A Subspace Approach to Molecular Markov State Models via a New Infinitesimal Generator

Marcus Weber

January 7, 2010

Contents

1 Introduction **2**
- 1.1 Problem formulation . . . 2
- 1.2 Outline . . . 5

2 Molecular kinetics in the canonical ensemble **7**
- 2.1 From molecular dynamics to Markov State Models . . . 7
- 2.2 Transfer operator approach . . . 11
- 2.3 Thermostated molecular dynamics simulations . . . 12
- 2.4 Periodic boundary condition . . . 14
- 2.5 Hierarchical, adaptive, and meshless discretization . . . 15
- 2.6 Markovian milestoning with Voronoi tessellations . . . 20
- 2.7 Committor functions . . . 22

3 Design of a generalized Markov operator **25**
- 3.1 Desirable properties . . . 25
- 3.2 From a transfer operator to a Markov operator . . . 28
- 3.3 Subspace approach . . . 29
- 3.4 Discretization error . . . 34
- 3.5 An infinitesimal generator . . . 39
- 3.6 Application of Gauß' theorem . . . 42

4 Algorithmic Details **45**
- 4.1 Meshless basis functions . . . 45
- 4.2 Adaptive and hierarchical refinement . . . 48
- 4.3 Efficient free energy calculation . . . 49
- 4.4 Approximation of the surface flux . . . 54
- 4.5 Efficiency of the algorithmic approach . . . 57

5 Numerical Examples **61**
- 5.1 Illustrative Example . . . 61
- 5.2 Real-World Examples . . . 68
  - 5.2.1 State of the art molecular simulations . . . 69
  - 5.2.2 Binding path of an APN inhibitor . . . 70

Conclusion **75**

Chapter 1

Introduction

1.1 Problem formulation

The computer simulation of molecular systems is a research area with a fine tradition. Nowadays, supercomputers like ANTON [66] 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)[29]. 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) [84]. Thus, in order to simulate one second in "real life" ANTON needs about 270 years of computing time[1]. 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. [20, 18, 21, 63, 19] introduced this set-based approach to conformation dynamics, which was inspired by ideas of Michael Dellnitz et al. [17] 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] 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.

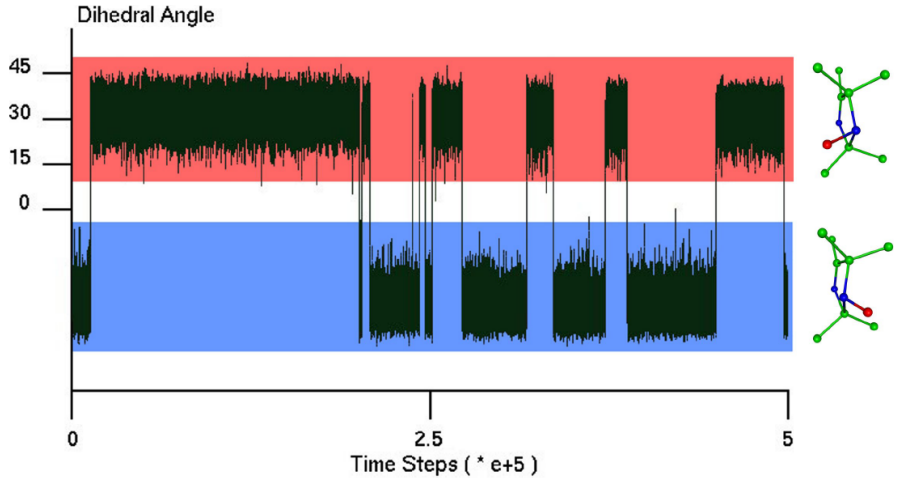

Figure 1.1: A typical time-series plot of an observable from a molecular dynamics simulation of a small molecule.

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 of this thesis. 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 [19]. 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

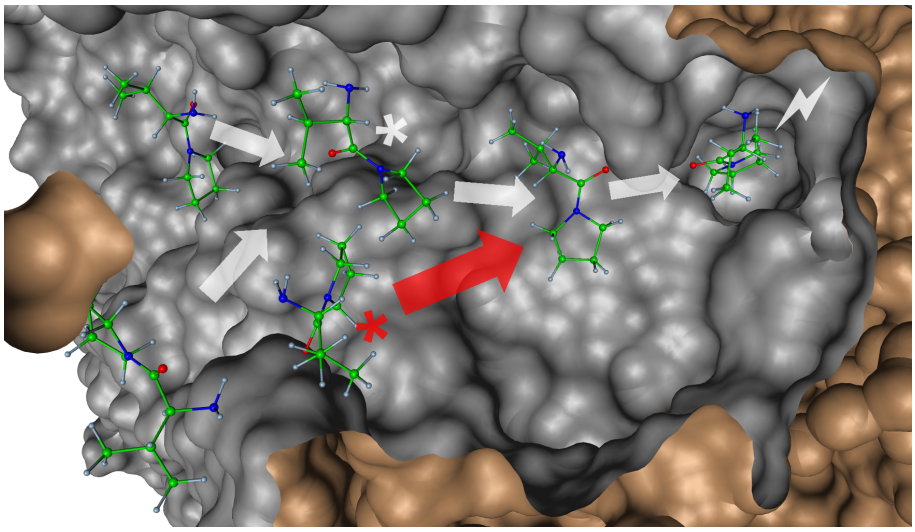

Figure 1.2: A binding path of a small ligand molecule from the surface of a protein into the active site. This is a result of a molecular kinetics simulation and cannot be computed in reasonable time using standard MD methods [9].

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 [81]. 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 focusses 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 and will be the main part of the present thesis. The first requirement R1 can be implemented by using a special sampling scheme,

which will be shown in chapter 4.3.

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

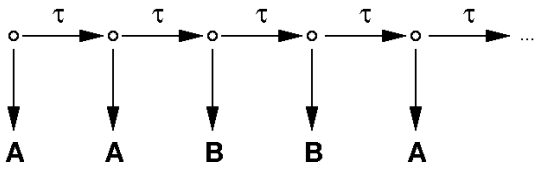

Figure 1.3: The states of a trajectory of a molecular dynamics simulation are assigned to two conformations A and B.

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 [76, 80]. 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 thesis 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 [63]. 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.

1.2 Outline

The thesis is organized as follows. Chapter 2 presents the basic assumptions of molecular kinetics on the basis of a canonical ensemble. The difference between

molecular dynamics and molecular kinetics will be derived and common concepts will be illustrated. It will be shown that these concepts directly lead to a definition of a transfer operator in equation (2.3). The definition of the transfer operator is depending on the dynamical model of the molecular system. One example is given by Schütte's transfer operator for Hamiltonian dynamics in chapter 2.2 equation (2.4). Since Hamiltonian dynamics does not lead to self-equilibration of the canonical ensemble, a generalization of the transfer operator approach is derived in chapter 2.3 equation (2.7) for arbitrary dynamical models - including stochastic molecular processes. In chapter 2.5, it will be shown, how an existing algorithmic approach of Röblitz to the discretization of Schütte's transfer operator can be used for the generalized transfer operator, too. This algorithmic approach can be computationally expensive. In chapters 2.6 and 2.7 alternative, efficient algorithmic approaches to the simulation of molecular kinetics are analyzed. It is shown that the common theoretical basis for all efficient approaches is the assumption of an infinitesimal generator of the generalized transfer operator. Schütte's transfer operator does not meet this assumption. Therefore, a new generalized transfer operator is designed in chapter 3. Instead of selecting one special dynamical model in order to simulate molecular processes, desirable properties of the transfer operator are identified. Theoretical results for the designed, generalized transfer operator are deduced in chapter 3.1. In chapter 3.3 it is shown, how this new generalized transfer operator leads to a valid Markov State Model. Especially, the importance of the Robust Perron Cluster Analysis (PCCA+) in the context of correct time-extrapolation will be shown in Theorem 2. For this purpose, the PCCA+ algorithm will be extended from discrete transition matrices to continuous operators in chapter 3.4. The aforementioned assumption of an existing infinitesimal generator of the generalized transfer operator is one important aspect for a correct time-extrapolation, this assumption may be too rigorous for realistic molecular models. This thesis offers different optional approaches to the transition rate computation even if an infinitesimal generator does not exist, see chapter 3.5. From an algorithmic point of view, Theorem 4 in chapter 3.6 (based on Gauß' Theorem) will be used to effectively compute the discretization of the infinitesimal generator in high-dimensional spaces. Some algorithmic details of an efficient simulation of molecular kinetics are given in chapter 4. Instead of estimating the transition pattern by a time-consuming realization of the metastable dynamical model (and by "waiting" for the rare transition events), only the computation of Boltzmann distributions at the surfaces of Voronoi cells is needed which can be done with accelerated (rapidly mixing) sampling schemes. The algorithmic point of view will be changed with this new concept: It is not important to simulate a statistically relevant number of transitions in order to compute a Markov State Model, it is even not necessary to *simulate* a molecular system according to a metastable dynamical model. Instead of applying molecular dynamics simulations, it is only important to figure out free energy differences between certain subsets of the configurational space efficiently. The latter requirement will be solved by Jarzynski's Identity in chapter 4.3. Finally, two different numerical examples will be shown in chapter 5: An illustrative two-dimensional example and a real-world molecular simulation of a binding process.

Chapter 2

Molecular kinetics in the canonical ensemble

All considerations in the present thesis 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 in chapter 3.

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 of this thesis, 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 $q \in \Omega \subset R^{3N}$ and momentum state $p \in R^{3N}$. 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, $V : \Omega \to R$, only depending on q, and its kinetic energy, $K : R^{3N} \to R$, only depending on p. Taking all these conditions into account, Boltzmann derived the probability density function $\pi : \Omega \times R^{3N} \to R_+$ of states (q,p) of the subsystems as

$$\pi(q,p) = \frac{1}{Z} \exp(-\beta\, H(q,p)), \tag{2.1}$$

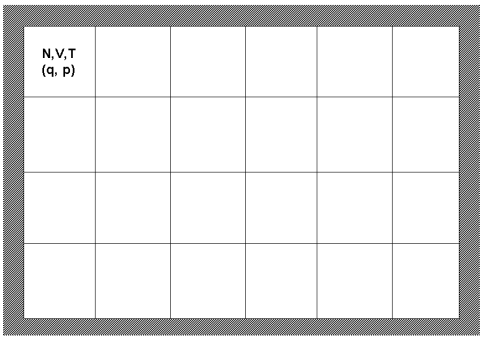

Figure 2.1: An isolated system divided into a large number of closed subsystems. Each subsystem is a copy of the same molecular system with the same number of atoms and the same volume, but with a different molecular state given by configuration q and momentum coordinates p.

where $Z > 0$ is the normalization constant (also called *partition function*), and β 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 [42] of the isolated system[1]. The factor β 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*[2]. In this dynamical model, the time-evolution of the states is given by a first order differential equation

$$\begin{aligned} \dot{q}(t) &= \nabla_p K(p(t)), \\ \dot{p}(t) &= -\nabla_q V(q(t)), \end{aligned} \tag{2.2}$$

[1] 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.

[2] 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 π 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[3]. 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

[3] An 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 $\rho : R \times \Omega \times R^{3N} \to R$ of states $(q, p) \in \Omega \times R^{3N}$ at time t and a *lag-time* $\tau > 0$, the evolution of the probability density function at time $t + \tau$ can be expressed as:

$$\rho(t + \tau, \cdot, \cdot) = \overline{\mathcal{P}}_s(\tau)\, \rho(t, \cdot, \cdot), \tag{2.3}$$

where $\overline{\mathcal{P}}_s(\tau) : L^1(\Omega \times R^{3N}) \to L^1(\Omega \times R^{3N})$ is a lag-time-dependent (but not time-dependent) operator which propagates probability density functions[4]. In chapter 3.2 it will be shown that our approach defines a Markov operator $\overline{\mathcal{P}}(\tau)$, a projection of $\overline{P}_s(\tau)$ to configurational space Ω. 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 $w(t) \in R^n$. The elements of this vector are given by the statistical weights of the n conformations at time t. In Figure 5.10 at the end of this thesis, a plot of the time-dependent evolution of a w-vector (with 3 elements) is shown as an example for this kind of complexity reduction. 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 $P_c^\top(\tau)$ via $w(t + \tau) = P_c^\top(\tau)\, w(t)$. In order to get a commuting

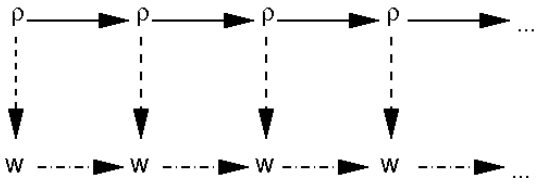

Figure 2.2: Complexity reduction of molecular kinetics. Solid lines: propagation of probability density functions via $\overline{\mathcal{P}}_s(\tau)$. Dashed lines: projection of probability density functions to a low-dimensional vector w. Dash-dotted lines: propagation of statistical weights via an $n \times n$-matrix $P_c^\top(\tau)$.

diagram in Figure 2.2, the matrix $P_c^\top(\tau)$ 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 $P_c(\tau)$ 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

[4]Note that for a non-negative probability density function ρ the normalization $1 = \int_{R^{3N}} \int_\Omega \rho(t, q, p)\, dq\, dp$ holds, i.e., $\rho(t, \cdot, \cdot) \in L^1(\Omega \times R^{3N})$. 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.

Markov state of the system. The transition behavior is given by $P_c(\tau)$. The aim of conformation dynamics is to compute this Markov State Model[5].

2.2 Transfer operator approach

The set-based concept of the transfer operator approach [63] 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 of this thesis, the total energy H is the sum of kinetic and potential energy. The Boltzmann distribution π can be decomposed, $\pi = \pi_p \pi_q$, into a probability density function $\pi_p : R^{3N} \to R$ for the kinetic energy in momentum space and a probability density function $\pi_q : \Omega \to R$ 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 $\mathcal{T}(\tau) : L^{1,2}_{\pi_q}(\Omega) \to L^{1,2}_{\pi_q}(\Omega)$ on the basis of Hamiltonian dynamics as

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

Equation (2.4) can be understood as follows: Given an initial state (q,p), a backward Hamiltonian dynamics for a lag-time τ is investigated. The final state is denoted as $\Psi^{-\tau}(q,p)$. Via Π_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 π_p. The definition of conformations as metastable sets leads to a decomposition of the configurational space Ω. In the set-based approach, the conformations $\chi_1, \ldots, \chi_n$ are given by the characteristic functions of the corresponding subsets of Ω, i.e., in terms of functions $\chi_i : \Omega \to \{0,1\}$. The conformations χ_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 χ in practice, the configurational space is decomposed into a larger (but finite) number of subsets represented by characteristic functions $\Phi_i : \Omega \to \{0,1\}$, $i = 1, \ldots, m$, 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 χ_i of the configurational space. The discretization scheme of the transfer operator approach to conformation dynamics can be written as

$$\mathcal{T} \to P \to P_c. \tag{2.5}$$

A continuous transfer operator $\mathcal{T}(\tau)$ 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 Φ leads to a transition matrix $P(\tau)$ which is used to identify the conformations χ and the Markov State Model $P_c(\tau)$. On the basis of the transfer operator $\mathcal{T}(\tau)$, the element

[5] If $P_c^\top(\tau)$ is column-stochastic, then $P_c(\tau)$ is a *row-stochastic matrix*. However, $P_c^\top(\tau)$ is not necessarily column-stochastic. Nevertheless, the technical term "Markov State Model" will be used for $P_c(\tau)$ throughout this thesis.

(i,j) of the transition matrix $P(\tau)$ can be computed as

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

where $\langle f, g \rangle_{\pi_q}$ is the π_q-scalar product defined as $\int_\Omega f(q)\, g(q)\, \pi_q(q)\, dq$. The function $e : \Omega \to \{1\}$ is a constant. The nominator in (2.6) counts the number of states which undergo a transition from set Φ_i to set Φ_j in time τ. This number is divided by the statistical weight of set Φ_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 Φ_i to set Φ_j. The matrix P_c can be derived from the operator $\mathcal{T}$ in a similar way. Just exchange the Φ-basis functions by χ-basis functions in (2.6). $\mathcal{T}(\tau)$ in (2.4) is a transfer operator. $\mathcal{T}(\tau)$ is not acting on density functions, it is acting on membership functions. Thus, stationarity is characterized by the equation $e = \mathcal{T}(\tau)\, e$, where e is the constant function $e \equiv 1$ in Ω. In contrast to $\mathcal{T}$, the operator $\overline{\mathcal{P}}_s(\tau)$ in (2.3) is acting on density functions. In this case, stationarity is characterized by $\pi = \overline{\mathcal{P}}_s(\tau)\, \pi$, with the Boltzmann density π. Consequently, $\mathcal{T}$ is discretized with a set of basis functions Φ or χ, which is a partition of *unity* (and not a partition of π_q). $\mathcal{T}$ and the projection of $\overline{\mathcal{P}}_s$ to configurational space are *adjoint* operators, which can be seen by the fact that a χ-discretization of $\mathcal{T}$ leads to P_c and a χ-discretization of $\overline{\mathcal{P}}_s$ 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 τ 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 [30, 36]. Another example is the Berendsen thermostat [5] which does not generate the canonical ensemble exactly. Other time-reversible deterministic thermostats can be found in [37]. 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 [38] and Langevin dynamics [61], the class of hybrid Monte-Carlo methods (HMC) [22] is an example for a stochastic dynamical model[6]. In HMC, the system is mainly propagated according to (2.2). Sole exception: After a certain time-span the 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 $(q, p) \in \Omega \times R^{3N}$, 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 τ-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 $\mathcal{P}(\tau) : L^{1,2}_{\pi_q}(\Omega) \to L^{1,2}_{\pi_q}(\Omega)$ can be written as:

$$\mathcal{P}(\tau)\, f(q) = \int_{R^{3N}} \Big(\int_{\Omega} f(\widetilde{q})\ \Psi_{-\tau}\big(\widetilde{q}|(q,p)\big)\, d\widetilde{q} \Big)\, \pi_p(p)\, dp. \tag{2.7}$$

In equation (2.7), the initial state (q, p) determines a *probability density function* $\Psi_{-\tau}\big(\cdot|(q,p)\big)$ for the possible evolutions of the system in configurational space in time τ. 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[7] – like Smoluchowski dynamics. $\mathcal{T}$ in (2.4) is a special case[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 Φ) of Ω in a long-term molecular dynamics trajectory generated by one of the above ergodic dynamical models for lag-time τ. This gives the nominator in (2.6). Dividing this number by the equilibrium population of each set Φ_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 which will be discussed in chapter 2.5.

