ackermannweg 10
55128 Mainz

+49 (0)6131-379148
ruehle@mpip-mainz.mpg.de

Resume