Chapter 1

Introduction into Molecular Simulation

The computer simulation of molecular systems is a research area with a fine tradition. Nowadays, supercomputers like ANTON [28] are specifically designed to generate classical molecular dynamics trajectories. Researchers are interested in classical molecular simulations in order to understand protein folding processes and interactions between molecules like ligands and proteins.

In classical molecular dynamics, the equations of motion of an $N$-atoms molecular system are solved in configurational space ($3N$ cartesian coordinates) and in momenta space ($3N$ momentum variables). One problem with the simulation of molecular systems is the gap between the fastest oscillations of covalent bonds of a molecule (some femtoseconds) and the time-scale of the interesting processes (some microseconds for fast folding processes) [12]. Another problem is the evaluation of forces acting on the atoms of the molecular system. This evaluation is very expensive. ANTON can generate a trajectory of 10,000 ns per day for a large protein structure (23,000 atoms plus explicit water) [34]. Thus, in order to simulate one second in “real life” ANTON needs about 270 years of computing time. Even if we can generate and analyze a trajectory of this length, it is not clear that the generated statistical data contains enough information about the molecular processes and time-scales under consideration. In Figure 1.1, e.g., a typical time-series plot of an observable taken from a dynamics simulation of a small molecule (head group of a cholestane spin-probe) is shown.

Simply speaking, this molecule has two different conformations. The relative position of the oxygen atom, as measured by means of an internal coordinate, determines its conformation. Here, the internal coordinate is a dihedral angle. In the time-series plot, one can see that the molecule jumps between two conformations. The conformations are indicated by blue and red coloured stripes. Deuflhard, Schütte et al. [9, 7, 10, 27, 8] introduced this set-based approach to conformation dynamics, which was inspired by ideas of Michael Dellnitz et al. [6] for hyperbolic dynamical systems. The observation in Figure 1.1 is that the molecular system stays in a conformation for a very long time compared to the

\[^{1}\text{In fact, ANTON is the first computer which can reach this efficiency. At the time ANTON was designed, other comparable parallel computers could only simulate about 100 ns per day.}\]
Figure 1.1: A typical time-series plot of an observable from a molecular dynamics simulation of a small molecule.

the time-scale of simulation before it jumps to another conformation. Consequently, Deuflhard et al. used this kind of observation in order to define conformations as metastable sets in the configurational space of the molecular system. Their approach is the theoretical framework in the following. The set-based concept can be extended to more than two conformations. It can also be extended to the situation of Figure 1.2. The binding process of a ligand into an active site can be seen as a transition process between different conformations along the binding path. An observable can, e.g., characterize the relative position of the ligand to the active site. In this case, a similar behavior compared to that in Figure 1.1 can be expected for the observable in MD simulations. Due to rare events during this simulation, the complete binding path in Fig. 1.2 cannot be simulated with standard MD methods. These are the interesting questions to be solved for the protein-ligand binding example:

Q1 What are the statistical weights of the conformations? In other words, what is the probability for the system to be in one of the conformations?

Q2 How can we describe the transition pattern between the conformations? In other words, how can we estimate the probability to stay in one of the conformations, as well as the probability to move between certain conformations?

Based on the conformation dynamics approach, the questions Q1 and Q2 can be answered by a Markov State Model. The transition pattern between the conformations is given by a row-stochastic transition matrix and the statistical weights are given by the dominant left eigenvector of this transition matrix [8]. This is a statement about the mathematical formulation of the answers to Q1 and Q2. It is not a statement about how to get the statistical information to build up the Markov State Model. Figure 1.1 clearly shows that a direct sampling approach leads to redundant statistical data. E.g., follow the trajectory in Figure 1.1 when the system is in the red conformation for the first time. During the simulation a lot of redundant data about the local distribution of the dihedral
Figure 1.2: A binding path of a small ligand molecule from the surface of a protein into the active site. This is a result of a molecular kinetics simulation and cannot be computed in reasonable time using standard MD methods [3].

angle inside this “red conformation” (A) is collected until the trajectory jumps to the “blue conformation” (B) again. Although the overall generated simulation data is by far sufficient to estimate the local distributions of the observable inside the two conformations, the sampling does not contain enough data in order to make a statistically relevant statement about the transition pattern of the system. There are too few jumps between the conformations A and B inside the trajectory. These jumps are rare events. Thus, the transition pattern (Q2) cannot be extracted from the time-series in Figure 1.1 within reasonable CPU-time. Additionally, even the statistical weights (Q1) of the two conformations cannot be extracted from the presented time-series. If the trajectory rarely jumps between its conformations, the weighting between the conformations may be wrong [33]. In order to answer the aforementioned questions of conformation dynamics, the applied simulation method has to balance between two different requirements.

R1 In order to figure out the statistical weights of the conformations, the sampling method cannot be based on rare jumps between the conformations.

R2 For an analysis of the transition pattern, the sampling method has to focus on the transition regions. In our example: The region between the two stripes in Figure 1.1.