[6] The data in Figure 1.1 have been generated with HMC.

[7] In this case $\Psi_{-\tau}$ is independent from p.

[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.

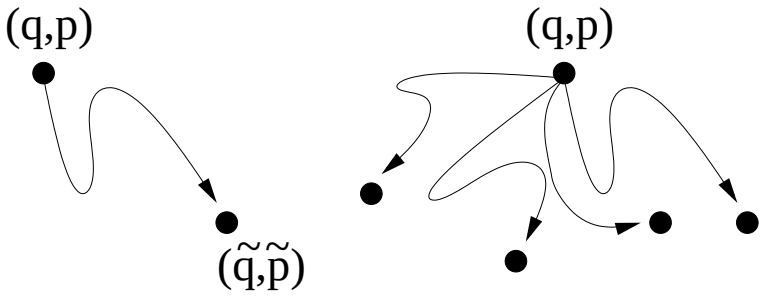

Figure 2.3: *Left.* In the case of (2.4), Hamiltonian dynamics is deterministic. A given initial state (q, p) leads to a fixed propagated state $(\widetilde{q}, \widetilde{p}) = \Psi^{-\tau}(q, p)$. *Right.* In the general case of (2.7) with a stochastic differential equation or with a Markov chain, the initial state (q, p) is propagated to different states with a different probability. $\Psi_{-\tau}\big(\,\cdot\,|(q,p)\big)$ is the corresponding probability density function in Ω.

2.4 Periodic boundary condition

The potential energy function $V : \Omega \to R$ in this thesis is usually considered to be an empirical potential energy function (also called *molecular force field*) modelled on the basis of additive covalent and non-covalent energy contributions to the molecular system. A simulation of a molecular system with potential energy function V at constant temperature is only possible if the canonical ensemble is assumed. The canonical ensemble probability function of Boltzmann (2.1) only makes sense, if the partition function Z exists. Especially, the integral

$$\int_{\Omega} \exp(-\beta\, V(q))\, dq$$

must exist. If $\Omega = R^{3N}$, the integral expression can be divergent, because translation and rotation in general does not change the value of $V(q)$. One possibility for a corresponding definition of a compact set $\Omega \subset R^{3N}$ is connected to the iso-volume construction of the canonical ensemble. In practice, the atom coordinates are usually restricted to a certain three-dimensional box in position space. Thus, the partition function exists. The severe problem in this context is the construction of a potential energy function V. This problem can be explained as follows. Although the subsystems of the canonical ensemble are closed and do not interchange matter, they interchange "potential energy interaction" with their surrounding neighbors. In order to compute the potential energy of a certain molecular state $q \in \Omega$, one often assumes periodic boundary conditions for the box, as if the neighbors of the subsystem are copies of the same conformational state q. Using this modelling approach, boundary effects are excluded from the simulation.

Molecular dynamics. From a thermostated molecular dynamics point of view, this modelling is incorrect. If every subsystem is a copy of the same molecular state (q, p), then the corresponding ensemble is not canonical and "temperature" (connected to the Boltzmann factor β) can not be defined.

Molecular kinetics. In principle, the argument against periodic boundary conditions also holds for molecular kinetics. But there is a possibility to correct this inconsistency within the theoretical framework of the generalized transfer operator $\mathcal{P}$ in (2.7) and Figure 2.3. Instead of assuming that $V(q)$ models a potential energy of a conformational state $q \in \Omega$ physically, it is only assumed that $V(q) := -\frac{1}{\beta}\ln(\pi_q(q))$ is a (complicated) expression for the probability density of "finding" a certain conformational state q in the canonical ensemble[9]. In the framework of (2.7), it is allowed that a certain initial state (q,p) leads to a probabilty density $\Psi_\tau\big(\widetilde{q}|(q,b)\big)$ for the evolution of the system in time τ. E.g. in Hamiltonian dynamics this can account for the fact, that one initial state (q,p) can evolve differently in time depnding on the conformational state of the neighboring subsystems at time $t=0$. From this point of view, the generalized transfer operator (2.7) not only accounts for different dynamical models, it also accounts, theoretically, for the fact that the time-dependent evolution of an initial state (q,p) only depends on this state, not in the sense of one trajectory, but in the sense of a probability density function.

2.5 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 Ω. 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 Φ for the approximation P of $\mathcal{T}$ can be found adaptively and hierarchically with a set of meshless basis functions. This has been shown by Weber [75] and Röblitz [58]. Their concept can be extended to ergodic dynamical models, i.e. to the generalized operator $\mathcal{P}(\tau)$. 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 τ. 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 Φ in order to improve the estimated transition behavior [58]. This adaptive hierarchical scheme cannot be mesh-based, because of the high-dimensional configurational space Ω. A meshless discretization approach is mandatory. In this thesis, a *Voronoi tessellation* Φ of Ω is recommended as meshless approach, see also [75, 58]. A Voronoi tessellation is based on *nodes* $q_1,\ldots,q_m \in \Omega$ and on a *distance measure* $\mathrm{dist} : \Omega\times\Omega \to R_+$. The nodes are configurational states that represent the different configurations of the molecular system in a sufficient way. The basis functions are given by:

$$\Phi_i(q) = \begin{cases} 1, & \mathrm{dist}(q,q_i) = \min\limits_{k=1,\ldots,m} \mathrm{dist}(q,q_k) \\ 0, & \text{else} \end{cases} . \tag{2.8}$$

[9] In order to implement an approximate(!) mathematical model for the computation of the expression $V(q)$ numerically, one can use periodic boundaries. That is the trick.

In a hierarchical sampling approach, the set of basis functions Φ 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 [75, 58], a certain basis function Φ_i is determined, which will be refined. The basis function Φ_i is eliminated from the set of basis functions. A new Voronoi tessellation $\widetilde{\Phi}_1, \ldots, \widetilde{\Phi}_{\widetilde{m}}$ is defined with nodes $\widetilde{q}_1, \ldots, \widetilde{q}_{\widetilde{m}} \in \Omega$ stemming from subset Φ_i. Finally, the new basis functions $\Phi_i\widetilde{\Phi}_1, \ldots, \Phi_i\widetilde{\Phi}_{\widetilde{m}}$ are added to the set of basis functions. With this procedure, the partition of unity is preserved (even in the case where Φ_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

$$\begin{aligned} P(\tau)(i,j) &= \frac{\langle \Phi_i, \mathcal{P}(\tau)\Phi_j \rangle_{\pi_q}}{\langle \Phi_i, e \rangle_{\pi_q}} \\ &= \frac{\int_\Omega \Phi_i(q)\, \mathcal{P}(\tau)\Phi_j(q)\, \pi_q\, dq}{\int_\Omega \Phi_i(\bar{q})\, \pi_q(\bar{q})\, d\bar{q}} \\ &= \int_\Omega \int_{R^{3N}} \underbrace{\Big(\int_\Omega \Phi_j(\widetilde{q})\ \Psi_{-\tau}\big(\widetilde{q}|(q,p)\big)\, d\widetilde{q} \Big)}_{(I)} \underbrace{\frac{\Phi_i(q)\, \pi_q(q)}{\int_\Omega \Phi_i(\bar{q})\, \pi_q(\bar{q})\, d\bar{q}}\, \pi_p(p)}_{(II)}\ dp\, dq \end{aligned}$$

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):

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

π_i is a Boltzmann distribution of molecular states, where the configurational part is restricted to a subset Φ_i of Ω. The observable (I) itself is again an expectation value. The observable of this nested expectation value is $\Phi_j(\widetilde{q})$ and the distribution is given by $\Psi_{-\tau}(\widetilde{q}|(q,p))$. 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 Φ_i. With this set of initial states (q,p), one has to compute the mean value of $\Phi_j(\widetilde{q})$, where the states $\widetilde{q}$ are taken from different realizations of the dynamical model represented by $\Psi_{-\tau}(\cdot|(q,p))$. 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 $\widetilde{P}(\tau)$ 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 Φ. How many horizontal and how many vertical sampling points should be generated in order to estimate $\widetilde{P}$?

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

Röblitz [58] proposed a solution for these two questions. Her ideas were based on the adaptive sampling approach of Singhal Hinrichs and Pande [35]. 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 $\widetilde{P}$ is not yet available. For this sampling, Röblitz proposed a hybrid Monte-Carlo method for each subset Φ_i of Ω. She applied a distribution based Gelman-Rubin convergence indicator [32] as a stopping criterion (AS1). If a maximal number of sampling steps has been reached during the horizontal sampling of a subset Φ_i, this basis function has to be refined (AS2). After the horizontal sampling has converged for all basis functions Φ_i, the matrix elements of $\widetilde{P}$ can be sampled. Each horizontal sampling i can be used in order to start vertical samplings to compute one row $\widetilde{P}(\tau)(i,:)$ of $\widetilde{P}$. Thus, the matrix $\widetilde{P}$ is a row-wise correlated random matrix. This structure can be used to derive error bonds based on a stochastic error norm [58]. Röblitz computed an error bound for the dominant eigenspace of $\widetilde{P}$ in order to define a stopping criterion for the vertical samplings (AS1). This error bound also identifies the basis function Φ_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 Φ_i does not converge in a pre-defined maximal number of sampling steps, Φ_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 $\widetilde{P}$ of P is error-based and can be improved by adding basis functions to Φ. 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 χ on the basis of P, the adaptive sampling approach is also effective and robust, because χ is computed from the dominant eigenspace of P (chapter 3.3), which is the error-controlled object in the algorithm. Especially in transition regions, the discretization of Ω 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 $w_i := \langle \chi_i, e \rangle_{\pi_q}$ of the conformation χ_i cannot be estimated by counting the states in the set χ_i. The reason is that the number of sampling points generated per discretization set Φ_i is not determined by the weight d_i, it is the result of a convergence criterion. Because of the following equation

$$\begin{aligned}(d^\top P)_j &= \sum_{i=1}^{m} d_i \cdot \frac{\langle \Phi_i, \mathcal{P}(\tau)\Phi_j \rangle_{\pi_q}}{d_i} \\ &= \langle e, \mathcal{P}(\tau)\Phi_j \rangle_{\pi_q}\end{aligned}$$

$$
\begin{aligned}
&= \int_\Omega \int_{R^{3N}} \int_\Omega \Phi_j(\widetilde{q})\ \Psi_{-\tau}\big(\widetilde{q}|(q,p)\big)\, \pi_p(p)\, \pi_q(q)\, d\widetilde{q}\, dp\, dq \\
&= \int_\Omega \Phi_j(\widetilde{q}) \Big(\int_{R^{3N}} \int_\Omega \Psi_{-\tau}\big(\widetilde{q}|(q,p)\big)\, \pi_p(p)\, \pi_q(q)\, dq\, dp \Big) d\widetilde{q} \\
(*) \quad &= \int_\Omega \Phi_j(\widetilde{q}) \pi_q(\widetilde{q})\, d\widetilde{q}, \\
&= \langle e, \Phi_j \rangle_{\pi_q} \\
&= d_j, \qquad (2.9)
\end{aligned}
$$

for the j-th element of the vector matrix product $d^\top P(\tau)$, the weights d can be estimated by computing the left eigenvector $\widetilde{d}$ of the approximation $\widetilde{P}(\tau)$ of $P(\tau)$ for the eigenvalue 1. The step $(*)$ in (2.9) uses the fact that $\pi_p \pi_q$ 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 [33] for a row-stochastic matrix P. If the matrix P is *irreducible*[10] and *primitive*[11], then the theorem of Frobenius and Perron [56] 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 $\widetilde{d}$ of d via solving an eigenvalue problem can have a unique solution, the condition number of the eigenproblem $\widetilde{d}^\top = \widetilde{d}^\top \widetilde{P}$ can be very high. Weber et al. [81] 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 $\|P - \widetilde{P}\|_\infty$ can lead to a very large error $\|d - \widetilde{d}\|_\infty \leq \kappa \, \|P - \widetilde{P}\|_\infty$. Simply adding more functions to the basis Φ, 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 Φ. For the Meyer condition number κ of the statistical weight computation, the following estimation holds:

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

where $\lambda_1, \ldots, \lambda_m$ are the sorted eigenvalues of P, see [31, 51, 75] and (3.3) in [11]. Thus, even if d is taken as the error-controlled object of the adaptive sampling approach (see [48]), 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 Φ_i or between the conformations χ_i [74, 81] 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 Φ_i and the weight d_j of subset Φ_j can be written in terms of a *free energy difference* ΔA_{ij} between Φ_i and Φ_j. This free energy difference is connected to the ratio via

$$
\Delta A_{ij} = -\frac{1}{\beta}\, \ln\Big(\frac{d_i}{d_j}\Big). \qquad (2.11)
$$

[10] It cannot be decomposed into independent block matrices.

[11] For a primitive matrix A, there is a number $k \in N$ with $A^k > 0$.

The estimation of the statistical weights can be done with methods for the computation of free energy differences, see also chapter 4.3. Note that for the computation of free energy differences there are a lot of efficient sampling approaches which can also be applied in the context of this thesis. For an excellent overview see Chipot [10]. 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).

Increasing number of Voronoi cells. The construction of a rapidly mixing Markov chain can solve the problem of computing statistical weights of the conformations. A rapidly mixing sampling scheme can also accelerate the horizontal sampling part of the adaptive algorithm. The horizontal sampling is used in order to generate a restricted Boltzmann distribution. It is not important to use a certain dynamical model for this issue. Thus, it is possible to construct an artificial Markov chain for this part of the procedure. The vertical sampling part, however, cannot be replaced by an artificial Markov chain. The vertical sampling has to correctly reproduce the dynamical model $\Psi_{-\tau}$. From this point of view, the increasing number of Voronoi cells during the adaptive sampling algorithm is a further problem. Transitions in lag-time τ do not only occur between neighboring cells. Therefore, the statistics of the vertical sampling algorithm has to be sufficient for an increasing number of possible transitions in order to approximate $P(\tau)$. This can lead to a slower convergence rate of the vertical sampling part at a higher level of the hierarchical refinement and can therefore introduce further refinements. This may be a vicious circle in the aforementioned adaptive sampling algorithm if convergence of the vertical sampling is crucial. For very short lag-times τ, however, the transitions between non-neighboring cells can be neglected. Thus, the trick is to take the derivative of $\mathcal{P}(\tau)$ for lag-time $\tau = 0$. In the following, we will assume that the operator $\mathcal{P}(\tau)$ defines an infinitesimal operator

$$\mathcal{Q} = \lim_{\tau \to 0+} \frac{\mathcal{P}(\tau) - \mathcal{I}}{\tau}, \tag{2.12}$$

where $\mathcal{I}$ is the identity operator. $\mathcal{P}$ in (2.7) has been defined on the basis of a dynamical model. The realizations of the dynamical model are time-continuous trajectories in configurational space. In this case, $\mathcal{Q}$ is well-defined via (2.12) [45]. In chapter 3.5, we will assume that the transfer operator $\mathcal{P}$ meets the Chapman-Kolmogorov equation $\mathcal{P}(\tau+\sigma) = \mathcal{P}(\tau)\,\mathcal{P}(\sigma)$ for all lag-times $\tau, \sigma > 0$. According to [45] $\mathcal{Q}$ is an infinitesimal generator of $\mathcal{P}$

$$\mathcal{P}(\tau) = \exp(\tau\,\mathcal{Q}). \tag{2.13}$$

Hamiltonian dynamics projected to configurational space (represented by the transfer operator $\mathcal{T}$) does not have an infinitesimal generator [38], because the Chapman-Kolmogorov equation does not hold. Smoluchowsky dynamics (also called Brownian dynamics) is an example for a dynamical model that has an infinitesimal generator [38]. Conceptually, $\mathcal{Q}$ is connected to the computation of *transition rates*. Many modern sampling approaches estimate transition rates between conformations instead of estimating transition probabilities. *Transition path sampling* [16], its more efficient variant called *transition interface sampling* [26], and its workable variant called *Markovian milestoning* [71] are important

examples of estimating transition rates. The basis for these algorithms is the assumption of the existence of an infinitesimal generator $\mathcal{Q}$.

2.6 Markovian milestoning with Voronoi tessellations

The transfer operator $\mathcal{P}$ is acting on membership functions. For a time-dependent membership function $f : R \times \Omega \to [0, 1]$, the following equations hold

$$\begin{aligned} & f(t+\tau,\cdot) = \mathcal{P}(\tau)\, f(t,\cdot) \\ \Leftrightarrow \quad & \frac{f(t+\tau,\cdot) - f(t,\cdot)}{\tau} = \frac{\mathcal{P}(\tau) - \mathcal{I}}{\tau}\, f(t,\cdot) \\ \Rightarrow \quad & \dot{f} = \mathcal{Q}\, f, \end{aligned} \tag{2.14}$$

where $\dot{f}$ is the time derivative of f. Thus, $\mathcal{Q}$ is the infinitesimal generator of a time-continuous Markov process, see [45]. This insight can be used in order to estimate $\mathcal{Q}$ on the basis of molecular dynamics simulations.