These two requirements are contradictory, because for R1 the sampled trajectory avoids spending time in transition regions whereas for R2 it focuses on these regions. In the opinion of the author, these two requirements lead to different sampling approaches. The second requirement R2 can be realized by a deeper analysis of transition regions. This requirement is the most difficult part of conformation dynamics.
Conceptual change. At this stage, everything seems to be solvable using the given Markov State Model framework of conformation dynamics. Answering the two questions (Q1 and Q2) seems to be a purely technical problem: The requirements (R1 and R2) simply lead to different sampling schemes. However, conformation dynamics has a conceptual problem. In Figure 1.3, a sketch of the set-based concept of conformation dynamics is shown. The state space is decomposed into metastable sets (in our example: A and B). Each state of a time-discrete trajectory (circles in the first row) can be assigned to one of these sets, i.e., to one of the conformations. Note that in this concept, the states in the transition region in Figure 1.1 have to be uniquely assigned to A or B as well. Whereas, the first row in Figure 1.3 is a Markov chain in state space, the projected time-series (AABBA...) does not possess a Markov property [31, 32]. Thus, presenting a Markov State Model as a solution of the conformation dynamics problem conceals that the projection from a dynamics simulation in a continuous state space onto a finite number of sets spoils the Markov property. There are several approaches in literature, which try to correct the results of conformation dynamics in this direction. A rigorous approach is given by the computation of committor functions. Unfortunately, this approach is only valid for the case of two conformations. Therefore, the present text focuses in analyzing the projection in Figure 1.3 from a different point of view. Instead of analyzing molecular dynamics trajectories a transfer operator concept will be used as, in principle, introduced by Schütte [27]. However, a transfer operator different from Schütte's operator will be defined. Desirable properties of such a new transfer operator will be derived. In order to deduce a valid projection of this new transfer operator to a low dimensional Markov State Model, conformations are not defined by sets any more. Rather, conformations are given by membership functions computed as a linear combination of dominant eigenfunctions of the transfer operator. This approach will include the key for an efficient sampling of the state space. A direct estimation of the infinitesimal generator of the new transfer operator will be used to adaptively and hierarchically sample from the state space, i.e., generate sampling data wherever “more information” is needed.

Figure 1.3: The states of a trajectory of a molecular dynamics simulation are assigned to two conformations A and B.
Chapter 2

Molecular kinetics in the canonical ensemble

All considerations in the followings assume a canonical ensemble, i.e., the simulation results are valid for systems with a constant number of particles $N$, constant volume $V$ and constant temperature $T$. The theoretical construction of a canonical ensemble as well as the state of the art methods to characterize its dynamical behavior are derived in the upcoming sections. This is necessary because the basic assumptions should be clear before the mentioned conceptual change is introduced.

2.1 From molecular dynamics to Markov State Models

In Figure 2.1, an isolated system is indicated by a box with a thick-walled barrier. The system cannot interact with its surroundings, the transfer of matter and of heat is blocked. Imagine a thermos flask. This isolated system cannot be modeled with computational methods because it consists of too many particles. A certain homogeneity is assumed: The system is divided into a large number of identical, closed subsystems. A closed system can exchange heat with the surroundings, but it cannot transfer matter. Each subsystem is a copy of a certain molecular system. E.g., imagine a certain protein with its ligand in a water box as a subsystem in Figure 2.1. In the classical framework, each subsystem has the same number $N$, same types, and same connectivities of atoms as well as the same volume. However, they are different with regard to their molecular state. Each subsystem has its own configurational state $q \in \Omega \subset \mathbb{R}^{3N}$ and momentum state $p \in \mathbb{R}^{3N}$. Thus, the total energy $H$ of each subsystem is different. The total energy $H(q, p)$ of such a classical molecular system is the sum $H(q, p) = V(q) + K(p)$ of its potential energy, $V : \Omega \to \mathbb{R}$, only depending on $q$, and its kinetic energy, $K : \mathbb{R}^{3N} \to \mathbb{R}$, only depending on $p$. Taking all these conditions into account, Boltzmann derived the probability density function $\pi : \Omega \times \mathbb{R}^{3N} \to \mathbb{R}_+$ of states $(q, p)$ of the subsystems as

$$\pi(q, p) = \frac{1}{Z} \exp(-\beta H(q, p)),$$  \hspace{1cm} (2.1)
where $Z > 0$ is the normalization constant (also called partition function), and $\beta$ is derived from the Lagrange multiplier of the above constraints. In order to compute the Boltzmann distribution as a solution of an optimization problem, one can either ask for the most probable distribution or maximize the entropy [21] of the isolated system. The factor $\beta$ can be related to the inverse temperature $T$ of the system $\beta = (k_B T)^{-1}$, where $k_B$ is the Boltzmann constant. Thus, besides the number of particles and the volume of the subsystems, temperature is a further common property (see Figure 2.1). The states of the subsystems are time-dependent. The Boltzmann distribution is a dynamical equilibrium of the system. There are two different ways to characterize this dynamical process: molecular dynamics and molecular kinetics.

**Molecular dynamics.** A molecular dynamics simulation of the system in Figure 2.1 picks out only one subsystem of the canonical ensemble and determines its time-dependent evolution. This is done independently from the states of the other subsystems. In the context of the canonical ensemble, molecular dynamics can be seen as a simulation of a closed molecular (sub)system. For this simulation a dynamical model has to be defined. This model has to qualify and quantify the interchange of energy between the subsystems of the canonical ensemble. One important dynamical model is Hamiltonian dynamics. In this dynamical model, the time-evolution of the states is given by a first order differential equation

$$
\dot{q}(t) = \nabla_p K(p(t)),
\dot{p}(t) = -\nabla_q V(q(t)),
$$

The Boltzmann distribution is the equilibrium distribution of the states of the subsystems in Figure 2.1. Entropy increases as long as the system is not in an equilibrium state. In an equilibrium state, entropy reaches its maximum.

Hamiltonian dynamics is a dynamical model of an isolated subsystem, but an isolated subsystem is just a special case of a closed subsystem without interchange of energy.
where $\dot{q}$ and $\dot{p}$ are the time-derivatives of $q(t)$ and $p(t)$. A short calculation shows that $\frac{d}{dt} \pi(q(t), p(t)) = 0$ in the case of Hamiltonian dynamics (2.2). Furthermore, the phase space volume is time-invariant with regard to the symplectic dynamics. Therefore, the density $\pi$ of states is preserved assuming Hamiltonian dynamics. Hamiltonian dynamics is a valid dynamical model with regard to the canonical ensemble (it preserves the equilibrium state of the system dynamically). This insight contradicts a common opinion. In fact, Hamiltonian dynamics is a valid dynamical model for a simulation at constant temperature. The important insight is that temperature is not a property of a molecular state, it is a property of an ensemble. However, Hamiltonian dynamics is not recommended for constant-temperature simulations, because it is not ergodic in this context. Given an initial state $(q(0), p(0))$, the Hamiltonian dynamics trajectory does not come arbitrarily close to all the states of the subsystems. It keeps its initial energy level, whereas the subsystems have different total energy levels in Figure 2.1. As a consequence, self-equilibration of the isolated system, i.e., convergence against the Boltzmann distribution of states, cannot be based on Hamiltonian dynamics as a dynamical model for the canonical ensemble. In chapter 2.3, modifications of Hamiltonian dynamics are presented, which are often used for an ergodic molecular dynamics simulation at constant temperature.

**Molecular kinetics.** In molecular dynamics a single trajectory is analyzed. A conformational transition takes place along this trajectory in the moment when the trajectory leaves a certain subset of the state space and enters a different one. In molecular kinetics, however, an ensemble of trajectories is analyzed, i.e. the propagation of probability densities is observed. From this point of view, a “transition” takes place if an arbitrary of the observed trajectories leaves a certain subset of the state space and enters a different one. In particular, the transition pattern of molecular kinetics does not hold for the behavior of single subsystems. Statements like “transition rates between the conformations correspond to a long-term dynamics trajectory” are not possible with regard to this approach, because such a statement mixes two different points of view. It is not possible to claim that the molecular kinetics transition pattern represents a single realization of a dynamical model. The computed transition rates of molecular kinetics can also not be validated by experiments observing the dynamics of a single molecule. Although, there are two main assumptions in the molecular dynamics approach which are relevant for molecular kinetics, too. The first assumption is a kind of a Markov property: Equation (2.2) is a first order deterministic differential equation. In order to predict the future evolution of the system it is sufficient to know the current state $(q(t), p(t))$. Note that for all dynamical models in chapter 2.3, i.e. also for first order stochastic differential equations and for time-continuous time-harmonic Markov processes, this Markov property holds. The second assumption of molecular dynamics, which also holds for molecular kinetics, is given by the independence of the energy transfer of the subsystems with regard to the rest of the ensemble. In other words, the subsystems do not “see” if the ensemble is equilibrated or not. An important consequence: Molecular kinetics is always (at each point of time) driven by the same dynamical model, no matter whether the ensemble is on its

---

3An example is given by the conformational transitions in Figure 1.1.
way to equilibrium or has already reached its dynamical equilibrium.

In molecular kinetics, the evolution of the probability density of the states of the subsystems is determined. In this context, the basic assumptions can be expressed mathematically by an operator equation. Given a time-dependent probability density function \( \rho : \mathbb{R} \times \Omega \times \mathbb{R}^{3N} \to \mathbb{R} \) of states \( (q, p) \in \Omega \times \mathbb{R}^{3N} \) at time \( t \) and a lag-time \( \tau > 0 \), the evolution of the probability density function at time \( t + \tau \) can be expressed as:

\[
\rho(t + \tau, \cdot, \cdot) = \mathcal{P}_s(\tau) \rho(t, \cdot, \cdot),
\]