In this section, an algorithm is worked out which takes only transitions between neighboring Voronoi cells into account. Milestoning is an idea of Faradjian and Elber [27]. The title of this section is also the title of an article of Vanden-Eijnden and Venturoli [71] about milestoning. In their article, the authors describe an algorithmic approach to compute transition rates between conformations. This algorithm is very similar to the aforementioned adaptive sampling algorithm of Röblitz except that it is based on $\mathcal{Q}$ instead of $\mathcal{P}(\tau)$. The first step of Markovian milestoning is a discretization of Ω into m Voronoi cells $\Phi_1, \ldots, \Phi_m$. For these Voronoi cells an $m \times m$-rate matrix Q^* is computed on the basis of a sampling. In Markovian milestoning, only the restricted horizontal sampling part of the adaptive sampling algorithm is needed. However, an artificial, rapidly mixing sampling scheme is not applicable for this part, because the trajectories have to represent the dynamical model Ψ_τ of the molecular system correctly in order to extract the rates Q^*. Whenever a horizontal sampling trajectory hits an edge of a Voronoi cell, the momenta are reversed. The corresponding state is denoted as a *hitting point*. In a hitting point, the trajectory is reflected at the edge of the cell and thus stays in the same Voronoi cell Φ_i throughout the simulation. Using this kind of restricted sampling, an ergodic dynamical model generates a local Boltzmann distribution of states inside Φ_i. Plotting the histogram of the hitting points for each edge of the Voronoi cells provides a sampling of the Boltzmann distribution of states restricted to these edges. Intuitively, one would expect the edge-restricted Boltzmann distribution to include all relevant information[12] for the rate estimation between cell Φ_i and a neighboring cell Φ_j, i.e. for the matrix element $Q^*(i, j)$. One has to stress that, in general, this intuitive approach is incorrect. This has been shown by Vanden Eijnden et al. [72, 71]. However, in chapter 3.6 we will show that the intuitive approach is correct in a special situation. The key is, one has to understand what the term "incorrect" *means* in this context.

[12] Note that this thesis proposes to use exactly this information in order to approximate a Galerkin discretization Q of $\mathcal{Q}$ in chapter 3.6.

Correct computation of transition rates. In order to define transition rates between conformations, the assumption of the existence of an infinitesimal generator $\mathcal{Q}$ of the semi-group of transfer operators $\mathcal{P}(\tau)$ is important. If Φ is a Voronoi tessellation, a Galerkin discretization Q of this operator $\mathcal{Q}$ is given by

$$Q(i,j) = \frac{\langle \Phi_i, \mathcal{Q}\Phi_j \rangle_{\pi_q}}{\langle \Phi_i, e \rangle_{\pi_q}}. \tag{2.15}$$

Although Q is a Galerkin discretization of the infinitesimal generator $\mathcal{Q}$ of $\mathcal{P}(\tau)$, Q is not the infinitesimal generator of the Galerkin discretization $P(\tau)$ of $\mathcal{P}(\tau)$ defined in (2.6). This is the most important fact to know when using a set-based decomposition approach of Ω in order to derive transition rates. The reason for this "incorrectness" is simple: The set of operators $P(\tau), \tau > 0$, does not have an infinitesimal generator, because the Chapman-Kolmogorov equation does not hold for the discretized operators [60, 77]. In general, $P(\tau + \sigma) \neq P(\tau)P(\sigma)$, for $\tau, \sigma > 0$, because the discretization $P(\tau)$ is based on a projection. Thus, there is no way to find a "correct" transition rate matrix Q, because it does not exist in the set-based approach. Furthermore, there is no way to find the "correct" Markov State Model $P(\tau)$ in the set-based approach, because an iterative application $\big(P(\tau)\big)^k$ of $P(\tau)$ is different from the Galerkin discretization of $\big(\mathcal{P}(\tau)\big)^k$ and different from a Galerkin discretization of $\mathcal{P}(k\,\tau)$, see [80, 76, 47]. This means, the desirable *time-extrapolation*, $\tau \to k\tau$, is not possible in this way. Two different errors can be defined in the framework of transition rate estimation.

Definition 1. *For a given lag-time $\tau > 0$ the* iteration error *is defined as the difference between the Galerkin discretization of $\big(\mathcal{P}(\tau)\big)^k$ denoted as $G\big(\big(\mathcal{P}(\tau)\big)^k\big)$ and the iterative application $\big(P(\tau)\big)^k$ of the Galerkin discretization of $\mathcal{P}(\tau)$ with regard to a suitable matrix norm, i.e. the iteration error is given as $\|G\big(\big(\mathcal{P}(\tau)\big)^k\big) - \big(G\big(\mathcal{P}(\tau)\big)\big)^k\|$.*

Definition 2. *For a given lag-time $\tau > 0$ the* time-extrapolation error *is defined as the difference between the Galerkin discretization of $\big(\mathcal{P}(k\,\tau)\big)$ and the iterative application $\big(P(\tau)\big)^k$ of the Galerkin discretization of $\mathcal{P}(\tau)$ with regard to a suitable matrix norm, i.e. the time-extrapolation error is given as $\|G\big(\mathcal{P}(k\,\tau)\big) - \big(G\big(\mathcal{P}(\tau)\big)\big)^k\|$.*

In the case of an existing infinitesimal generator $\mathcal{Q}$ the equation $\mathcal{P}(k\,\tau) = \big(\mathcal{P}(\tau)\big)^k$ holds. The iteration error and the time-extrapolation error are identical. There exist very different ways to handle these errors in literature:

- In the Markovian milestoning approach, Vanden-Eijnden and Venturoli [71] only count hitting points on the edge between Φ_i and Φ_j if the corresponding trajectory has hit a *different* cell (different from Φ_j) before. These hitting points are denoted as *first hitting points*. The rate matrix Q^* computed on the basis of first hitting points is different from Q. **In Markovian milestoning, the matrix Q^* is not used as an infinitesimal generator.** It is used in order to compute the transition rates $k_{A\to B}$ and $k_{B\to A}$ between two selected Voronoi cells denoted as "conformations" A and B, respectively. Note that in the approach of Vanden-Eijnden

and Venturoli there is a distinction between conformations and transition regions which is not desired in the conformation dynamics approach of Deuflhard and Schütte. The Markovian milestoning can be seen as an efficient analysis of transition regions.

- Sarich, Noé and Schütte give an **error bound [60] for the iteration error.** They conclude that the error is small if the sets Φ_i correlate well with the dominant eigenfunctions of $\mathcal{P}(\tau)$. In chapter 3.3, the conformations χ are defined as a linear combination of the dominant eigenfunctions of $\mathcal{P}(\tau)$. It will be shown that the iteration error theoretically vanishes in this case. The existence of an infinitesimal generator $\mathcal{Q}$ is not a necessary condition in this context.

- In the case of an ergodic dynamical model, the asymptotic result of iteration, $P^{\infty}(\tau)$, is a matrix with identical rows. These rows are given by the vector d (statistical weights of the sets Φ)[4]. The same holds for a Galerkin discretization $P(k\,\tau)$ of $\mathcal{P}(k\,\tau)$ for $k \to \infty$. Thus, asymptotically the time-extrapolation error vanishes. Many researchers test the *Markovianity* of their Markov State Model [34, 12, 69]. They determine if

$$P(2k\,\tau) \approx P(k\,\tau)P(k\,\tau)$$

holds, e.g., by computing the eigenvalues of $P(k\tau)$ and $P(2k\,\tau)$. Having the asymptotical result in mind, it is clear that there is a $k > 0$ for which the approximation is good. The statement in this context is that the **iterative application of $P(\sigma)$ is correct only for lag-times $\sigma \geq k\tau$**, whereas for lag-times $\sigma < k\tau$ it is incorrect. In chapter 3.3 it is shown that a function based approach does not need this concept of an optimal lag-time if an infinitesimal generator $\mathcal{Q}$ exists.

Summarizing, the interpretation of the matrix Q as an infinitesimal generator and the interpretation of $P(\tau)$ as a Markov State Model are not adequate. This is why, in the following, these interpretations will not be used for these matrices any more. It is not appropriate to ignore the continuous nature of $\mathcal{P}$ and $\mathcal{Q}$ by taking P and Q as a starting point of the conformation dynamics analysis of a molecular system. In fact, there is already a rigorous theoretical approach to transition rates taking the *continuous* operator $\mathcal{Q}$ into account – the Transition Path Theory based on a *committor function* approach [49].

2.7 Committor functions

Transition Path Theory (TPT)[23, 50, 83] is a rigorous way to compute transition rates between two conformations A and B. The basis for TPT is the computation of *committor functions* $f_A, f_B : \Omega \to [0, 1]$. The theoretical fundament of committor functions and of committor probabilities has been developed by Onsager [53]. In the first step, the two conformations A and B are defined as (small) subsets of Ω, whereas the rest $\Omega - (A \cup B)$ is the transition region between A and B. In the second step the following boundary value problem is

solved [49]:

$$\begin{cases} \mathcal{Q}f_A = 0, \ \text{ in } \Omega - (A \cup B), \\ f_A = 1, \text{ on } \partial A, \\ f_A = 0, \text{ on } \partial B, \\ f_A := 1, \text{ in } A, \\ f_A := 0, \text{ in } B. \end{cases} \tag{2.16}$$

For solving the boundary value problem, the infinitesimal generator $\mathcal{Q}$ has to be known explicitly. In the case of Smoluchowsky dynamics, $\mathcal{Q}$ is an elliptic partial differential operator [38, 49]. The function $f_B = 1 - f_A$ solves the above equation when changing the roles of A and B. The important statistical information used for TPT can be derived from the committor functions: The function value $f_A(q)$ is the probability that a realization of the dynamical model (according to $\mathcal{Q}$) starting in $q \in \Omega$ reaches set A before it reaches set B. Some properties of this approach do not fit into the conformation dynamics framework given in the Introduction:

(i) The committor function concept is designed for the case of two conformations $n = 2$.

(ii) The conformations are subsets of Ω and have to be pre-defined (they are not a result of the committor function calculation).

(iii) Like in Markovian milestoning, there is a distinction between conformations and transition regions, which is not the case for the desired partition-of-unity decomposition of Ω into conformations proposed by Deuflhard and Schütte.

In order to use the committor function concept in the conformation dynamics framework, some conceptual changes have to be made. The distinction (iii) between conformations and transition regions can be neglected if the sets A and B in (2.16) are replaced by two different points A and B in Ω. In this case, the committor functions f_A and f_B themselves *define* the conformations in a fuzzy sense. f_A and f_B are interpreted as *membership functions*. The partition-of-unity property, $f_A + f_B = e$, holds. For the case of two metastable conformations $n = 2$ and for the case of using points A and B instead of sets, Berezhkovskii and Szabo [6] have shown that (2.16) is approximately solved by a linear combination of the leading two eigenfunctions of $\mathcal{Q}$ corresponding to the eigenvalues $0 = \xi_1 \geq \xi_2 \approx 0$. The leading eigenfunction of $\mathcal{Q}$ is the constant function $e \equiv 1$. The second eigenfunction $X_2 : \Omega \to R$ is assumed to be real and bounded. The linear combination is done in such a way that the constraints of the equation are satisfied:

$$f_A \approx \frac{X_2 - \min_{q\in\Omega} X_2(q)}{\max_{q\in\Omega} X_2(q) - \min_{q\in\Omega} X_2(q)}, \qquad f_B = 1 - f_A. \tag{2.17}$$

It is easy to check that these functions approximately solve the system of equations (2.16) because of the approximately zero eigenvalue of X_2. Note that computing f_A and f_B on the basis of a linear combination of the leading eigenfunctions of $\mathcal{Q}$ also defines(!) the points A and B in (2.17). In this equation, the point A is a maximum of the eigenfunction X_2, while B is a minimum of X_2. This solves the problem of the identification of conformations (ii). The approximate solutions (2.17) of f_A and f_B are exactly the same as the membership

functions χ_1 and χ_2 identified by Robust Perron Cluster Analysis (PCCA+) for the case of two conformations, see Deuflhard and Weber [21, 75],

$$\chi_1 = \frac{X_2 - \min_{q\in\Omega} X_2(q)}{\max_{q\in\Omega} X_2(q) - \min_{q\in\Omega} X_2(q)}, \qquad \chi_2 = 1 - \chi_1. \tag{2.18}$$

Via PCCA+, the generalization (i) of the committor function concept to more than $n = 2$ conformations is simple. For the case of $n > 2$ conformations, the membership functions $\chi_1, \ldots, \chi_n$ are determined as linear combination of the leading n eigenfunctions of $\mathcal{Q}$ such that certain constraints are satisfied. For the concept of Robust Perron Cluster Analysis (PCCA+) see chapter 3.3. Since χ_i can be seen as a membership function of conformation i, these functions themselves will be denoted as conformations in the followings. At this stage, the picture is clear. For all of the aforementioned algorithms, the main subject of conformation dynamics is the analysis of the dominant eigenfunctions of the infinitesimal generator $\mathcal{Q}$. This can be seen as follows:

- Vanden-Eijnden and Venturoli have shown [71] that the restricted horizontal sampling of the Markovian milestoning algorithm is theoretically valid only if successive transitions between the Voronoi cells are statistically independent. This is the case only if Φ provides a good set of basis functions to discretize equation (2.16), because the optimal choice for the Voronoi cells in Markovian milestoning is given by the isocommittor surfaces [71]. Solving equation (2.16) is very similar to computing the dominant eigenfunctions of $\mathcal{Q}$ as shown before.

- Sarich et al. [60] have shown, that a Markov State Model has a small iteration error only if the discretization sets Φ provide a good approximation of the eigenfunctions of $\mathcal{P}$. The eigenfunctions of $\mathcal{Q}$ are identical to the eigenfunctions of $\mathcal{P}(\tau)$.

- Röblitz [58] has introduced an error bound as a stopping criterion of the adaptive sampling approach. The adaptive sampling approach terminates if the dominant subspace of $P(\tau)$ is determined sufficiently. In case of a self-adjoint transfer operator, the dominant eigenfunctions of $\mathcal{Q}$ span the dominant invariant subspace of $\mathcal{P}(\tau)$.

Analyzing the dominant invariant subspace of an infinitesimal generator $\mathcal{Q}$ can be seen as a possible unification of molecular kinetics theories [54]. In the next chapter, the corresponding theory will be derived.

Chapter 3

Design of a generalized Markov operator

In equation (2.7), a generalized transfer operator $\mathcal{P}(\tau) : L^{1,2}_{\pi_q}(\Omega) \to L^{1,2}_{\pi_q}(\Omega)$ has been defined which characterizes the transfer of membership functions $f : \Omega \to [0,1]$ with regard to different dynamical models $\Psi_{-\tau}$. In the above sections, molecular simulation always starts with the selection of a dynamical model. Instead of selecting one special dynamical model in the following, a generalized transfer operator will be designed which meets certain desirable properties. This operator will be discretized. However, realizations of a dynamical model cannot be applied for the computation of this discretization as long as the dynamical model is undefined. Fortunately, we are only interested in the dominant invariant subspace of the operator $\mathcal{P}(\tau)$ which is identical to the dominant invariant subspace of $\mathcal{Q}$. It will be shown that a discretization of $\mathcal{Q}$ can be computed without realizations of a dynamical model.

3.1 Desirable properties

In this chapter two desirable properties are defined: self-adjointness and discrete dominant eigenvalues for the operator $\mathcal{P}$. From a mathematical point of view, these properties assure that a small perturbation of the input data leads to a small perturbation of eigenvalues and eigenvectors. Besides these mathematical advantages, the desired properties can be motivated physically.

Self-adjointness. A very important property of Hamiltonian dynamics is its *time-reversibility*. Changing the sign of the initial momentum vector $p(0)$ and then solving the equations of motion is the same as starting in $(q(0), p(0))$ and going backward in time. In the canonical ensemble, the probability for a momentum vector $p(0)$ is equal to the probability for $-p(0)$, i.e. for every Hamiltonian dynamics trajectory in the canonical ensemble there exists a reverse trajectory. Physically, the Boltzmann density of states in the canonical ensemble is a *detailed balanced* dynamical equilibrium with regard to this dynamical model. This property should be valid for the designed operator $\mathcal{P}$, too. Detailed balance is the first desirable property of $\mathcal{P}$. Thus, the frequency

of transitions $\widetilde{q} \to q$ is equal to the frequency of transitions $q \to \widetilde{q}$ for all pairs of states $(q, \widetilde{q}) \in \Omega$ in the canonical ensemble. This can be expressed by $\pi_q(\widetilde{q}) \cdot P(\widetilde{q} \to q) = \pi_q(q) \cdot P(q \to \widetilde{q})$, where $P(q \to \widetilde{q})$ is the conditional probability density for a transition from q to $\widetilde{q}$. In the general case of $\Psi_{-\tau}$, detailed balance can be written as

$$\pi_q(\widetilde{q}) \cdot \underbrace{\int_{R^{3N}} \Psi_{-\tau}\big(q|(\widetilde{q},p)\big)\, \pi_p(p)\, dp}_{P(\widetilde{q}\to q)} = \pi_q(q) \cdot \underbrace{\int_{R^{3N}} \Psi_{-\tau}\big(\widetilde{q}|(q,p)\big)\, \pi_p(p)\, dp}_{P(q\to\widetilde{q})}. \quad (3.1)$$

Theorem 1. *Given the operator $\mathcal{P}(\tau) : L^2_{\pi_q}(\Omega) \to L^2_{\pi_q}(\Omega)$ defined in (2.7) and the detailed balance condition (3.1), then $\mathcal{P}(\tau)$ is a linear, π_q-self-adjoint operator with $\|\mathcal{P}(\tau)\, f\|_{\pi_q} \le \|f\|_{\pi_q}$. In particular, $\mathcal{P}$ has a real-valued spectrum $\sigma(\mathcal{P})$ with $\sigma(\mathcal{P}) \subset [-1, 1]$.*

Proof. Linearity is easy to check. Self-adjointness is a direct consequence of equation (3.1):

$$\begin{aligned}
\langle g, \mathcal{P}(\tau)\, f\rangle_{\pi_q} &= \int_\Omega g(q)\, \mathcal{P}(\tau)\, f(q)\, \pi_q(q)\, dq \\
&= \int_\Omega g(q) \int_{R^{3N}} \int_\Omega f(\widetilde{q})\, \Psi_{-\tau}\big(\widetilde{q}|(q,p)\big)\, d\widetilde{q}\, \pi_p(p)\; dp\; \pi_q(q)\; dq \\
&= \int_{R^{3N}} \int_\Omega \int_\Omega g(q)\, f(\widetilde{q})\, \Psi_{-\tau}\big(\widetilde{q}|(q,p)\big)\, \pi_q(q)\, \pi_p(p)\; d\widetilde{q}\; dq\; dp \\
&= \int_{R^{3N}} \int_\Omega \int_\Omega g(q)\, f(\widetilde{q})\, \Psi_{-\tau}\big(q|(\widetilde{q},p)\big)\, \pi_q(\widetilde{q})\, \pi_p(p)\; dq\; d\widetilde{q}\; dp \\
&= \int_\Omega \mathcal{P}(\tau)\, g(\widetilde{q})\, f(\widetilde{q})\, \pi_q(\widetilde{q})\, d\widetilde{q} \\
&= \langle \mathcal{P}(\tau)\, g, f\rangle_{\pi_q}.
\end{aligned}$$