where \( \mathcal{P}_s(\tau) : L^1(\Omega \times \mathbb{R}^{3N}) \to L^1(\Omega \times \mathbb{R}^{3N}) \) is a lag-time-dependent (but not time-dependent) operator which propagates probability density functions. Equation (2.3) is too complex to be solved for high-dimensional molecular systems. The complexity of this equation is reduced by using the aforementioned conformation dynamics approach. The probability density function \( \rho(t, \cdot, \cdot) \) is projected to a time-dependent low-dimensional vector \( w(t) \in \mathbb{R}^n \). The elements of this vector are given by the statistical weights of the \( n \) conformations at time \( t \). In Figure 2.2, a sketch of the complexity reduction can be seen. Whereas the projection shown in Figure 1.3 is based on a molecular dynamics simulation, Figure 2.2 presents the conformation dynamics approach in the desired molecular kinetics framework. The propagation of the vectors \( w \) in Figure 2.2 is done by an \( n \times n \) matrix \( \mathcal{P}_c^\top(\tau) \) via \( w(t + \tau) = \mathcal{P}_c^\top(\tau) w(t) \). In order to get a commuting diagram in Figure 2.2, the matrix \( \mathcal{P}_c^\top(\tau) \) has to preserve the non-negativity of \( w \) (at least). Furthermore, the sum of the elements of \( w \) has to be 1 for every time-step in Figure 2.2. Column-stochastic matrices having positive eigenvalues present one possible(!) class of matrices in this context. Column-stochastic matrices have non-negative elements and their column sums are equal to 1. In this special case, the transposed \( \mathcal{P}_c(\tau) \) can be interpreted as a transition matrix of a Markov chain. This is the reason why this matrix is called Markov State Model in conformation dynamics. Each conformation is denoted as one possible Markov state of the system. The transition behavior is given by \( \mathcal{P}_c(\tau) \). The aim of conformation dynamics is to compute this Markov State Model.

---

4Note that for a non-negative probability density function \( \rho \) the normalization \( 1 = \int_{\mathbb{R}^{3N}} \int_\Omega \rho(t, q, p) dq dp \) holds, i.e., \( \rho(t, \cdot, \cdot) \in L^1(\Omega \times \mathbb{R}^{3N}) \). The connection of (2.3) to ordinary reaction kinetics is visible if \( \rho(t, q, p) \) is interpreted as “concentration” of “species” \( (q, p) \) at time \( t \).

5If \( \mathcal{P}_c^\top(\tau) \) is column-stochastic, then \( \mathcal{P}_c(\tau) \) is a row-stochastic matrix. However, \( \mathcal{P}_c^\top(\tau) \) is not necessarily column-stochastic. Nevertheless, the technical term “Markov State Model”
2.2 Transfer operator approach

The set-based concept of the transfer operator approach [27] of Schütte et al. is a method to find a Markov State Model $P_c(\tau)$. This approach is described in this section. As mentioned before, the investigations can be restricted to the case of an equilibrated system. In the classical framework, the total energy $H$ is the sum of kinetic and potential energy. The Boltzmann distribution $\pi$ can be decomposed, $\pi = \pi_p \pi_q$, into a probability density function $\pi_p : R^{3N} \rightarrow R$ for the kinetic energy in momentum space and a probability density function $\pi_q : \Omega \rightarrow R$ for the potential energy in configurational space.

The common algorithmic realization of the transfer operator approach is as follows. In order to generate a Markov State Model $P_c$, the molecular kinetics information is taken from a molecular dynamics simulation with a pre-defined dynamical model. Hamiltonian dynamics has been identified above as a valid dynamical model for the canonical ensemble. In his thesis, Schütte defined a transfer operator $T(\tau) : L^1_{\pi_q}(\Omega) \rightarrow L^1_{\pi_q}(\Omega)$ on the basis of Hamiltonian dynamics as

$$T(\tau) f(q) = \int_{R^{3N}} f(\Pi_q \Psi^{-\tau}(q, p)) \pi_p(p) \, dp.$$  \hspace{1cm} (2.4)

Equation (2.4) can be understood as follows: Given an initial state $(q, p)$, a backward Hamiltonian dynamics for a lag-time $\tau$ is investigated. The final state is denoted as $\Psi^{-\tau}(q, p)$. Via $\Pi_q$, this final state is projected to position space. The integral in (2.4) averages over all possible initial momentum variables with given Boltzmann distribution $\pi_p$. The definition of conformations as metastable sets leads to a decomposition of the configurational space $\Omega$. In the set-based approach, the conformations $\chi_1, \ldots, \chi_n$ are given by the characteristic functions of the corresponding subsets of $\Omega$, i.e., in terms of functions $\chi_i : \Omega \rightarrow \{0, 1\}$. The conformations $\chi_i$ form a partition of unity via $\sum_{i=1}^n \chi_i(q) = 1$ for all $q \in \Omega$.

In order to identify the conformations $\chi$ in practice, the configurational space is decomposed into a larger (but finite) number of subsets represented by characteristic functions $\Phi_i : \Omega \rightarrow \{0, 1\}$, $i = 1, \ldots, m$, with $m \gg n$. A transition probability matrix $P(\tau) \in R^{m \times m}$ between these small subsets is used to identify the metastable subsets $\chi_i$ of the configurational space. The discretization scheme of the transfer operator approach to conformation dynamics can be written as

$$T \rightarrow P \rightarrow P_c.$$  \hspace{1cm} (2.5)

A continuous transfer operator $T(\tau)$ is defined to characterize the collective transfer of initial states for a certain dynamical model of the molecular system. A discretization of this operator via basis functions $\Phi$ leads to a transition matrix $P(\tau)$ which is used to identify the conformations $\chi$ and the Markov State Model $P_c(\tau)$. On the basis of the transfer operator $T(\tau)$, the element $(i, j)$ of the transition matrix $P(\tau)$ can be computed as

$$P(\tau)(i, j) = \frac{\langle \Phi_i, T(\tau) \Phi_j \rangle_{\pi_q}}{\langle \Phi_i, e \rangle_{\pi_q}},$$  \hspace{1cm} (2.6)

where $\langle f, g \rangle_{\pi_q}$ is the $\pi_q$-scalar product defined as $\int_{\Omega} f(q) g(q) \pi_q(q) \, dq$. The function $e : \Omega \rightarrow \{1\}$ is a constant. The nominator in (2.6) counts the number will be used for $P_c(\tau)$ throughout this text.
of states which undergo a transition from set $\Phi_i$ to set $\Phi_j$ in time $\tau$. This number is divided by the statistical weight of set $\Phi_i$ in equilibrium, given by $d_i := \langle \Phi_i, e \rangle_{\pi_q}$. Thus, the expression (2.6) denotes the conditional probability for a transition from set $\Phi_i$ to set $\Phi_j$. The matrix $P_c$ can be derived from the operator $\mathcal{T}$ in a similar way. Just exchange the $\Phi$-basis functions by $\chi$-basis functions in (2.6). $T(\tau)$ in (2.4) is a transfer operator. $T(\tau)$ is not acting on density functions, it is acting on membership functions. Thus, stationarity is characterized by the equation $e = T(\tau)e$, where $e$ is the constant function $e \equiv 1$ in $\Omega$. In contrast to $\mathcal{T}$, the operator $\mathcal{P}_s(\tau)$ in (2.3) is acting on density functions. In this case, stationarity is characterized by $\pi = \mathcal{P}_s(\tau)\pi$, with the Boltzmann density $\pi$. Consequently, $\mathcal{T}$ is discretized with a set of basis functions $\Phi$ or $\chi$, which is a partition of unity (and not a partition of $\pi_q$). $\mathcal{T}$ and the projection of $\mathcal{P}_s$ to configurational space are adjoint operators, which can be seen by the fact that a $\chi$-discretization of $\mathcal{T}$ leads to $P_c$ and a $\chi$-discretization of $\mathcal{P}_s$ leads to $P_c^\top$ in chapter 2.1. The transfer operator approach is a powerful concept for molecular kinetics investigations. For the computation of $P$ and $P_c$, however, high dimensional integrals have to be solved, see (2.6). In this equation, the computation of the term $\mathcal{T}(\tau)\Phi_j(\tau)$ can be based on short-time $\tau$ molecular dynamics simulation data. As mentioned before, Hamiltonian dynamics is not ergodic. Therefore, we will show how to compute the transition matrix $P$ on the basis of molecular dynamics simulation data for different dynamical models. This requires a generalization of $\mathcal{T}$.

### 2.3 Thermostated molecular dynamics simulations

Many researchers have created possible dynamical models for a canonical ensemble, such that the distribution of simulation data of a single long-term trajectory converges to (2.1). They have been inspired by the equations of motion (2.2). There are two main approaches used in practice.

1. A deterministic approach: Instead of (2.2) an alternative but similar deterministic dynamical system is defined which converges against Boltzmann distribution. A well-known example is the time-reversible Nosé-Hoover dynamics \cite{13, 18}. Another example is the Berendsen thermostat \cite{2} which does not generate the canonical ensemble exactly. Other time-reversible deterministic thermostats can be found in \cite{19}. It should be mentioned, that the term “deterministic approach” is only of academic interest. From a numerical point of view, the Ljapunov exponent of the dynamical systems is usually very high: Long-term deterministic dynamical systems, say over more than about 100 femtoseconds, are numerically chaotic. This is the reason why many researchers prefer molecular dynamics simulations for generating Boltzmann distributed ensembles.

2. A stochastic approach: Besides Smoluchowski \cite{20} and Langevin dynamics \cite{26}, the class of hybrid Monte-Carlo methods (HMC) \cite{11} is an example for a stochastic dynamical model. In HMC, the system is mainly propagated according to (2.2). Sole exception: After a certain time-span the data in Figure 1.1 have been generated with HMC.
momentum coordinates are refreshed randomly and a Metropolis-like acceptance step assures the convergence of the system towards (2.1). Since a total refreshing of momentum variables seems to be unphysical, there are alternative variants of this method. In these variants, momentum variables are more or less conserved, e.g., like in targeted shadow HMC [1].

Besides possible physical inconsistencies of the above dynamical models, there is always an unknown additional parameter which defines how fast the trajectories can change between the energy levels of $H$. From a physical point of view, this parameter determines the quality of the energy transfer between the molecular system and its environment in order to equilibrate temperature. This parameter is difficult to define and often appears arbitrarily. However, the transfer operator approach can be extended to these different dynamical models. The two classes of dynamical models have an important property in common – the Markov property. Given a starting point $(q, p) \in \Omega \times \mathbb{R}^{3N}$, one can determine a probability for the possible future evolution of the system. These probabilities only depend on the starting point $(q, p)$. From this point of view, a $\tau$-time-discretized computation of one of the mentioned dynamical models is nothing else but a realization of a Markov chain in phase space. A generalized transfer operator $\mathcal{P}(\tau) : \mathcal{L}_{\pi^\tau}(\Omega) \to \mathcal{L}_{\pi^\tau}(\Omega)$ can be written as:

$$\mathcal{P}(\tau) f(q) = \int_{\mathbb{R}^{3N}} \left( \int_{\Omega} f(\tilde{q}) \, \Psi^{-\tau}(\tilde{q} | (q, p)) \, d\tilde{q} \right) \pi_p(p) \, dp. \quad (2.7)$$