In order to prove $\|\mathcal{P}(\tau)\, f\|_{\pi_q} \le \|f\|_{\pi_q}$, note that according to the Cauchy-Schwarz inequality, $(\int_\Omega f(x)\, \pi(x)\, dx)^2 \le \int_\Omega f^2(x)\, \pi(x)\, dx$ for a probability density function $\pi : \Omega \to R$. Thus, the following holds:

$$\begin{aligned}
\|\mathcal{P}(\tau)\, f\|^2_{\pi_q} &= \int_\Omega \big(\mathcal{P}(\tau)\, f(q)\big)^2\, \pi_q(q)\, dq \\
&= \int_\Omega \Big(\int_{R^{3N}} \int_\Omega f(\widetilde{q})\, \Psi_{-\tau}\big(\widetilde{q}|(q,p)\big)\, d\widetilde{q}\, \pi_p(p)\, dp\Big)^2 \pi_q(q)\, dq \\
&\le \int_\Omega \int_{R^{3N}} \Big(\int_\Omega f(\widetilde{q})\, \Psi_{-\tau}\big(\widetilde{q}|(q,p)\big)\, d\widetilde{q}\Big)^2 \pi_q(q)\, \pi_p(p)\, dp\; dq \\
&\le \int_\Omega \int_{R^{3N}} \int_\Omega f^2(\widetilde{q})\, \Psi_{-\tau}\big(\widetilde{q}|(q,p)\big)\, \pi_q(q)\, \pi_p(p)\, d\widetilde{q}\; dp\; dq \\
(*) \quad &= \int_\Omega \int_{R^{3N}} \int_\Omega f^2(\widetilde{q})\, \Psi_{-\tau}\big(q|(\widetilde{q},p)\big)\, \pi_q(\widetilde{q})\, \pi_p(p)\, d\widetilde{q}\; dp\; dq \\
&= \int_\Omega f^2(\widetilde{q})\, \pi_q(\widetilde{q})\, d\widetilde{q} \\
&= \|f\|^2_{\pi_q}.
\end{aligned}$$

The detailed balance condition has been used in $(*)$. The spectral properties of $\mathcal{P}$ are a consequence of self-adjointness and the above estimation. □

Discrete dominant eigenvalues. The second desirable property of a generalized transfer operator $\mathcal{P}(\tau)$ is the existence of a dominant *discrete* spectrum of eigenvalues $1 = \lambda_1 \geq \lambda_2 \geq \lambda_3 \ldots \lambda_n$. The dominant eigenvalue is $\lambda_1 = 1$ because of Theorem 1 and because of $\mathcal{P}(\tau)\, e = e$. The eigenvalues are real according to Theorem 1. A conjecture by Schütte [63] says that in realistic applications of the transfer operator approach, the dominant discrete spectrum of $\mathcal{T}$ is bounded away from the essential spectrum of $\mathcal{T}$. Huisinga [38] has derived properties of a generalized transfer operator such that the essential spectrum is well separated from the dominant eigenvalues near $\lambda_1 = 1$. For the operator $\mathcal{P}(\tau)$ we will assume that there is a set of eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ which are close to $\lambda_1 = 1$ and corresponding eigenvectors $\{X_1, \ldots, X_n\}$. This assumption can be justified by a perturbation approach. The inequality $\|\mathcal{P}(\tau)\, f\|_{\pi_q} \leq \|f\|_{\pi_q}$ in Theorem 1 is sharp. It can be shown that certain indicator functions I_M are eigenfunctions of $\mathcal{P}$ corresponding to the eigenvalue $\lambda = 1$.

Definition 3. *A* stable subset $M \subset \Omega$ *of* $\Psi_{-\tau}$ *is defined as follows: If a position state* q *is in the closure of* M*, then* $\int_{R^{3N}} \Psi_{-\tau}\big(\widetilde{q}|(q,p)\big)\pi_p(p)\, dp = 0$ *for all* $\widetilde{q} \notin M$ *and* $\tau \in R$*.*

Lemma 1. *If* $I_M : \Omega \to \{0,1\}$ *is the indicator function of a stable subset* $M \subset \Omega$*, then* $\mathcal{P}(\tau)\, I_M = I_M$*. In particular, the indicator functions of stable subsets of* Ω *span an eigenspace of* $\mathcal{P}(\tau)$ *according to the dominant eigenvalue* $\lambda = 1$*.*

Proof. If $M \subset \Omega$ is a stable subset, then its complement $\Omega \backslash M$ is also stable. This can be shown by applying the detailed balance condition (3.1) for $\widetilde{q} \notin M$ and $q \in M$. The definition of stability means that there is no transition between M and its complement $\Omega \backslash M$. Stability of M and $\Omega \backslash M$ directly implies $\mathcal{P}(\tau)\, I_M(q) = 0$ for all $q \notin M$, because the integrand

$$I_M(\widetilde{q}) \int_{R^{3N}} \Psi_{-\tau}\big(\widetilde{q}|(q,p)\big)\, \pi_p(p)\, dp$$

is zero for all $\widetilde{q} \in \Omega$. For $q \in M$:

$$\begin{aligned}
\mathcal{P}(\tau)\, I_M(q) &= \int_{R^{3N}} \Big(\int_{\Omega} I_M(\widetilde{q})\ \Psi_{-\tau}\big(\widetilde{q}|(q,p)\big)\, d\widetilde{q} \Big)\, \pi_p(p)\, dp \\
&= \int_{R^{3N}} \Big(\int_{M} I_M(\widetilde{q})\ \Psi_{-\tau}\big(\widetilde{q}|(q,p)\big)\, d\widetilde{q} \Big)\, \pi_p(p)\, dp \\
&= \int_{R^{3N}} \Big(\int_{M} \Psi_{-\tau}\big(\widetilde{q}|(q,p)\big)\, d\widetilde{q} \Big)\, \pi_p(p)\, dp \\
&= 1.
\end{aligned}$$

The last equality uses the fact, that $\Psi_{-\tau}$ is a density function that is normalized in M. □

In the case of *metastable subsets*, a perturbation approach justifies the existence of eigenfunctions X_i with eigenvalues $\lambda_i \approx 1$. The identification of these metastable subsets is an important aim of conformation dynamics. Note that the two mentioned desirable properties are also valid for the transfer operator $\mathcal{T}$. In chapter 3.5 a further desirable property of $\mathcal{P}$ will be given which is not valid for $\mathcal{T}$.

3.2 From a transfer operator to a Markov operator

In this section, we will see that defining a transfer operator $\mathcal{P}(\tau)$ via listing desirable properties also defines a corresponding projected Markov operator $\overline{\mathcal{P}}(\tau)$ acting in configurational space.

Markov operator. $\mathcal{P}(\tau)$ is a transfer operator which acts on membership functions. It is not a desired Markov operator[1] $\overline{\mathcal{P}}(\tau)$ propagating density functions. If $\overline{\mathcal{M}}$ denotes the adjoint operator of $\mathcal{M}$ with regard to the scalar product $\langle f, g\rangle = \int_\Omega f(q)\, g(q)\, dq$, then the following equations hold

$$\begin{aligned} \overline{\mathcal{P}} \circ \pi_q &= \overline{\mathcal{P}} \circ \overline{\pi_q} \\ &= \overline{\pi_q \circ \mathcal{P}} \\ (*) &= \pi_q \circ \mathcal{P}, \end{aligned} \tag{3.2}$$

where π_q is the Nemytskii operator multiplying a function f point-wise with the configurational Boltzmann density function π_q. For the reason of simplicity the same expression π_q is used for the operator and for the function. Equation $(*)$ is valid because of the π_q-self-adjointness of $\mathcal{P}$. From equation (3.2) one can derive an expression for the propagator $\overline{\mathcal{P}}$:

$$\overline{\mathcal{P}} = \pi_q \circ \mathcal{P} \circ \pi_q^{-1}. \tag{3.3}$$

It can be shown that $\overline{\mathcal{P}}(\tau)$ is a linear Markov operator because it is preserving the $\|\cdot\|_1$-norm for a non-negative function f:

$$\begin{aligned} \|\overline{\mathcal{P}}(\tau)\, f\|_1 &= \langle e, |\overline{\mathcal{P}}(\tau)\, f|\rangle \\ (*_1) &= \langle e, \overline{\mathcal{P}}(\tau)\, f\rangle \\ &= \langle e, \mathcal{P}(\tau)\, (\pi_q^{-1}\, f)\rangle_{\pi_q} \\ (*_2) &= \langle \mathcal{P}(\tau)\, e, (\pi_q^{-1}\, f)\rangle_{\pi_q} \\ &= \langle e, (\pi_q^{-1}\, f)\rangle_{\pi_q} \\ &= \langle e, f\rangle = \|f\|_1, \end{aligned}$$

where the positivity of $\overline{\mathcal{P}}$ is used in $(*_1)$ and the π_q-self-adjointness of $\mathcal{P}$ is used in $(*_2)$. Thus, equation (3.3) means that the space of membership functions and the space of density functions is connected by multiplication with π_q. In the upcoming sections, only the operator $\mathcal{P}(\tau)$ will be investigated. Keep in mind that there is a simple relation between $\mathcal{P}$ and $\overline{\mathcal{P}}$. If $\mathcal{P}$ has an eigenfunction $X : \Omega \to R$ corresponding to an eigenvalue λ, then $\overline{\mathcal{P}}$ has an eigenfunction $(\pi_q\, X)$ with eigenvalue λ. Thus, the spectral properties in Theorem 1 are also valid for $\overline{\mathcal{P}}$.

[1] In this thesis a transfer operator and its adjoint Markov operator are defined. In literature often the technical terms *forward* and *backward* operator are used instead.

Markov property. In chapter 2.1, the propagator $\overline{\mathcal{P}}_s$ has been defined in state space, whereas in (3.3) the propagator is only acting in configurational space Ω. Is it sufficient for conformation dynamics to assume a Markov operator with the aforementioned desirable properties in configurational space (and not in state space)? In fact there exists an argument against this simplification: The Markov property of molecular dynamics is valid for all dynamical models, but only in state space. E.g., being a first order differential equation, Hamiltonian dynamics is Markovian in state space. The equivalent Newton dynamics in configurational space is a second order differential equation in q. Given an initial configurational state $q(0)$, it is not possible to predict the future evolution $q(t)$ of the system. Thus, Newton dynamics is not Markovian in configurational space. In the molecular kinetics approach to conformation dynamics, however, the ensemble of states is important and not a single trajectory. For the evaluation of $\mathcal{P}(\tau)$ it is assumed that the momentum states p are equilibrated according to the Boltzmann distribution π_p. Thus, one can predict the future evolution $\rho(t,\cdot)$ of an initial density $\rho(0,\cdot) : \Omega \to R_+$ assuming equilibrated momentum states. Using the molecular kinetics approach, a non-linear dynamical model (like Hamiltonian dynamics) which is only Markovian in state space can therefore be transformed into a linear Markov operator $\overline{\mathcal{P}}$ acting in configurational space. The Markov property does not hold for the dynamical model in configurational space, but the Markov property holds for the evolution of density functions in configurational space assuming equilibrated momentum states.

3.3 Subspace approach

A projection of $\mathcal{P}(\tau)$ to a low-dimensional Markov State Model $P_c(\tau)$ is based on a Galerkin discretization. A Galerkin discretization on the basis of sets is shown in equation (2.6). As mentioned in chapter 2.7, the set-based approach will be replaced by a function-based definition of the conformations. These conformations $\chi_1, \ldots, \chi_n : \Omega \to [0,1]$ are not sets, they are membership functions with values between 0 and 1. In this situation the Galerkin basis functions $\chi = \{\chi_1, \ldots \chi_n\}$ are not orthogonal. A Galerkin discretization of $\mathcal{P}$ has the following form[2]

$$P_c(\tau) = G\big(\mathcal{P}(\tau)\big) := \big(\langle \chi, \chi \rangle_{\pi_q}\big)^{-1} \langle \chi, \mathcal{P}(\tau)\, \chi \rangle_{\pi_q}. \tag{3.4}$$

In equation (3.4), the expression $\langle \chi, \chi \rangle_{\pi_q}$ is an $n \times n$-matrix. The element (i,j) of this matrix is given by the scalar product $\langle \chi_i, \chi_j \rangle_{\pi_q}$. The matrix $\langle \chi, \mathcal{P}(\tau)\, \chi \rangle_{\pi_q}$ is defined analogously. The construction of a Markov State Model $P_c(\tau)$ based on $\mathcal{P}(\tau)$ is an important step in conformation dynamics. Time-extrapolation and complexity reduction are the main intentions, see Figure 2.2 in chapter 2.1. The iteration error of $P_c(\tau)$ is zero, if the Galerkin discretization of $\big(\mathcal{P}(\tau)\big)^k$ is equal to the iteration $\big(P_c(\tau)\big)^k$. In other words, the iteration error is zero if the diagram in Figure 3.1 commutes. As mentioned before, there is an intermediate step between the operator $\mathcal{P}$ and the Markov State Model P_c. This intermediate

[2] In the case that the conformations $\chi_i, i = 1, \ldots, n$, define separated sets, equation (3.4) has the same form like equation (2.6), because in this case $\chi_i^2 = \chi_i$ and $\chi_i \chi_j = 0$ for $i \neq j$.

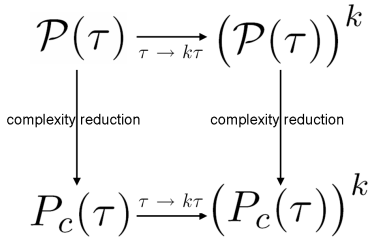

Figure 3.1: If iteration and complexity reduction commute, the iteration error of the Markov State Model $P_c(\tau)$ is zero. With a set-based concept of conformation dynamics this is not possible.

step is a Voronoi tessellation Φ of Ω. The basis extension of χ using the set Φ of basis functions is given by:

$$\chi_i(q) = \sum_{j=1}^{m} \chi_{disc}(i,j)\, \Phi_j(q), \tag{3.5}$$

where χ_{disc} is a $n \times m$-matrix. The element (i,j) of χ_{disc} defines whether the subset Φ_j is part of conformation χ_i. In the early articles about the transfer operator approach, the elements of χ_{disc} were defined as 0 or 1 [19]. In general, this set-based Markov State Model has a non-vanishing iteration error because the diagram in Figure 3.1 does not commute for this choice of χ. However, if we abandon the set-based concept of conformations and define a set of membership functions $\chi_1, \ldots, \chi_n$ which can have values within the interval $[0,1]$, there is a possible Galerkin discretization of $\mathcal{P}(\tau)$ such that the diagram in Figure 3.1 commutes, see also [47] and Theorem 2. Thus, in the recent years the set-based concept was abandoned in favour of a function-based approach. In the first step, elements between 0 and 1 were allowed for the matrix χ_{disc} in (3.5), as long as the conformations still form a partition of unity. This has led to the Robust Perron Cluster Analysis algorithm (PCCA+) by Deuflhard and Weber [21] for the identification of χ_{disc} on the basis of $P(\tau)$. This approach will be described in this section. In the second step, the subspace based discretization of Ω was extended to a basis expansion (3.5) with a set of basis functions Φ_i which form a non-negative partition of unity and have values between 0 and 1, see [75]. The second step was necessary in order to compute the transition matrix P numerically [82]. This numerical trick will be used in chapter 3.6 again. According to the desirable properties of $\mathcal{P}(\tau)$, there is a set of eigenfunctions $X = \{X_1, \ldots, X_n\}$ corresponding to the real eigenvalues $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. $\mathcal{P}$ is π_q-self-adjoint. Thus, the set of eigenfunctions can be assumed to be normalized via $\langle X, X \rangle_{\pi_q} = I$, where I is the $n \times n$ unity matrix.

Theorem 2. *Given the π_q-self-adjoint transfer operator $\mathcal{P}(\tau)$ in the situation of Theorem 1 and a set $X = \{X_1, \ldots, X_n\}$ of normalized eigenfunctions with $\mathcal{P}(\tau)X = X\Lambda$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix of eigenvalues. Furthermore, given a set of functions $\chi = X\mathcal{A}$ that is a linear combination of the eigenvectors X with a regular $n \times n$-transformation matrix $\mathcal{A}$, then the iteration error for the Galerkin discretization $P_c(\tau) = G\big(\mathcal{P}(\tau)\big)$ defined in (3.4) vanishes.*

Proof. By inserting the eigenvalue equation into the Galerkin discretization:

$$\begin{aligned} G\big(\big(\mathcal{P}(\tau)\big)^k\big) &= \big(\langle\chi,\chi\rangle_{\pi_q}\big)^{-1}\langle\chi,\big(\mathcal{P}(\tau)\big)^k\chi\rangle_{\pi_q} \\ &= \big(\mathcal{A}^\top\langle X,X\rangle_{\pi_q}\mathcal{A}\big)^{-1}\mathcal{A}^\top\langle X,\big(\mathcal{P}(\tau)\big)^k X\rangle_{\pi_q}\mathcal{A} \\ &= \big(\mathcal{A}^\top\mathcal{A}\big)^{-1}\mathcal{A}^\top\langle X,X\rangle_{\pi_q}\Lambda^k\mathcal{A} \\ &= \mathcal{A}^{-1}\Lambda^k\mathcal{A}. \end{aligned}$$

In particular, $G\big(\mathcal{P}(\tau)\big) = \mathcal{A}^{-1}\Lambda\mathcal{A}$. Furthermore,

$$\begin{aligned} G\big(\big(\mathcal{P}(\tau)\big)^k\big) &= \mathcal{A}^{-1}\Lambda^k\mathcal{A} \\ &= \big(\mathcal{A}^{-1}\Lambda\mathcal{A}\big)^k \\ &= \big(G\big(\mathcal{P}(\tau)\big)\big)^k. \end{aligned}$$

□

The diagram in Figure 3.1 commutes if the Galerkin discretization of the transfer operator is based on eigenfunctions. This also holds for the corresponding Markov operator $\overline{\mathcal{P}}(\tau)$. In terms of density propagation, the results of the propagated full density compared to the propagated projected density are identical if the projection of the initial density can be done without any error, see Figure 2.2. Let Σ_* be the invariant subspace spanned by the dominant eigenfunctions of $\mathcal{P}$. The initial densities which allow for a correct propagation stem from the function space $\pi_q\Sigma_* = \mathrm{span}(\{\pi_q\chi_1,\ldots,\pi_q\chi_n\})$ spanned by the partially equilibrated, restricted Boltzmann densities.