In equation (2.7), the initial state $(q, p)$ determines a probability density function $\Psi^{-\tau}(- | (q, p))$ for the possible evolutions of the system in configurational space in time $\tau$. For an explanation see Figure 2.3. $\Psi^{-\tau}$ is a Dirac delta function in the case of deterministic dynamics, because the initial state $(q(0), p(0))$ exactly defines the final configurational state $q(-\tau)$. Equation (2.7) can be used to define a generalized transfer operator for any of the dynamical models (deterministic and stochastic) mentioned above, even in the case of a dynamical model which is independent from momentum variables\textsuperscript{7} – like Smoluchowski dynamics. $\mathcal{T}$ in (2.4) is a special case\textsuperscript{8} of $\mathcal{P}$. There is a very simple way to derive a transition matrix $P$ from the continuous operator $\mathcal{P}$ via (2.6), where $\mathcal{T}$ is replaced by $\mathcal{P}$. One can simply count the transitions between subsets (defined by $\Phi$) of $\Omega$ in a long-term molecular dynamics trajectory generated by one of the above ergodic dynamical models for lag-time $\tau$. This gives the nominator in (2.6). Dividing this number by the equilibrium population of each set $\Phi_i$ directly leads to a transition matrix $P$. This direct sampling approach is very common, but it does not solve the conformation dynamics problem, as mentioned in the introduction. An adaptive discretization approach for the computation of $P$ is needed.

\textsuperscript{7}In this case $\Psi^{-\tau}$ is independent from $p$.

\textsuperscript{8}The lag-time dependence $\mathcal{P}(\tau)$ is omitted sometimes. In this case, a common property of all operators $\mathcal{P}(\tau)$ for $\tau > 0$ is addressed.
2.4 Hierarchical, adaptive, and meshless discretization

As mentioned in the introduction, long-term molecular dynamics trajectories are not suitable for solving the conformation dynamics problem. Long-term trajectories include a lot of redundant data concerning the local distribution of states inside the metastable subsets of $\Omega$. They do not contain enough data for the evaluation of the transition pattern between these sets or of the statistical weights of the conformations. Suitable discretizations $\Phi$ for the approximation $P$ of $T$ can be found adaptively and hierarchically with a set of meshless basis functions. This has been shown by Weber [30] and Röblitz [25]. Their concept can be extended to ergodic dynamical models, i.e., to the generalized operator $P(\tau)$. Instead of computing one long-term trajectory, one can estimate the transition matrix $P(\tau)$ on the basis of many short-time trajectories of length $\tau$. Once one has estimated a transition matrix $P(\tau)$ in this way, it is also possible to initiate an hierarchical refinement of the set of basis functions $\Phi$ in order to improve the estimated transition behavior [25]. This adaptive hierarchical scheme cannot be mesh-based, because of the high-dimensional configurational space $\Omega$. A meshless discretization approach is mandatory. A Voronoi tessellation $\Phi$ of $\Omega$ is recommended as meshless approach, see also [30, 25]. A Voronoi tessellation is based on nodes $q_1, \ldots, q_m \in \Omega$ and on a distance measure $\text{dist} : \Omega \times \Omega \rightarrow R_+$. The nodes are configurational states that represent the different configurations of the molecular system in a sufficient way. The basis functions are given by:

\[
\Phi_i(q) = \begin{cases} 
1, & \text{dist}(q, q_i) = \min_{k=1,\ldots,m} \text{dist}(q, q_k) \\
0, & \text{else} \end{cases} \quad (2.8)
\]

In a hierarchical sampling approach, the set of basis functions $\Phi$ is not extended by simply adding nodes to the Voronoi tessellation. This would also cause the “old” basis functions and integrals in (2.6) to have to be recomputed. In a hierarchical approach [30, 25], a certain basis function $\Phi_i$ is determined, which will be refined. The basis function $\Phi_i$ is eliminated from the set of basis functions.
Given a set of basis functions $\Phi$, finally, the new basis functions $\Phi, \Phi_1, \ldots, \Phi_m$ are added to the set of basis functions. With this procedure, the partition of unity is preserved (even in the case where $\Phi_i$ can have values between 0 and 1, which will be described later). The other basis functions $\Phi_j, j \neq i$, are unchanged. Given a set of basis functions $\Phi_i, i = 1, \ldots, m$, the matrix element $P(i, j)$ of $P(\tau)$ can be computed via (2.6) as

$$P(\tau)(i, j) = \frac{\langle \Phi_i, \mathcal{P}(\tau) \Phi_j \rangle_{\pi_q}}{\langle \Phi_i, e \rangle_{\pi_q}} = \frac{\int_\Omega \Phi_i(q) \mathcal{P}(\tau) \Phi_j(q)_{\pi_q} dq}{\int_\Omega \Phi_i(q)_{\pi_q} dq} = \int_\Omega \int R^{3N} \left( \frac{\int \Phi_j(\bar{q}) \Psi_{-\tau}(\bar{q})(q, p) dq}{\int_\Omega \Phi_i(q)_{\pi_q} dq} \right) \frac{\Phi_i(q)_{\pi_q} dq}{\pi_p(p)} dp dq$$