Robust Perron Cluster Analysis (PCCA+). Theorem 2 holds for arbitrary regular transformations $\mathcal{A}$ of X. However, the theorem only makes sense in the context of conformation dynamics if $\chi = X\mathcal{A}$ can be interpreted as membership functions. PCCA+ (see [21, 75]) determines a transformation matrix $\mathcal{A} \in R^{n\times n}$ such that non-negativity $\chi_i \geq 0, i = 1,\ldots,n$, and the partition-of-unity property, $\sum_{i=1}^n \chi_i(q) = 1, q \in \Omega$, hold. In this situation the transformation matrix $\mathcal{A}$ has a special structure [75]. Note that in standard PCCA+ only finite transition matrices have been investigated so far. In the following, the results of PCCA+ are discussed for the continuous operator $\mathcal{P}(\tau)$ and the corresponding algorithmic change is derived in chapter 3.4. For the purposes of this thesis the following property of $\mathcal{A}$ is important.

Lemma 2. *The partition-of-unity property of the conformations $\chi_1,\ldots,\chi_n$ is equivalent to $\mathcal{A}e = e_1$ for the constant vector $e \in R^n, e(i) \equiv 1$, and the first unit vector $e_1 = (1,0,\ldots,0)^\top \in R^n$.*

Proof. The partition-of-unity relation $\sum_{i=1}^n \chi_i = X_1$ can be expressed by $\chi e = X_1$ which is equivalent to $X\mathcal{A}e = X_1$. For a linear independent set of eigenfunctions this is equivalent to $\mathcal{A}e = e_1$. □

In order to derive properties about the Markov State Model $P_c(\tau)$ a further preparation is needed.

Lemma 3. *Let W be the $n \times n$ diagonal matrix of the statistical weights $w_i := \langle\chi_i, e\rangle_{\pi_q}$ of the conformations, $W = \mathrm{diag}(w_1,\ldots,w_n)$. In this situation the vector $w^\top = (w_1,\ldots,w_n)$ solves the eigenproblems $w^\top = w^\top W^{-1}\langle\chi,\mathcal{P}(\tau)\chi\rangle_{\pi_q}$ and $w^\top = w^\top W^{-1}\langle\chi,\chi\rangle_{\pi_q}$.*

Proof. Note that the matrices $M_1 = \langle \chi, \mathcal{P}(\tau)\chi \rangle_{\pi_q}$ and $M_2 = \langle \chi, \chi \rangle_{\pi_q}$ are symmetric. Note that $w^\top W^{-1} = e^\top$ and that the row sums of the matrices M_1 and M_2 are given by the statistical weights of the conformations due to the partition-of-unity property of the basis functions χ. □

With these preparations the next Theorem shows that the Markov State Model P_c has the expected properties.

Theorem 3. *In the situation of Theorem 2, with a transformation matrix $\mathcal{A}$ defined via PCCA+, the Markov State Model $P_c(\tau)$ has the following properties:*

(i) The row-sum of $P_c(\tau)$ is 1, i.e. $P_c(\tau)e = e$ for the constant vector $e \in R^n, e_i \equiv 1$.

(ii) $P_c(\tau)$ is a Petrov-Galerkin discretization of the Markov operator $\overline{\mathcal{P}}(\tau)$.

(iii) w is an eigenvector of $P_c^\top(\tau)$ corresponding to the eigenvalue $\lambda = 1$ (i.e. $w^\top P_c(\tau) = w^\top$).

(iv) In chapter 3.5, we will assume that $\mathcal{P}$ has an infinitesimal generator $\mathcal{Q}$. If $\mathcal{Q}$ is the infinitesimal generator of the set of operators $\mathcal{P}(\sigma), \sigma > 0$, then P_c has an infinitesimal generator Q_c with $Q_c = G\big(\mathcal{Q}\big)$.

Proof. According to Lemma 2 the relation $\mathcal{A}e = e_1$ holds, where $\mathcal{A}$ is regular. Furthermore, $\lambda_1 = 1$. Thus, the following equations hold $P_c(\tau)\, e = \mathcal{A}^{-1}\Lambda\mathcal{A}e = \mathcal{A}^{-1}\Lambda e_1 = \mathcal{A}^{-1}e_1 = e$. This shows that **(i)** is true. **(ii)**: In terms of the Markov operator $\overline{\mathcal{P}}$ in (3.3), the following equation holds

$$\big(\langle \chi, \chi \rangle_{\pi_q}\big)^{-1} \langle \chi, \mathcal{P}(\tau)\, \chi \rangle_{\pi_q} = \big(\langle \chi, (\pi_q\, \chi) \rangle\big)^{-1} \langle \chi, \overline{\mathcal{P}}(\tau)\, (\pi_q\, \chi) \rangle.$$

Thus, the $\langle \cdot, \cdot \rangle_{\pi_q}$-Galerkin discretization of $\mathcal{P}$ can be seen as a $\langle \cdot, \cdot \rangle$-Petrov-Galerkin discretization of $\overline{\mathcal{P}}$ with membership functions χ in the test space and density functions $(\pi_q\, \chi)$ in the trial space. **(iii)** is a consequence of Lemma 3 because

$$P_c = (W^{-1}\langle \chi, \chi \rangle_{\pi_q})^{-1}(W^{-1}\langle \chi, \mathcal{P}(\tau)\chi \rangle_{\pi_q}).$$

In the situation of **(iv)** an infinitesimal generator $\mathcal{Q}$ of $\mathcal{P}(\sigma)$ exists. The eigenfunctions of $\mathcal{Q}$ are identical to the eigenfunctions of $\mathcal{P}(\tau)$. The corresponding eigenvalues $\xi_1, \ldots, \xi_n$ of $\mathcal{Q}$ meet the property $\exp(\tau\xi_i) = \lambda_i$. A computation similar to the proof of Theorem 2 shows that the Galerkin discretization $G\big(\mathcal{Q}\big)$ of $\mathcal{Q}$ is given by $G\big(\mathcal{Q}\big) = \mathcal{A}^{-1}\Xi\mathcal{A}$, where Ξ is the diagonal matrix of eigenvalues $\Xi = \mathrm{diag}(\xi_1, \ldots, \xi_n)$. Note, that this expression is an infinitesimal generator of $P_c(\sigma)$. □

Optimization problem. The non-negativity property and the partition-of-unity property define a *feasible set* $\mathcal{F} \subset R^{n \times n}$ of transformation matrices $\mathcal{A}$. As shown by Weber [75], this feasible set $\mathcal{F}$ is a convex polytope in $R^{n \times n}$. PCCA+ is formulated as a maximization problem on the set $\mathcal{F}$. The solution of the maximization problem is based on a transformation of the constraint optimization problem in n^2 variables into an unconstraint optimization problem in $(n-1)^2$ variables. This transformation uses the fact, that the objective function of PCCA+ is convex. In this case, the maximum of the objective function is attained at a vertex of $\mathcal{F}$. For $n = 2$ conformations (discarding

permutation of indices) the convex polytope $\mathcal{F}$ has only one vertex leading uniquely to the solution (2.18) of PCCA+. For $n > 2$ conformations, the optimal tranformation matrices $\mathcal{A}$ can depend on the choice of the convex objective function. Two different points of view are discussed in this context in order to define a suitable optimization problem: The interpretation of $P_c(\tau)$ as a transition matrix and the interpretation of χ as a committor function.

$P_c(\tau)$ as a transition matrix. Although $P_c(\tau)$ is the correct Markov State Model, $P_c(\tau)$ in equation (3.4) cannot be interpreted as a transition matrix between conformations. The reason is that is does not have the form of equation (2.6). In equation (2.6) the number of transitions is divided by the weight of the corresponding Voronoi cell. The matrix $W^{-1}\langle\chi, \mathcal{P}(\tau)\chi\rangle_{\pi_q}$ in Lemma 3 has an appropriate form that can be interpreted as a transition matrix. The difference between $P_c(\tau)$ and this matrix is given by

$$\left(W^{-1}\langle\chi,\chi\rangle_{\pi_q}\right) P_c = W^{-1}\langle\chi, \mathcal{P}(\tau)\chi\rangle_{\pi_q}.$$

$S := W^{-1}\langle\chi,\chi\rangle_{\pi_q}$ is a row-stochastic $n \times n$-matrix. The maximal value for the trace of S is given by n. If the trace of S is n, then S is the identity matrix. If the matrix S is the identity matrix, then $P_c(\tau)$ can be interpreted as a transition matrix. Thus, the objective function for the optimal choice of $\mathcal{A}$ is given by the trace of S. Röblitz [58] has introduced this objective function. She has shown that optimizing trace(S) is equivalent to optimizing the *crispness* of the conformations χ. In other words, by optimizing trace(S), the interpretation of χ as sets is also intended. In order to derive an expression for trace(S) in terms of the transformation matrix $\mathcal{A}$, there is a further property of the transformation matrix $\mathcal{A}$ which is a consequence of Theorem 3.

Corollary 1. *Assume that the second eigenvalue of $\mathcal{P}(\tau)$ is not 1, $\lambda_2 \neq 1$. In the situation of PCCA+, the first row of $\mathcal{A}$ is the vector $w^\top$ of the statistical weights of the conformations, i.e. $e_1^\top \mathcal{A} = w^\top$.*

Proof. This corollary is a consequence of Theorem 3 (iii):

$$\begin{aligned} & w^\top P_c = w^\top \\ \Leftrightarrow \quad & w^\top \mathcal{A}^{-1}\Lambda\mathcal{A} = w^\top \\ \Leftrightarrow \quad & w^\top \mathcal{A}^{-1}\Lambda = w^\top \mathcal{A}^{-1}. \end{aligned}$$

The only vector which meets $v^\top\Lambda = v^\top$ in the situation of $\lambda_2 \neq 1$ is the vector $v = e_1$, i.e. $e_1^\top = w^\top\mathcal{A}^{-1}$ and $e_1^\top\mathcal{A} = w^\top$. □

From Corollary 1, the objective function for an interpretation of $P_c(\tau)$ as a transition matrix is given by

$$I_R[\mathcal{A}] = \text{trace}(\text{diag}(\mathcal{A}e_1)^{-1}\mathcal{A}^\top\mathcal{A}) = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\left(\mathcal{A}(j,i)\right)^2}{\mathcal{A}(1,i)}. \tag{3.6}$$

Being a sum of convex functions I_R is convex. For the proof of convexity the positivity of $\mathcal{A}(1,i)$ shown in Corollary 1 is important.

χ as a committor function. In chapter 2.7, comittor functions have been defined for the case of $n = 2$ conformations. The resulting membership functions of PCCA+ for the case of $n > 2$ conformations form a partition-of unity discretization of Ω. In realistic molecular applications of the conformation dynamics approach, mainly one or two different membership functions contribute to this partition-of-unity locally. In transition regions, the configurational space Ω can be seen locally as a transition region between two conformations. This allows for an interpretation of χ as a set of committor functions. In order to interpret χ_i as a committor function, the function χ_i should be an approximate solution of the corresponding equation (2.16). The condition $\mathcal{Q}\chi_i \approx 0$ holds if χ_i is a linear combination of the eigenfunctions of $\mathcal{Q}$ corresponding to eigenvalues $\xi \approx 0$. Furthermore, there should exist a point $q_i \in \Omega$ such that $\chi_i(q_i) = 1$ (see chapter 2.7). This value is the maximum value of χ_i according to the constraints of the optimization problem of PCCA+. The corresponding objective function is given by the sum of the maximum values of the membership functions $\chi_i, i = 1, \ldots, n$. This objective function $I_W[\mathcal{A}] = \sum_{i=1}^{n} \max_{q\in\Omega} \chi_i(q)$ has been introduced by Weber [75]. It is a convex function. Weber has shown that in the case that the optimal function value is $I_W[\mathcal{A}] = n$, the solution of PCCA+ is unique. $\mathcal{F}$ has only one vertex (neglecting the permutation of indices of the conformations). Note, that χ_i need not be an indicator function of a set if its maximum value is 1. Whereas, χ_i must be an indicator function if the corresponding diagonal element (i, i) of S is 1. The following Lemma shows the corresponding relation between I_R and I_W.

Lemma 4. *In the case of PCCA+, for a given solution matrix $\mathcal{A}$ the relation $I_R[\mathcal{A}] \leq I_W[\mathcal{A}]$ holds.*

Proof.

$$\begin{aligned} I_R[\mathcal{A}] &= \sum_{i=1}^{n} \frac{\langle \chi_i, \chi_i \rangle_{\pi_q}}{\langle \chi_i, e \rangle_{\pi_q}} \leq \sum_{i=1}^{n} \frac{\langle \chi_i, \left(\max_{q\in\Omega} \chi_i(q) \right) e \rangle_{\pi_q}}{\langle \chi_i, e \rangle_{\pi_q}} \\ &= \sum_{i=1}^{n} \left(\max_{q\in\Omega} \chi_i(q) \right) \frac{\langle \chi_i, e \rangle_{\pi_q}}{\langle \chi_i, e \rangle_{\pi_q}} = I_W[\mathcal{A}]. \end{aligned}$$

□

Thus, for an optimal transition matrix $\mathcal{A}$ with $I_R[\mathcal{A}] \approx n$ the function value $I_W[\mathcal{A}]$ is also approximately or equal to n. This means that an optimization of I_R is suggested in the following.

3.4 Discretization error

Theorem 2 and 3 hold for $P_c(\tau)$ only if the matrix $\mathcal{A}$ and the eigenvalues Λ are computed without error. Given a set of Voronoi cells Φ, the Galerkin discretization $P(\tau)$ of $\mathcal{P}(\tau)$ can be computed according to an equation of the type (2.6). In a numerical approach, PCCA+ is applied to the discrete eigenvectors and eigenvalues of $P(\tau)$ leading to the discrete conformations χ_{disc}. χ_{disc} is used for a basis expension of χ as shown in (3.5). For a given discretization Φ, it can be assumed that $P(\tau)$ has been estimated in a stable way with controllable small sampling error, e.g. by using adaptive sampling of Röblitz or

of Pande and Singhal-Hinrichs. However, if we have computed the correct discretization $P(\tau)$, there is still a difference between the dominant spectrum Λ_{disc} of $P(\tau)$ and the true spectrum Λ of $\mathcal{P}(\tau)$. There is also a difference between the transformation matrix $\mathcal{A}_{disc}$ for the discrete eigenvectors X_{disc} of $P(\tau)$ and the transformation matrix $\mathcal{A}$ of X. This error is denoted as the *discretization error*. Whereas the true Markov State Model is given by $P_c = \mathcal{A}^{-1}\Lambda\mathcal{A}$, the matrix $\widetilde{P}_c = \mathcal{A}_{disc}^{-1}\Lambda_{disc}\mathcal{A}_{disc}$ is the result of a numerical method which is based on a Voronoi tessellation Φ of Ω. The matrix $\widetilde{P}_c$ depends on the computed eigenvalues Λ_{disc} and on the computed transformation matrix $\mathcal{A}_{disc}$. If we can find an hierarchical adaptive discretization scheme which minimizes the errors $\|\Lambda - \Lambda_{disc}\|$ and $\|\mathcal{A} - \mathcal{A}_{disc}\|$, then the true Markov State Model is approximated by this approach. Does this approach also approximate the metastability of $P_c(\tau)$ and the statistical weights of the conformations? The answer is yes, as we can see in the following.

Discretization error in terms of a metastability. In literature there exist different definitions for the *metastability* of a Markov chain. On the basis of the Markov State Model $P_c(\tau)$, Huisinga [38] defines metastability as the trace of the matrix $P_c(\tau)$. In the case of Theorem 2,

$$\text{trace}(P_c(\tau)) = \text{trace}(\Lambda), \quad \text{trace}(\widetilde{P}_c(\tau)) = \text{trace}(\Lambda_{disc}).$$

Thus, the error in terms of metastability only depends on the approximation error of the eigenvalues and not on the approximation error of the transformation matrix. Weber [75] defines metastability as the determinant of $P_c(\tau)$. The reason for this definition becomes clear if the existence of an infinitesimal generator Q_c of P_c is assumed (see Theorem 3 (iv)):

$$\det(P_c(\tau)) = \det(\exp(\tau\, Q_c)) = \exp(\text{trace}(\tau\, Q_c)) = \prod_{i=1}^{n} \exp(\tau\, Q_c(i,i)).$$

If Q_c is interpreted as an infinitesimal generator of a Markov jump process, then the term $\exp(\tau Q_c(i,i))$ can be seen as the *holding probability* $h_i(\tau)$ [45, 73] of conformation χ_i, i.e. it is the probability that a process starting in χ_i stays in χ_i during the simulation time τ. Thus, the determinant of $P_c(\tau)$ is the product of the holding probabilities. Furthermore,

$$\det(P_c(\tau)) = \det(\Lambda), \quad \det(\widetilde{P}_c(\tau)) = \det(\Lambda_{disc}).$$

Thus, also for the determinant-based definiton of metastability the discretization error only depends on the approximation error of the eigenvalues.

Discretization error in terms of the statistical weights. Corollary 1 has shown that the computation of the first row of $\mathcal{A}$ is equivalent to the determination of the statistical weights of the conformations. Thus, the correct computation of the statistical weights directly depends on the correct approximation of $\mathcal{A}$. We have seen in chapter 2.5 that the computation of statistical weights $d_i, i = 1, \ldots, m$, of the Voronoi cells is ill-conditioned in the case of a decomposition approach. Whenever a metastable sampling scheme is used to determine statistical weights, their computation is ill-conditioned. Assume that

we can compute the weights d_j with controllable small error (see chapter 4.3). In this case, the statistical weights of the conformations in equilibrium are given by $w_{disc}(i) = \sum_{j=1}^{m} \chi_{disc}(i,j)\, d_j$ in the numerical approach.

Controlling the discretization error of the transformation matrix. If the objective function $I_R[\mathcal{A}]$ is used for an optimization of the transformation matrix $\mathcal{A}$, then this objective function is independent from the eigenfunctions and eigenvalues of $\mathcal{P}(\tau)$, see (3.6). The discretization error of $\mathcal{A}$ only depends on the constraints of the optimization problem. In order to understand the influence of the discretization Φ to the construction of $\mathcal{A}_{disc}$ via PCCA+, one has to understand the PCCA+ algorithm. PCCA+ solves the optimization problem with an iterative optimization routine. Given a non-optimal feasible transformation matrix $\mathcal{A}_{disc}^{(s)}$ in step s of the algorithm a feasible transformation matrix $\mathcal{A}_{disc}^{(s+1)}$ is proposed by an unconstraint local optimization step followed by a mapping step. In the optimization step, the elements of an infeasible transformation matrix $\overline{\mathcal{A}}_{disc}^{(s+1)}$ are determined except for the elements of its first row and its first column. In order to derive the next step of the iterative scheme, the infeasible transformation matrix is mapped to a feasible matrix $\mathcal{A}_{disc}^{(s+1)}$. In this mapping $h : \overline{\mathcal{A}}_{disc}^{(s+1)} \mapsto \mathcal{A}_{disc}^{(s+1)}$ the eigenvector data X_{disc} is used in the algorithm. This is the only part of the algorithm where the discretization can influence the transformation matrix $\mathcal{A}_{disc}$. The following three steps are used for the mapping h (see [75]):

1. For $j = 2, \ldots, n$ the element $\overline{\mathcal{A}}_{disc}^{(s+1)}(j,1)$ is defined as $\overline{\mathcal{A}}_{disc}^{(s+1)}(j,1) := -\sum_{i=2}^{n} \overline{\mathcal{A}}_{disc}^{(s+1)}(j,i)$. This is done because of Lemma 2. The row-sums for the rows $j = 2, \ldots, n$ of a feasible transformation matrix are zero. Note that the row-sums are still zero if $\overline{\mathcal{A}}_{disc}^{(s+1)}$ is multiplied with a scalar $\gamma > 0$. This step is independent from the discretization of Ω.

2. For $i = 1, \ldots, n$ the element $\overline{\mathcal{A}}_{disc}^{(s+1)}(1,i)$ is defined as $\overline{\mathcal{A}}_{disc}^{(s+1)}(1,i) := -\min_{l=1,\ldots,m} \sum_{j=2}^{n} \overline{\mathcal{A}}_{disc}^{(s+1)}(j,i)\, X_{disc}(l,i)$. This step is also invariant against positive scaling of $\overline{\mathcal{A}}_{disc}^{(s+1)}$. It assures feasibility of the transformation matrix according to the non-negativity constraint. This step depends on the discretization of Ω.

3. The row-sum of the first row is determined

$$\gamma := \sum_{i=1}^{n} \overline{\mathcal{A}}_{disc}^{(s+1)}(1,i).$$

The matrix $\overline{\mathcal{A}}_{disc}^{(s+1)}$ is multiplied with γ^{-1}:

$$\mathcal{A}_{disc}^{(s+1)} := \gamma^{-1} \overline{\mathcal{A}}_{disc}^{(s+1)}.$$

By this scaling, the row-sum of the first row of $\mathcal{A}_{disc}^{(s+1)}$ is 1. According to Lemma 2, $\mathcal{A}_{disc}^{(s+1)}$ is a feasible transformation matrix.

A PCCA+ algorithm which provides an optimal transformation matrix $\mathcal{A}$ in the continuous case is simply given by a replacement of the term

$$\overline{\mathcal{A}}_{disc}^{(s+1)}(1,i) := -\min_{l=1,\ldots,m} \sum_{j=2}^{n} \overline{\mathcal{A}}_{disc}^{(s+1)}(j,i)\, X_{disc}(l,i)$$

of (2.) by the term[3]

$$\overline{\mathcal{A}}^{(s+1)}(1,i) := -\inf_{q\in\Omega} \sum_{j=2}^{n} \overline{\mathcal{A}}^{(s+1)}(j,i)\, X_i(q).$$

The only difference between the discretized and the continuous algorithm is given by this replacement. Note, that the above mapping h is not a projection in the feasible set $\mathcal{F}$ of transformation matrices. A transformation matrix $\mathcal{A}_{disc}$ is mapped to an edge of $\mathcal{F}$. Convexity of the objective function assures that the optimal value is element of the image of the mapping h. For the computation of the transformation matrix $\mathcal{A}$ in the continuous case it is important to approximate the eigenfunctions X_i of $\mathcal{P}(\tau)$ well.

Controlling the discretization error of the eigenvalues. For the analysis of the discretization error of the eigenvalue matrix Λ_{disc}, one step of an hierarchical refinement is investigated. We will assume that an arbitrary Voronoi cell is refined. Without loss of generality we will assume, that the Voronoi cell Φ_m is decomposed into two new cells $\Phi_m\widetilde{\Phi}_1$ and $\Phi_m\widetilde{\Phi}_2$ according to chapter 2.5. The eigenvalues of a matrix $P(\tau) \in R^{m\times m}$ discretized according to $\Phi = \{\Phi_1, \ldots, \Phi_m\}$ are compared with the eigenvalues of a matrix $P^*(\tau) \in R^{(m+1)\times(m+1)}$ discretized according to $\Phi^* = \{\Phi_1, \ldots, \Phi_m\widetilde{\Phi}_1, \Phi_m\widetilde{\Phi}_2\}$. If this refinement does not change the n dominant eigenvalues, the discretization error vanishes for Φ. The analysis of the discretization error is done in the following way. Solving the eigenvalue problem for $P^*(\tau)$ (i.e. on the finer level) via $P^*X^*_{disc} = X^*_{disc}\Lambda^*_{disc}$, where $X^*_{disc} \in R^{(m+1)\times n}$ and $\Lambda^*_{disc} \in R^{n\times n}$, provides a method for finding a good approximation $\widetilde{X}_{disc}$ of the eigenvectors X_{disc} of $P(\tau)$. Using this approximation, the difference of the eigenvalues of $P(\tau)$ and $P^*(\tau)$ can be estimated. The estimation is based on the following lemma.

Lemma 5. *Let $A \in R^{m\times m}$ be a matrix with real eigenvalues $\lambda_1, \ldots, \lambda_m$ and corresponding eigenvectors $x_1, \ldots, x_m \in R^m$. Further, assume that the eigenvectors are orthonormal with regard to a vector norm $\|\cdot\|$. For a vector $y \in R^m$ with $y \neq 0$ and for a number $\widetilde{\lambda} \in R$ let $\Delta := Ay - \widetilde{\lambda}y$, then*

$$\min_{i=1,\ldots,m} |\lambda_i - \widetilde{\lambda}| \leq \frac{\|\Delta\|}{\|y\|}.$$

Proof. This result is well known, it is a simple computation of the basis extension of Δ in terms of the normalized vectors x_i. □

Let $d^*_m := \langle \Phi_m\widetilde{\Phi}_1, e\rangle_{\pi_q}$ and $d^*_{m+1} := \langle \Phi_m\widetilde{\Phi}_2, e\rangle_{\pi_q}$ be the weights of the new basis functions. With $\alpha := d^*_m/(d^*_m + d^*_{m+1})$ the elements of the last row and the last column of the coarse matrix $P(\tau)$ can be computed via

$$P(\tau)(m,k) \quad = \quad \alpha\, P^*(\tau)(m,k) + (1-\alpha)\, P^*(\tau)(m+1,k),$$

[3]In the following it is assumed that the eigenfunctions of $\mathcal{P}(\tau)$ are bounded.

$$P(\tau)(k,m) \quad = \quad P^*(\tau)(k,m) + P^*(\tau)(k,m+1),$$

for $k = 1, \ldots, m$. This is a consequence of (2.6) and $\Phi_m = \Phi_m \widetilde{\Phi}_1 + \Phi_m \widetilde{\Phi}_2$. The relation can be written in matrix form as $P(\tau) = EP^*(\tau)F$ with matrices $E \in R^{m\times(m+1)}$ and $F \in R^{(m+1)\times m}$. For $i = 1, \ldots, m$, the matrix elements are $E(i,i) = F(i,i) = 1$. Furthermore, $E(m,m) = \alpha$, $E(m,m+1) = 1-\alpha$, and $F(m,m) = F(m+1,m) = 1$. All remaining elements of E and F are zero. For the approximation of the eigenvectors, $\widetilde{X}_{disc} := EX^*_{disc}$ is defined similar to y in Lemma 5. The following equations hold

$$\begin{aligned} P(\tau)\widetilde{X}_{disc} &= EP^*(\tau)FEX^*_{disc} = EP^*(\tau)\cdot\big(X^*_{disc} + \underbrace{(FE-I)}_{=:G}X^*_{disc}\big) \\ &= EP^*(\tau)X^*_{disc} + EP^*(\tau)GX^*_{disc} = \widetilde{X}_{disc}\Lambda^*_{disc} + EP^*(\tau)GX^*_{disc}. \end{aligned}$$

Equivalently,

$$P(\tau)\widetilde{X}_{disc} - \widetilde{X}_{disc}\Lambda^*_{disc} = EP^*(\tau)GX^*_{disc}.$$

According to Lemma 5 the difference between the eigenvalues Λ_{disc} of $P(\tau)$ and the eigenvalues Λ^*_{disc} of $P^*(\tau)$ is in the order of $\|EP^*(\tau)GX^*_{disc}\|$. This norm is small if $(P^*(\tau)G)$ approximately vanishes. A short calculation shows that all non-vanishing elements of this matrix have the form

$$\pm\big((1-\alpha)\,P^*(\tau)(k,m) - \alpha\,P^*(\tau)(k,m+1)\big), \quad k = 1, \ldots, m+1.$$

Thus, the matrix $(P^*(\tau)G)$ would be zero if either

(P1) $P^*(\tau)(k,m+1)$ and $P^*(\tau)(k,m)$ are zero or

(P2)
$$\frac{P^*(\tau)(k,m)}{P^*(\tau)(k,m+1)} = \frac{d^*_m}{d^*_{m+1}}, \quad \forall k = 1, \ldots, m+1.$$

In order to find a good decomposition of Ω on this basis, assume a discretization $\Phi = \{\Phi_1, \ldots, \Phi_m\}$. Furthermore, assume that every Voronoi cell $\Phi_i, i = 1, \ldots, m$ has one of the following properties:

1) **The statistical weight of Φ_i is small.** The conditional probability for a transition from any set $\Phi_k, k = 1, \ldots, m$, into this set Φ_i in time τ is neglectable small. This means that for a further refinement of Φ_i property P1 is valid.

2) **Inside the Voronoi cell Φ_i the dynamics is rapidly mixing.** Assume a realization of the dynamical model $\Psi_{-\tau}$ starting in any Voronoi cell Φ_k having an end point q in cell Φ_i. Because of the rapidly mixing dynamics inside Φ_i, the probability for a certain end point q is approximately given by the restricted Boltzmann distribution proportional to $\pi_q\Phi_i(q)$. A further refinement of Φ will have the property P2: The probability to end in a certain subset of Φ_i is given by the relative statistical weight of this subset.

Thus, for such a kind of decomposition the discretization error of the eigenvalues is small. If $P(\tau)$ is used, Voronoi cells Φ_i with a small weight d_i need not be refined in an adaptive hierarchical scheme in order to approximate Λ.

3.5 An infinitesimal generator

In chapter 3.3 it has been shown that the iteration error for a transfer operator having the aforementioned properties is vanishing for an appropriate selection of χ. Thus, iterative application of $P_c(\tau)$ in terms of the diagram in Figure 3.1 provides the correct transition pattern of the system. In general, an iterative application of $\mathcal{P}(\tau)$ is not a time-extrapolation in terms of the dynamical model which is in the "heart" of $\mathcal{P}$. To give an example: The two aforementioned desirable properties, self-adjointness and discrete dominant eigenvalues, also hold for the transfer operator $\mathcal{T}(\tau)$ in (2.4) which is based on Hamiltonian dynamics. According to chapter 3.2, there is the corresponding Markov operator $\overline{\mathcal{T}}(\tau)$ propagating configurational density functions. Given an initial density function $\rho(0,\cdot) : \Omega \to R_+$, the propagated density is $\rho(\tau,\cdot) = \overline{\mathcal{T}}(\tau)\,\rho(0,\cdot)$. For this propagation an equilibrated density π_p of initial momentum states is assumed. At time τ the initial equilibrium distribution of momentum variables is not preserved by Hamiltonian dynamics if the initial density ρ is different from π_q. Whereas, an iterative application of $\overline{\mathcal{T}}(\tau)$ always assumes an equilibrated density of momentum variables at the starting point. In general,

$$\big(\overline{\mathcal{T}}(\tau)\big)^2\,\rho(0,\cdot) \neq \overline{\mathcal{T}}(2\tau)\,\rho(0,\cdot)$$

because the iterative application of $\overline{\mathcal{T}}$ is not a continuation of the Hamiltonian dynamics (in particular it is not a continuation in momentum space). The Chapman-Kolmogorov equation does not hold for the set of operators $\mathcal{T}(\tau), \tau > 0$. Although for a special choice of conformations the iteration error vanishes (see Theorem 2), the time-extrapolation error is not zero, because $\big(\mathcal{T}(\tau)\big)^k \neq \mathcal{T}(k\,\tau)$. In this strict sense: time-extrapolation, $\tau \to k\tau$, is not possible in conformation dynamics by using a discretization $G\big(\mathcal{T}(\tau)\big)$ as a Markov State Model if the Chapman-Kolmogorov equation does not hold. The Markov property of a dynamical model (like Hamiltonian dynamics) is lost when going from state space to configurational space. The configurational space based transfer operator $\mathcal{T}$ has not lost its Markov property, it has lost its *Chapman-Kolmogorov property*. For the upcoming sections, the Chapman-Kolmogorov property is added to the desirable properties of $\mathcal{P}$ which allows for a time-extrapolation on the basis of transition rates. In this case an infinitesimal generator $\mathcal{Q}$ exists. The Smoluchowsky dynamics is one possible(!) dynamical model which leads to a π_q-self-adjoint transfer operator having an infinitesimal generator $\mathcal{Q}$ [38]. In the followings, some optional approaches are listed which can circumvent the assumption of a Chapman-Kolmogorov property. From a mathematical point of view these options can only lead to approximate time-extrapolations.

Option 1. As an operator acting on the space of membership functions $\mathcal{T}(\tau) : L^2_{\pi_q} \to L^2_{\pi_q}$ does not meet the Chapman-Kolmogorov property, but as an operator $\mathcal{T}(\tau) : \{e\} \to \{e\}$ it meets this property for obvious reasons. In terms of the Markov operator $\overline{\mathcal{T}}(\tau)$ this simply means that the Boltzmann density π_q is invariant with regard to Hamiltonian dynamics (and with regard to any of the aforementioned dynamical models) for any lag-time τ. The restriction to π_q means that an iterative application of the operator $\overline{\mathcal{T}}(\tau)$ is a correct continuation of Hamiltonian dynamics, because at any time $\tau > 0$ the momentum

states are distributed according to π_p. In the case of a Boltzmann equilibrated ensemble of initial states, time-extrapolation is trivial because the equilibrated density is the stationary density. The argumentation is still valid, if we go from $\{e\}$ to the space $\Sigma = \text{span}(\{I_M; M \,\text{is stable}\})$ spanned by the indicator functions I_M of the stable subsets M of Ω. Approximately, it also holds for a space of metastable membership functions $\Sigma_* = \text{span}(\{\chi_1, \ldots, \chi_n\})$, because $\mathcal{T}(\tau)\chi \approx \chi$ for all lag-times τ which are not too large. In the space Σ_*, the Chapman-Kolmogorov property is approximately valid for $\mathcal{T}(\tau) : \Sigma_* \to \Sigma_*$. Consequently, transition rates between conformations exist, which allows for time-extrapolation. Although the transfer operator $\mathcal{T}$ does not have an infinitesimal generator, rates between the conformations (and only between the conformations) can be defined.

This is a mathematical description of Option 1. The physical interpretation is as follows. The conformation dynamics framework of Deuflhard and Schütte does not distinguish between conformations and transition regions. A transition from one conformation χ_i to a different conformation χ_j can be considered as instantaneous. The assumption $\Sigma \approx \Sigma_*$ means that the conformations are like traps. Transitions between the conformations are rare, instantaneous events. In contrast to that, self-equilibration within the conformations is a rapidly mixing process. Thus, we can consider the approach as a computation of rates between always self-equilibrated subsystems of Ω. This point of view is the basis of a very efficient parallel sampling algorithm of Voter [73]: If rates between conformations exist, then holding probabilities h_i can be defined and related to the diagonal elements of Q_c, $h_i(t) = \exp(t\, Q_c(i,i))$, see also page 35. These holding probabilities of the conformations can be determined with the parallel sampling approach of Voter. In contrast to Theorem 2, Voter defines conformations as sets. This is necessary, because holding probabilities can only be defined rigorously if they are based on sets.