The above expression is an expectation value of an observable $O(q, p)$ according to a distribution $\pi_i(q, p)$ of states $(q, p)$. The distribution is given by the term (II):

$$\pi_i(q, p) := \frac{\Phi_i(q)_{\pi_q}}{\int_\Omega \Phi_i(q)_{\pi_q} dq} \pi_p(p).$$

$\pi_i$ is a Boltzmann distribution of molecular states, where the configurational part is restricted to a subset $\Phi_i$ of $\Omega$. The observable (I) itself is again an expectation value. The observable of this nested expectation value is $\Phi_j(\bar{q})$ and the distribution is given by $\Psi_{-\tau}(\bar{q})(q, p)$. A very common and efficient method to evaluate continuous expectation values numerically is the following approach: Generate a set of states according to the given distribution. Then compute the mean value of the observable for the generated set of states. In the case of $P(\tau)(i, j)$ one has to generate a set of Boltzmann distributed states $(q, p)$, where the $q$-variable is restricted to a subset $\Phi_i$. With this set of initial states $(q, p)$, one has to compute the mean value of $\Phi_j(\bar{q})$, where the states $\bar{q}$ are taken from different realizations of the dynamical model represented by $\Psi_{-\tau}(\bar{q})(q, p))$. In our group, the technical term for generating restricted states $(q, p)$ is horizontal sampling. The different realizations of the dynamical model based on states $(q, p)$ are called vertical sampling. In this Monte Carlo quadrature approach for the estimation of $P(\tau)(i, j)$, the approximated transition matrix $\tilde{P}(\tau)$ is a random matrix due to truncated (finite) sampling. There are two main questions to be solved in an adaptive sampling scheme for the estimation of $P(\tau)$:

**AS1** Given a discretization $\Phi$. How many horizontal and how many vertical sampling points should be generated in order to estimate $\tilde{P}$?

**AS2** Given a sampling and an estimation $\tilde{P}$. How should the discretization set $\Phi_i$ be determined that has to be refined?

Röblitz [25] proposed a solution for these two questions. Her ideas were based on the adaptive sampling approach of Singhal Hinrichs and Pande [17]. The horizontal and the vertical sampling have to provide enough data for the statistics.
in order to approximate \( P(\tau) \) well. During the horizontal sampling, a transition matrix \( \tilde{P} \) is not yet available. For this sampling, Röblitz proposed a hybrid Monte-Carlo method for each subset \( \Phi_i \) of \( \Omega \). She applied a distribution based Gelman-Rubin convergence indicator \cite{15} as a stopping criterion (AS1). If a maximal number of sampling steps has been reached during the horizontal sampling of a subset \( \Phi_i \), this basis function has to be refined (AS2). After the horizontal sampling has converged for all basis functions \( \Phi_i \), the matrix elements of \( \tilde{P} \) can be sampled. Each horizontal sampling \( i \) can be used in order to start vertical samplings to compute one row \( \tilde{P}(\tau)(i,:) \) of \( \tilde{P} \). Thus, the matrix \( \tilde{P} \) is a row-wise correlated random matrix. This structure can be used to derive error bonds based on a stochastic error norm \cite{25}. Röblitz computed an error bound for the dominant eigenspace of \( \tilde{P} \) in order to define a stopping criterion for the vertical samplings (AS1). This error bound also identifies the basis function \( \Phi_i \) which mainly contributes to this error. The vertical sampling of this basis function has to be extended. If a vertical sampling based on the horizontal sampling of \( \Phi_i \) does not converge in a pre-defined maximal number of sampling steps, \( \Phi_i \) has to be refined (AS2). The procedure has to be repeated until every horizontal and every vertical sampling has converged.

**Advantages of an adaptive sampling.** In fact, the adaptive, hierarchical, meshless sampling approach by Röblitz can be seen as the solution of the conformation dynamics problem, especially of the question Q2 in the introduction. The approximation \( \tilde{P} \) of \( P \) is error-based and can be improved by adding basis functions to \( \Phi \). The algorithm avoids the sampling of redundant data, because of its adaptive structure. It also avoids long-term dynamics simulations and can be applied to non-ergodic dynamical models like Hamiltonian dynamics. Concerning the identification of conformations \( \chi \) on the basis of \( P \), the adaptive sampling approach is also effective and robust, because \( \chi \) is computed from the dominant eigenspace of \( P \), which is the error-controlled object in the algorithm. Especially in transition regions, the discretization of \( \Omega \) is refined in the adaptive sampling algorithm. The algorithm satisfies requirement R2 given in the introduction.