Option 2. Researchers testing Markovianity [34, 12, 69] claim that the Chapman-Kolmogorov property for a transfer operator $\mathcal{P}_1$ is violated only for small lag-times. $\mathcal{P}_1$ can be any transfer operator meeting the desirable properties in chapter 3.1 if it corresponds to an *ergodic* dynamical model. In the case of an ergodic dynamical model, there is an optimal lag-time σ: Some observables $O(\mathcal{P}_1(\tau))$ of $\mathcal{P}_1$ (e.g. the dominant eigenvalues) meet $O(\mathcal{P}_1(k\,\tau)) \approx O((\mathcal{P}_1(\tau))^k)$ if these observables are tested only for transfer operators $\mathcal{P}_1(\tau)$ with $\tau > \sigma$. For these observables, the Chapman-Kolmogorov relation approximately holds. In molecular applications, this optimal lag-time σ is much smaller than the slowest time-scale of the system. The transfer operator $\mathcal{P}_1(\sigma)$ defines metastable conformations. On the basis of the π_q-self-adjoint transfer operator $\mathcal{P}_1(\sigma)$ having a discrete dominant spectrum and a fixed lag-time σ, a possible transfer operator $\mathcal{P}_2(\tau)$ which meets the Chapman-Kolmogorov property can be defined for arbitrary lag-times $\tau > 0$ as

$$\mathcal{P}_2(\tau) := \exp(\frac{\tau}{\sigma}\,(\mathcal{P}_1(\sigma) - \mathcal{I})).$$

If $\mathcal{P}_2$ is defined in this way it is π_q-self-adjoint and has a discrete dominant spectrum of eigenvalues. The eigenvectors of $\mathcal{P}_1(\sigma)$ and $\mathcal{P}_2$ are identical. Furthermore, the dominant eigenvalues of $\mathcal{P}_2(\sigma)$ and $\mathcal{P}_1(\sigma)$ are almost identical, because

$\exp(\lambda - 1) \doteq \lambda$ for $\lambda \approx 1$. From this point of view $\mathcal{P}_2(\sigma)$ is not only an alternative transfer operator (instead of $\mathcal{P}_1(\sigma)$) which meets the Chapman-Kolmogorov property, it also approximates the Markov State Model $G\big(\mathcal{P}_1(\sigma)\big) \approx G\big(\mathcal{P}_2(\sigma)\big)$ very well in the situation of Theorem 2. The reason is that the dominant eigenfunctions and eigenvalues of $\mathcal{P}_1(\sigma)$ and $\mathcal{P}_2(\sigma)$ are almost identical. Note that $G\big(\mathcal{P}_2(\sigma)\big)$ has a vanishing iteration error in Theorem 2. It has also a vanishing time-extrapolation error because $\big(\mathcal{P}_2(\sigma)\big)^k = \mathcal{P}_2(k\,\sigma)$. Summarizing, $\mathcal{P}_2(\sigma)$ meets all desirable properties and has a vanishing time-extrapolation error. From this insight one can conclude that the approximation $G\big(\mathcal{P}_1(\sigma)\big) \approx G\big(\mathcal{P}_2(\sigma)\big)$ is an appropriate Markov State Model for time-extrapolation $\sigma \to k\sigma$, too. This seems to be a good justification for testing Markovianity and generating Markov State Models for optimal lag-times σ. However, note that besides the dominant spectral properties, the transfer operators, $\mathcal{P}_1(\sigma)$ and $\mathcal{P}_2(\sigma)$, are very different.

Physically, $\mathcal{P}_1$ is constructed on the basis of a dynamical model. What kind of dynamical model is time-extrapolated by computing $G\big(\mathcal{P}_2(\sigma)\big)^k$? The infinitesimal generator $\mathcal{Q}^*$ of $\mathcal{P}_2$ is

$$\mathcal{Q}^* = \frac{\mathcal{P}_1(\sigma) - \mathcal{I}}{\sigma}.$$

This equation looks like an approximation of (2.12) using a finite difference (but with a large value of σ instead of a small one). Whereas Option 1 uses a short-time argument (for short lag-times the metastable subsets are traps), Option 2 uses a long-term argument (for large lag-times Markovianity holds). For the reason of a large lag-time, $\sigma \gg 0$, any discretization of $\mathcal{Q}^*$ based on Voronoi cells leads to non-vanishing "transition rates" $Q^*(i,j)$ between non-neighboring cells. This is unphysical, because instantaneous transitions between non-neighboring Voronoi cells do not exist if we consider a time-continuous dynamics trajectory in the "heart" of $\mathcal{P}_2$. Although this is a strong argument against this alternative approach, one can find examples for the construction of transition rates between non-neighboring sets in literature. One famous example is Markovian milestoning in chapter 2.6 which defines transition rates between two Voronoi cells A and B which are separated in Ω by a transition region. From a mathematical point of view, the assumption of transition rates between non-neighboring subsets of Ω allows for a valid time-extrapolation even if an infinitesimal generator does not exist. There is an important lesson to learn from Option 2. Two very different transfer operators $\mathcal{P}_1(\tau)$ and $\mathcal{P}_2(\tau)$ can have the same Galerkin discretization in Theorem 2, i.e. $G\big(\mathcal{P}_1(\tau)\big) = G\big(\mathcal{P}_2(\tau)\big)$. The operators can, therefore, lead to the same Markov State Model $P_c(\tau)$ of the molecular system. In the framework of instantaneous transitions between the conformations, this means that it is not so important to know how the dynamical model acts *within* the conformations. It is only important to know how it interacts *between* the conformations. This special aspect is used in a very important part of this thesis when exploiting Gauß' Theorem in chapter 3.6.

Option 3. Option 3 provides a possible interpretation of the results of the algorithm given in chapter 4 if it is assumed that $\mathcal{P}$ does not have an infinitesimal generator. For example, the Markov operator $\overline{\mathcal{T}}$ does not have an infinitesimal generator. The reason is the projection from state space to configurational

space. For a propagation of densities via Hamiltonian dynamics, there exists an infinitesimal generator in state space [30, 38] – the classical Liouville operator $\overline{\mathcal{L}}$. One could have derived the aforementioned results on the basis of a state space instead of using the projection to configurational space. In this case, Theorem 2 would be valid if the conformations χ_i are linear combinations of the dominant eigenfunctions $X_i : \Omega \times R^{3N} \to R$ of $\mathcal{L}$. For these state space based conformations, time-extrapolation would be possible because iteration error and time-extrapolation error would be zero. The approximation of Option 3 is that there exist[4] eigenfunctions $X_i(q,p) \equiv X_i(q)$ for all $p \in R^{3N}$ and $i = 1, \ldots, n$, i.e. there is a set of dominant eigenfunctions X_i of $\mathcal{L}$ which are independent from the momentum state. In order to approximate these eigenfunctions using a discretization approach only a decomposition of Ω is needed.

From a physical point of view, Option 3 assumes that the time-scale of transitions between conformations is much longer than the time-scale of momentum state equilibration. Note that conformations are defined in configurational space. The assumption of fast momentum state equilibration is a physical basis for all conformation dynamics algorithms.

3.6 Application of Gauß' theorem

For the algorithmic realization of the aforementioned theoretical approach, an adaptive hierarchical scheme will be applied. Recall the adaptive sampling algorithm of Röblitz in chapter 2.5. In Figure 3.2, the situation of a one-dimensional configurational space is shown. Ω is decomposed into Voronoi cells. One of it,

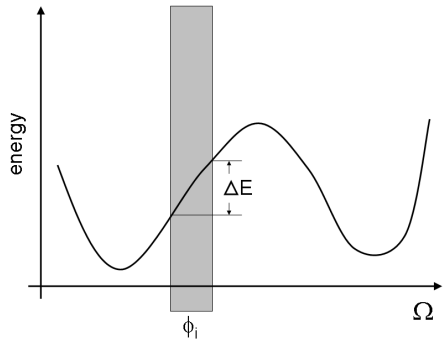

Figure 3.2: Voronoi tessellation of a one-dimensional configurational space Ω. A Voronoi cell Φ_i is an interval of Ω. The energy difference between the left and the right border of Φ_i determines the ratio between the transition frequency.

Φ_i, is shown in the figure. Assuming, that the restricted Boltzmann distribution according to Φ_i has already been sampled (horizontal sampling for Φ_i) and one wants to determine the transition behavior from Φ_i to the other subsets of Ω

[4]For Hamiltonian dynamics in state space, there exist *stable* subsets of $\Omega \times R^{3N}$. Every connected component of the preimage (with regard to the total energy H) of an interval in R is a stable subset of the state space. These subsets are not p-independent. The assumption is that there exist metastable subsets which are nearly p-independent.

in a vertical sampling. In this case, the dynamical process $\Psi_{-\tau}$ mainly leaves the cell Φ_i by crossing the left border rather than by crossing the right one. According to Boltzmann, the ratio r of observing a state at the left border with regard to observing a state at the right border is $r = \exp(-\beta\,\Delta E)$. The ratio exponentially decreases with ΔE. For $\Delta E = 15\,\mathrm{kJ/mol}$ and a temperature of $T = 300\,\mathrm{K}$, the ratio is $r \approx 0.002$. When starting M vertical samplings, only a small portion of $(M \cdot r)$ realizations of $\Psi_{-\tau}$ are expected to cross the right border. For a good statistics, an order of $O(1/r)$ vertical samplings is necessary, see Table 4.1 on page 58. This leads to a slowly converging sampling statistics for the transition matrix $\widetilde{P}$ if barriers are too high. A slow convergence rate implies further refinements in the adaptive sampling scheme of Röblitz. This might lead to a curse of dimensionality. However, if the matrix Q is computed in the adaptive approach (instead of the transition matrix $P(\tau)$), then it is sufficient to generate the horizontal sampling. This can be seen by the application of Gauß' theorem.

Theorem 4. *In the situation of chapter 3.5 the existence of an infinitesimal generator of the transfer operator $\mathcal{P}$ is assumed. $\mathcal{P}$ meets the desirable properties of chapter 3.1. Furthermore, assume a decomposition of the conformational space into a set of Voronoi cells $\Phi_1, \ldots, \Phi_m$ leading to a discretization $P(\tau)$ of $\mathcal{P}(\tau)$. In this situation, $Q(i,j) = \partial/\partial\tau\, P(\tau)(i,j)$ is given by a surface integral[5] of the form*

$$Q(i,j) = \oint_{\partial\Phi_i\Phi_j} z(q)\,\pi_i(q)\,dS, \tag{3.7}$$

where $\partial\Phi_i\Phi_j$ is the intersecting surface of Voronoi cell Φ_i and Voronoi cell Φ_j.

Note that Q is not an infinitesimal generator of $P(\tau), \tau > 0$. The elements of Q are not rates! This has been shown in chapter 2.6. Q is simply the Galerkin discretization of $\mathcal{Q}$ on the basis of a set of functions Φ. In (3.7)

$$z(q) = \int_{R^{3N}} \max\{\langle \bar{v}(q,p), \nu\rangle, 0\}\,\pi(p)\,dp,$$

where $\langle\cdot,\nu\rangle$ is the scalar product with the surface normal vector ν of the intersecting surface $\partial\Phi_i\Phi_j$. $z(q)$ is a result of a mean value computation for the set of momentum variables p distributed according to $\pi_p(p)$ and for the mean velocity vector $\bar{v}(q,p)$ depending on q and p. Equation (3.7) shows that the computation of transition rates is only based on information about the flux through certain surfaces. This is an important complexity reduction of molecular kinetics.

Proof. $\Psi_{-\tau}(\widetilde{q};(q,p))$ in (2.7) is a probability density function. Using the continuity equation [13], it can be shown that

$$\begin{aligned}
\frac{\partial}{\partial\tau}\Big|_{\tau=0} \Psi_{-\tau}(\widetilde{q};(q,p)) &= -\frac{\partial}{\partial\tau}\Big|_{\tau=0} \Psi_{\tau}(\widetilde{q};(q,p)) \\
&= \operatorname{div}\big(\Psi_0(\widetilde{q};(q,p)) \cdot \dot{q}\big) \\
&= \operatorname{div}\big(\delta_{q=\tilde{q}} \cdot \dot{q}\big).
\end{aligned} \tag{3.8}$$

[5] In this thesis, the expression $\oint$ is used in order to distinguish between volume and surface integrals. It is not a circular integral.

In (3.8) the term $\dot{q}$ denotes the velocity vector of the probability density function in the starting point (q,p) and $\delta_{q=\tilde{q}}$ is a Dirac delta function. From the Markovianity assumption given in the introduction, the velocity vector $\dot{q} = \bar{v}(q,p)$ only depends on the initial state (q,p). It can be an averaged vector of many realizations $v(q,p)$ for stochastic dynamical models. Let the restricted Boltzmann density inside the Voronoi cell Φ_i be denoted as

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

With this definition and (3.8) it can be shown that for $Q(i,j) = \partial/\partial\tau\, P(\tau)(i,j)$, with $i \neq j$, the following equations hold

$$\begin{aligned}
\frac{\partial}{\partial\tau}\Big|_{\tau=0} P(\tau)(i,j) &= \int_\Omega \int_{R^{3N}} \Big[\int_\Omega \Phi_j(\widetilde{q})\, \mathrm{div}\big(\delta_{q=\tilde{q}} \cdot \bar{v}(q,p)\big)\, d\widetilde{q} \Big]\, \pi_i(q)\,\pi(p)\,dp\,dq \\
(*^1) &= \int_\Omega \int_{R^{3N}} \Big[\int_{\Phi_j} \mathrm{div}\big(\delta_{q=\tilde{q}} \cdot \bar{v}(q,p)\big)\, d\widetilde{q} \Big]\, \pi_i(q)\,\pi(p)\,dp\,dq \\
(*^2) &= \int_\Omega \int_{R^{3N}} \Big[\oint_{\partial\Phi_j} \langle \delta_{q=\tilde{q}} \cdot \bar{v}(q,p), \nu \rangle\, dS \Big]\, \pi_i(q)\,\pi(p)\,dp\,dq \\
(*^3) &= \oint_{\partial\Phi_i\Phi_j} \underbrace{\int_{R^{3N}} \max\{\langle \bar{v}(q,p), \nu\rangle, 0\}\,\pi(p)\,dp}_{=:\, z(q)}\, \pi_i(q)\,dS. \qquad (3.9)
\end{aligned}$$

For $(*^1)$ it is used that $\Phi_j(q)$ is the indicator function of a Voronoi cell which is also denoted as Φ_j. $(*^2)$ is a consequence of Gauß' Theorem (divergence theorem [28]), where $\langle\cdot,\nu\rangle$ is the scalar product with the surface normal vector ν of the surface $\partial\Phi_j$ and $\oint_{\partial\Phi_j} dS$ is the corresponding surface integral. From $(*^2)$ to $(*^3)$ it is used, that the sets Φ_i and Φ_j have an intersecting interface and that only trajectories going from Φ_i to Φ_j can contribute to instantaneous transitions through this interface (positive sign of the scalar product with the surface normal vector ν). □

Chapter 4

Algorithmic Details

The algorithmic realization of the aforementioned theoretical results has to tackle some problems. An adaptive, meshless Galerkin discretization of a high-dimensional configurational space is needed. Furthermore, an efficient estimation of statistical weights is desired for the requirement R1 of the Introduction. Some algorithmic details concerning these aspects are discussed in this chapter.

4.1 Meshless basis functions

In high-dimensional cases a box discretization of Ω is not appropriate. Instead of using a box discretization, Voronoi cells are suitable for high-dimensional systems.

Simplification by neglecting fast degrees of freedom. For the construction of Voronoi cells, the definition of a distance function $\mathrm{dist} : \Omega \times \Omega \to R$ is required, see (2.8). From a mathematical point of view this distance function should be zero, $\mathrm{dist}(q, q^*) = 0$, only if two configurations are identical $q = q^*$. However, in realistic applications this would lead to an unnecessary combinatorial blow-up of the configurational space. Generated statistical information would never be sufficient for the evaluation of Q. Pre-knowledge about the molecular system should be incorporated into the distance function computation. The conformations $\chi_i(q)$ identify parts of configurational space which can be described by a much longer time-scale than the fast degrees of freedom of the dynamical model. Therefore, an efficient distance computation would provide $\mathrm{dist}(q, q^*) = 0$, if q and q^* are configurations which are connected by a rapidly mixing dynamical process. When discretizing Ω, the distance function can account for these fast degrees of freedom and exclude them from the computation of $\mathrm{dist}(q, q^*)$.

- **Rotation and translation.** There are many cases in which rotational and translational degrees of freedom of the whole system are not relevant for the time-scale of the chemical process to be investigated (e.g., protein folding). In this case, Kabsch [43, 44] has formulated an algorithm which can be used in order to align two configurational states q and q^* before computing their distance in configurational space.

- **Neglecting atoms.** The easiest way to simplify Voronoi cells in configurational space is to neglect atoms for the computation of the distances. Whereas, all atoms are regarded for the computation of energies, velocities etc. the definition of eigenfunctions is restricted to a subspace of Ω (this approach is comparable to neglecting the momentum coordinates in Option 3 in chapter 3.5). When neglecting atoms, it should be done in such a way that conformational changes can still be identified only on the basis of the observed atoms. It is not important to include all atoms that influence(!) the conformational change in the computation of the distance function. It is only important to include all atoms that allow for the identification(!) of conformational changes. Thus, it can be appropriate to neglect all explicit water molecules from distance computation if the conformational changes of the system can be observed in the remaining degrees of freedom. Although, the water molecules have an important influence in the chemical process, the conformational change can be seen in the movement of the remaining atoms.

- **Internal coordinates.** Instead of taking all cartesian coordinates of the observed atoms, one can restrict the point of view to a small set of internal coordinates which indicates conformational changes. This method can only be used if enough pre-knowledge about the molecular system has been collected. Bond lengths and bond angles are often preserved in a molecular simulation of high-dimensional molecular models. Therefore, the restriction to torsion angles may be appropriate. This reduction method should only be applied if possible conformational changes are known in advance.

Fast degrees of freedom can be excluded from the configurational space Ω with a combination of the aforementioned simplifications of the distance computation in Ω. The binding path simulation in chapter 5.2.2 has been identified with a distance function based on a set of internal coordinates.