**Computing the statistical weights with an adaptive sampling.** A direct sampling approach with an ergodic thermostated dynamical model provides statistical weights of the conformations by simply counting the states of the trajectory sampled per conformation. In an adaptive sampling approach, the statistical weight \( w_i := (\chi_i, e)_{\pi_q} \) of the conformation \( \chi_i \) cannot be estimated by counting the states in the set \( \chi_i \). The reason is that the number of sampling points generated per discretization set \( \Phi_i \) is not determined by the weight \( d_i \), it is the result of a convergence criterion. Because of the following equation

\[
(d^T P)_j = \sum_{i=1}^{n} d_i \frac{\langle \Phi_i, P(\tau) \Phi_j \rangle_{\pi_q}}{d_i} \\
= \langle e, P(\tau) \Phi_j \rangle_{\pi_q} \\
= \int_{\Omega} \int_{RRN} \int_{\Omega} \Phi_j(\tilde{q}) \Psi_{-\tau}(\tilde{q} | (q, p)) \pi_p(p) \pi_q(q) d\tilde{q} dp dq \\
= \int_{\Omega} \Phi_j(\tilde{q}) \left( \int_{RRN} \int_{\Omega} \Psi_{-\tau}(\tilde{q} | (q, p)) \pi_p(p) \pi_q(q) dq dp \right) d\tilde{q}
\]
\[ (*) = \int_{\Omega} \Phi_j(q) \pi_q(q) d\tilde{q}, \]
\[ = (e, \Phi_j) \pi_q, \]
\[ = d_j, \] (2.9)

for the \( j \)-th element of the vector matrix product \( d^T P(\tau) \), the weights \( d \) can be estimated by computing the left eigenvector \( \tilde{d} \) of the approximation \( \tilde{P}(\tau) \) of \( P(\tau) \) for the eigenvalue 1. The step \( (*) \) in (2.9) uses the fact that \( \pi_q \pi_q \) is the stationary Boltzmann density and, therefore, independent from a propagation via \( \Psi_{-\tau} \). The eigenvalue \( \lambda_1 = 1 \) is the dominant eigenvalue of \( P \) according to a Gerschgorin estimation [16] for a row-stochastic matrix \( P \). If the matrix \( P \) is irreducible\(^9\) and primitive\(^{10}\), then the theorem of Frobenius and Perron [24] says that the dominant left eigenvector of \( P \) is positive. Furthermore, it is geometrically and algebraically simple. Thus, the statistical weights of \( P \) are uniquely defined by (2.9). Although, the computation of an approximation \( \tilde{d} \) of \( d \) via solving an eigenvalue problem can have a unique solution, the condition number of the eigenproblem \( \tilde{d}^T = \tilde{d}^T \tilde{P} \) can be very high. Weber et al. [33] have shown, that in the case of a metastable dynamical system the computation of the stationary density by solving the eigenproblem (2.9) is ill-conditioned. A very small error \( \| P - \tilde{P} \|_\infty \) can lead to a very large error \( \| d - \tilde{d} \|_\infty \leq \kappa \| P - \tilde{P} \|_\infty \). Simply adding more functions to the basis \( \Phi \), cannot improve the estimation of the stationary density, because the condition of the weight computation depends on the eigenvalue structure of \( P \) which mainly depends on \( \mathcal{P} \) and not on the discretization \( \Phi \). For the Meyer condition number \( \kappa \) of the statistical weight computation, the following estimation holds:

\[ \frac{1}{m |1 - \lambda_2|} \leq \kappa \leq \frac{2(m - 1)}{\prod_{i=2}^{m} (1 - \lambda_i)}, \] (2.10)

where \( \lambda_1, \ldots, \lambda_m \) are the sorted eigenvalues of \( P \), see [14, 23, 30] and (3.3) in [5]. Thus, even if \( d \) is taken as the error-controlled object of the adaptive sampling approach (see [22]), this only can improve the result to a certain degree.

A possible solution of this condition problem has been mentioned in the introduction: One has to separate the estimation of the statistical weights from the computation of transition probabilities. This estimation of the statistical weights can be done by a Markov chain which jumps between the basis functions \( \Phi_i \) or between the conformations \( \chi_i \) [29, 33] rapidly. A rapidly mixing Markov chain can answer question Q1 given in the introduction and it satisfies the corresponding requirement R1. In fact, there are many possible approaches to a well-conditioned solution of the statistical weights problem in literature: The ratio of the statistical weight \( d_i \) of subset \( \Phi_i \) and the weight \( d_j \) of subset \( \Phi_j \) can be written in terms of a free energy difference \( \Delta A_{ij} \) between \( \Phi_i \) and \( \Phi_j \). This free energy difference is connected to the ratio via

\[ \Delta A_{ij} = -\frac{1}{\beta} \ln \left( \frac{d_i}{d_j} \right). \] (2.11)

The estimation of the statistical weights can be done with methods for the computation of free energy differences. Note that for the computation of free

\(^9\)It cannot be decomposed into independent block matrices.

\(^{10}\)For a primitive matrix \( A \), there is a number \( k \in N \) with \( A^k > 0 \).
energy differences there are a lot of efficient sampling approaches which can also be applied in the context of this text. For an excellent overview see Chipot [4]. All of the algorithms shown in that textbook circumvent the problem of rare events because of the aforementioned reasons (bad condition number for metastable systems).
Bibliography