Restricted sampling. The numerical computation of the matrix Q can be done by restricted samplings. Expanding equation (3.7) with the term

$$\oint_{\partial\Phi_i\Phi_j} \pi_q(q)\, dS$$

leads to:

$$Q(i,j) = \frac{\oint_{\partial\Phi_i\Phi_j} \pi_q(q)\, dS}{\int_{\Phi_i} \pi_q(q)\, dq} \cdot \oint_{\partial\Phi_i\Phi_j} z(q)\, \frac{\pi_q(q)}{\oint_{\partial\Phi_i\Phi_j} \pi_q(q)\, dS}\, dS = \frac{s_{ij}}{d_i} \cdot \langle z \rangle_{\partial\Phi_i\Phi_j}, \tag{4.1}$$

where s_{ij} is *the weight of the surface* $\partial\Phi_i\Phi_j$ and d_i is the weight of the Voronoi cell Φ_i. In chapter 4.3 it will be shown, that these two quantities can be estimated by a restricted sampling approach (restricted to Voronoi cells and restricted to the edges of Voronoi cells). In an algorithmic realization of the computation of (4.1), the ratio between the surface measure dS and the volume measure dq is unknown in general. This leads to a scaling factor μ of Q in (4.1), i.e. Q can be computed numerically except for the unknown scaling factor μ. Besides the weights s_{ij} and d_i, there is a further quantity in (4.1). $z(q)$ is not a constant function in general. $\langle z \rangle_{\partial\Phi_i\Phi_j}$ is the mean value of z with regard

to the restricted Boltzmann distribution on the surface $\partial\Phi_i\Phi_j$. In a numerical computation for a given configurational state q, the term $z(q)$ in (3.7) can be approximated with standard Monte Carlo quadrature methods based on known distributions of the variables. The value $\langle z\rangle_{\partial\Phi_i\Phi_j}$ in (4.1) is the averaged positive value of the scalar product $\langle\bar{v}(q,p),\nu\rangle$ determined for the set of π_p-distributed momentum vectors p and the Boltzmann distributed q-vectors, where this q-distribution is restricted to the surface $\partial\Phi_i\Phi_j$ between the Voronoi cells Φ_i and Φ_j. In order to determine $\langle z\rangle_{\partial\Phi_i\Phi_j}$ in (4.1) numerically, a Monte-Carlo sampling of the Boltzmann distribution of configurational states $q\in\partial\Phi_i\Phi_j$ is needed. This can be done by using a constraint molecular dynamics algorithm [59, 2] combined with a Hybrid Monte-Carlo approach (HMC) [22]. HMC generates a Markov chain of configurational states. It consists of a proposal step and an acceptance step. In the proposal step, a new configurational state $\widetilde{q}$ is generated on the basis of a given state q. This is done with a molecular dynamics simulation with randomized initial momenta. The constraints to be fixed for $\widetilde{q}$ are given by the following system of equalities and inequalities.

$$\begin{aligned}&\mathrm{dist}(\widetilde{q},q_i)=\mathrm{dist}(\widetilde{q},q_j)\\&\mathrm{dist}(\widetilde{q},q_k)\geq\mathrm{dist}(\widetilde{q},q_i),\forall k=1,\ldots,m,\end{aligned}\tag{4.2}$$

where $q_k, k=1,\ldots m$, are the centers of the Voronoi cells. The equaility in (4.2) restricts the sampling to the correct hyperplane between Φ_i and Φ_j, whereas the inequalities assure the restriction to the edge $\partial\Phi_i\Phi_j$. In an implementation of this sampling method, only the equalities are fixed by the constraint molecular dynamics. The inequalities are tested in the acceptance step of the HMC sampling algorithm. If $\widetilde{q}$ is accepted, it is the next step of the Markov chain. If it is rejected, the starting point q is repeated in the chain. For each sampling point (q,p) of a HMC sampling algorithm, one has to determine the averaged velocity vector $\bar{v}(q,p)$ for the selected dynamical model first. If it is assumed that the averaged velocity of the dynamical model is given by Hamiltonian dynamics, then $\bar{v}(q,p)=M^{-1}p$, where M is the diagonal matrix of atom masses. In a second step one has to determine the scalar product $\langle\bar{v}(q,p),\nu\rangle$. If one of the aforementioned simplifications is applied in the distance computation, the surface normal vector $\nu(q)$ depends on the configurational state $q\in\partial\Phi_i\,\Phi_j$. In the case of a Voronoi tessellation, this surface normal vector is given by the difference vector $\nu(q)\propto\nabla_q\mathrm{dist}(q,q_i)-\nabla_q\mathrm{dist}(q,q_j)$ (normalized to $\|\nu\|_2=1$).

Avoiding the computation of $z(q)$. Assuming $\bar{v}(q,p)=M^{-1}p$ combined with a special choice of $\mathrm{dist}(\cdot,\cdot)$ can be used in order to avoid the computation of $z(q)$. If the distance function computation is only based on one type of atoms (e.g. on a subset of carbon atoms of the system) by neglecting the rest of the atoms or by solely including corresponding internal coordinates in the distance function, then the surface normal vector ν has only non-vanishing components for the entries corresponding to the coordinates of these atoms. For these atoms, the averaged velocity distribution $M^{-1}p$ is isotropic (for p distributed according to π_p). I.e., z is a constant number, independent from the direction of ν. This constant number can be included in the unknown scaling factor μ by setting $z\equiv 1$.

4.2 Adaptive and hierarchical refinement

There are mainly two sources of numerical errors when approximating the infinitesimal generator in a numerical approach via (4.1). The first error is given by the numerical quadrature error, because the entries of Q are not computed analytically, they are computed by a quadrature rule. For high-dimensional applications, Monte-Carlo quadrature is used which leads to a stochastic error. Röblitz mainly concentrates on this type of error in her thesis [58]. For the present section, it will be assumed that the quadrature error can be controlled and is reasonably small. Besides the quadrature error there is the discretization error addressed in chapter 3.4. This error will be controlled by the adaptive refinement procedure. If the computation of Q is based on *soft basis functions*, the adaptive algorithm of Röblitz for stochastic matrices can be used.

Soft basis functions. Using Gauß' Theorem, theoretically a horizontal sampling is sufficient for the computation of Q, see (3.7). One can generate Boltzmann distributed states inside the Voronoi cell Φ_i and restrict this sampling to the boundaries afterwards. Obviously, a sampling of states inside a Voronoi cell will never exactly hit a boundary of this cell in a numerical routine. Instead of a crisp Voronoi tessellation $\Phi^{[\infty]}$ one can, therefore, use continuous membership functions $\Phi^{[\alpha]}$ with $\lim_{\alpha\to\infty} \Phi^{[\alpha]} = \Phi^{[\infty]}$. One possible class of membership functions has been defined by Weber [75]:

$$\Phi_i^{[\alpha]}(q) := \frac{\exp(-\alpha \operatorname{dist}^2(q, q_i))}{\sum_{k=1}^{m} \exp(-\alpha \operatorname{dist}^2(q, q_k))}. \tag{4.3}$$

Using the basis of membership functions (4.3), the Galerkin discretization Q in (3.7) can be formulated in terms of a regularization $Q^{[\alpha]}$. The outer diagonal elements of $Q^{[\alpha]}$ are given by

$$Q^{[\alpha]}(i,j) = \mu \int_\Omega z(q)\, \Phi_j^{[\alpha]}(q)\, \pi_i^{[\alpha]}(q)\, dq = \mu \frac{\langle \Phi_i^{[\alpha]}, z\Phi_j^{[\alpha]} \rangle_{\pi_q}}{\langle \Phi_i^{[\alpha]}, e \rangle_{\pi_q}}, \tag{4.4}$$

where μ is an unknown scaling factor which represents the difference between a volume and a surface measure. It is shown in chapter 4.4 that $Q^{[\alpha]}$ for $\alpha \gg 0$ is an approximation for Q. (4.4) can be seen as a mean value computation of the function $z(q)\Phi_j^{[\alpha]}(q)$ for a restricted Boltzmann distribution of states q according to $\pi_i^{[\alpha]}$ with its corresponding density function

$$\pi_i^{[\alpha]}(q) := \frac{\Phi_i^{[\alpha]}(q)\, \pi_q(q)}{\int_\Omega \Phi_i^{[\alpha]}(q)\, \pi_q(q)\, dq}.$$

Note that the outer diagonal elements of $Q^{[\alpha]}$ in (4.4) are identical to outer diagonal elements of the *stochastic mass matrix* defined by Röblitz in [58] except for the additional term z and the scaling factor μ. The stochastic mass matrix defines the overlap of the basis functions $\Phi^{[\alpha]}$ and measures their crispness. The adaptive sampling algorithm and the statistical error analysis of Röblitz can be used to generate the matrix $Q^{[\alpha]}$ without performing the vertical sampling. The binding process in chapter 5.2.2 has been constructed with this kind of approach

using the software code ZIBgridfree [82, 75, 58] for a given value $\alpha \gg 0$. For high dimensional molecular systems, this approach may be applicable, but it suffers from the fact that low-weighted boundaries of the Voronoi cells are not sampled sufficiently. What has been said about Figure 3.2 does not only hold for the vertical sampling, it may also be a problem of the horizontal sampling, because the sampling may be too sparse for a statistical relevant estimation of $Q^{[\alpha]}$.

Crisp basis functions. In this thesis it is suggested to generate a sampling not only inside the Voronoi cells. Additionally, one should generate a sampling on the edges of the Voronoi cells in order to estimate the transition behavior sufficiently. The restricted sampling has been described in chapter 4.1 for the computation of z. A feasible starting point $\widetilde{q}$ for this sampling meets the constraints (4.2). This can be achieved by using the function basis $\Phi^{[\alpha]}$ in (4.3). A feasible point, $\widetilde{q} \in \partial\Phi_i\Phi_j$, has the following property

$$\lim_{\alpha\to\infty} \frac{1}{\Phi_i^{[\alpha]}(\widetilde{q})\,\Phi_j^{[\alpha]}(\widetilde{q})} = s^2,$$

where s is the number of inequalities in (4.2) which are active in $\widetilde{q}$. For an infeasible point, $\widetilde{q} \notin \partial\Phi_i\Phi_j$ the limit tends to infinity. In order to find a feasible starting point with a low energy value $V(\widetilde{q})$ one can, therefore, minimize the function

$$I_C^{[\alpha]}(\widetilde{q}) := \frac{V(\widetilde{q})}{\Phi_i^{[\alpha]}(\widetilde{q})\,\Phi_j^{[\alpha]}(\widetilde{q})}$$

within a homotopy method for $\alpha \to \infty$. If the minimal value $I_C^{[\alpha]}$ does not converge for $\alpha \to \infty$, then the Voronoi cells Φ_i and Φ_j are not neighboring and do not share a common surface $\partial\Phi_i\Phi_j$. If additional samplings are performed for the surfaces of the Voronoi cells, the adaptive error analysis of Röblitz is not appropriate. The elements of Q are not row-wise correlated any more. A different refinement strategy is needed. For an adaptive refinement, the approximation of the eigenfunctions of $\mathcal{Q}$ is important. The space spanned by the eigenfunctions is identical to the space spanned by the conformations. A set of basis functions using a crisp Voronoi tessellation of Ω assumes a constant value of the conformations χ inside the cells. This assumption is incorrect for regions of Ω where χ has large gradients (transition regions). This fact can be used in order to refine the set of basis functions adaptively. In an illustrative example in chapter 5.1, a very simple adaptive refinement approach is shown, which hierarchically adds basis functions in transition regions.

4.3 Efficient free energy calculation

For high-dimensional examples, the numerical computation of $\langle z\rangle_{\partial\Phi_i\Phi_j}$ in (4.1) has been described so far. The missing computation of the expressions s_{ij}/d_i includes high-dimensional integrals. Only Monte-Carlo quadrature methods are suitable for this task. Because of the complexity of the Boltzmann distribution, standard Monte Carlo quadrature methods are not appropriate in order to determine the expressions s_{ij}/d_i in (4.1). The weights of the surfaces and

of the Voronoi cells have to be computed by a routine which can determine free energy differences in Ω. In other words, an algorithm is searched which meets requirement R1 of the introduction not only for the statistical weights of the conformations but also for the weights of discretization boxes and for the weights of the surfaces. Three different approaches will be discussed here. The first approach is called ConfJump which is a Metropolis-Monte-Carlo sampling scheme for the computation of free energy differences. The second approach is based on the *Jarzynski Identity*. The third approach is a heuristic method which only takes into account restricted Boltzmann samplings (restricted to cells and surfaces). The latter method is based on a *density estimation approach* introduced by Weber and Andrae in [78]. It is recommended to apply a combination of density estimation and Jarzynski's Identity which will be described at the end of this section.

Algorithm ConfJump**.** Instead of determining the absolute values of d_i and s_{ij} one can alternatively estimate the ratios s_{ij}/s_{kl} or d_i/d_j. If these ratios are computed, the absolute values of the weights are known for d_i by a normalization to $\sum_{i=1}^{m} d_i = 1$. The values s_{ij} can be determined with this procedure except for the scalar factor μ. There are a lot of methods in literature which estimate the ratios of weights of different parts of the configurational space by defining a Metropolis-Monte-Carlo like sampling approach. This sampling is based on jumps between the restricted parts of the space. For a survey see [8]. ConfJump [74] is one algorithm of this type, it is very similar to the Smart Darting Monte Carlo algorithm of Andricioaei et al. [3] and the Jump Between Wells (JBW) algorithm of Senderowitz et al. [64, 65]. The ConfJump method is embedded into a HMC-sampling of the restricted parts (Voronoi cells or surfaces) of the configurational space. During the HMC sampling, sometimes (with a given probability) the molecular dynamics proposal step is replaced by a jump from the current part (A) of the configurational space to another restricted part (B) of Ω. For the construction of the jump it is important that the conditional probability for being in A and proposing a jump from a certain state $q \in A$ to $q^* \in B$ is identical to the probability for being in B and proposing a jump from $q^* \in B$ to $q \in A$. There are many different ways to construct a suitable jump. One possible construction is as follows. For each restricted part i of the configurational space a representative configurational state q_i is selected in advance. If a jump from A to B is intended in the ConfJump algorithm, then the difference vector $\Delta := q_A - q$ between the current sampling point q and the representative q_A is determined. The proposed new configurational state q^* in B is given by $q^* = q_B - \Delta$. The jump is accepted, if the potential energy has decreased, $V(q) > V(q^*)$, during the jump or if a Metropolis like acceptance criterion holds for the energy difference. If the jump is rejected, the HMC sampling continues in the old part A of Ω. If the jump is accepted, the sampling continues in the new part B. When ConfJump has converged, the numbers of sampling points per part of Ω represent the relative weights of these parts. In order to improve the jump acceptance rate, one can shift the minimal potential energy level of all restricted parts to 0 in advance. In this case, the acceptance is improved because the jumps are performed between configurational states with similar potential energies. After convergence of the sampling, one has to reweight the number of sampling points per part with the

Boltzmann expression of the energy shift. Even by using the energy shift, the construction of a jump proposal step in this scheme is not a trivial task. In our experiments, for every molecular system, a different jump strategy has to be applied in order to yield a suitable acceptance ratio. A severe problem of all jump methods is given by the treatment of explicit water molecules. Whenever a jump between two configurational states of a molecular system is performed, it is not easy to rearrange the water molecules within a reversible(!) proposal step. Instead of a reversible jump, the next approach to weighting is based on a non-reversible transport of sampling points and can, therefore, tackle the problem of explicit water molecules.

Jarzynski's Identity. Determining the ratios of weights is like a computation of free energy differences, see (2.11). For the estimation of free energy differences, the Jarzynski's Identity[1] is very important [41, 14, 57]:

$$\langle \exp(-\beta\, W) \rangle = \exp(-\beta\, \Delta A), \tag{4.5}$$

where W is the work along a non-reversible process. On the right hand side of equation (4.5) there is the desired free energy difference between Voronoi cells or between surfaces of Voronoi cells. The left hand side of (4.5) can be understood as an algorithmic approach towards the computation of free-energy differences. A restricted Boltzmann distribution of states (restricted to Voronoi cells or restricted to surfaces) can be seen as a Boltzmann distribution according to a modified Hamiltonian. Assume $H(q,p,0)$ represents the Hamiltonian for the restricted sampling of the Voronoi cell Φ_j with π_p-distributed momenta, whereas, $H(q,p,t)$ for a fixed value $t>0$ represents the Hamiltonian according to the restricted Boltzmann distribution in Voronoi cell Φ_i with π_p-distributed momenta. The approach is a non-reversible biased density propagation method transporting a Boltzmann density according to the Hamiltonian $H(q,p,0)$ in time t. If the parameter t is fixed, the corresponding Hamiltonians $H(\cdot,\cdot,\tau)$ can be approximated by using the basis functions in (4.3) with $\alpha \gg 0$. For a transport from cell Φ_j to cell Φ_i the Hamiltonian is:

$$H(q,p,\tau) = -\frac{1}{\beta}\ln(\Phi_i^{[(\tau\alpha)/t]}(q)) - \frac{1}{\beta}\ln(\Phi_j^{[\alpha-(\tau\alpha)/t]}(q)) + V(q) + K(p). \tag{4.6}$$

For a transport from surface $\partial\Phi_i\Phi_j$ to surface $\partial\Phi_i\Phi_k$ the Hamiltonian is (see chapter 4.4):

$$\begin{aligned} H(q,p,\tau) \;=\; & -\frac{1}{\beta}\ln(\Phi_i^{[\alpha]}(q)) - \frac{1}{\beta}\ln(\Phi_k^{[(\tau\alpha)/t]}(q)) - \frac{1}{\beta}\ln(\Phi_j^{[\alpha-(\tau\alpha)/t]}(q)) \\ & + V(q) + K(p). \end{aligned} \tag{4.7}$$

[1] It is not possible to write down this equation without stressing out the incredible importance of Jarzynski's Identity for all fields of thermodynamics. Jarzynski's Identity replaces the second law of thermodynamics. This law has been $\langle W \rangle \leq \Delta A$ for non-reversible dynamical processes. It has been formulated in terms of an inequality (only for reversible processes the equality holds). Jarzynski has shown that the second law can be formulated in terms of an equality. Although the dynamical processes are assumed to be reversible on a microscopic and infinitesimal level (self-adjointness of $\mathcal{P}$), via Jarzynski's Identity it is possible to compute the statistical weights of conformations by a substitution of the reversible metastable process with a rapidly mixing non-reversible dynamics.

Assume that the restricted density of Φ_j has been sampled with HMC. Then, a molecular dynamics simulation according to the time-dependent Hamiltonian $H(\cdot,\cdot,\cdot)$ starting in $(q(0),p(0),0)$ and ending in $(q(t),p(t),t)$ transports the HMC-sampling point $(q(0),p(0))$ from one restricted region to the state $(q(t),p(t))$ of the other restricted part of configurational space. The left hand side in (4.5) is the averaged exponential work of this biased molecular dynamics simulation. The average is taken over the sampling points in Φ_i. Note that selecting Hamiltonian dynamics for the computation of free energy differences is not defining the dynamical model of $\mathcal{P}(\tau)$. Hamiltonian dynamics is selected, because it is a symplectic mapping. Any dynamical model can be used in order to propagate the density (even stochastic dynamical models). However, Hamiltonian dynamics offers a simple way to compute the work W along the propagation. Using that $W = H(q(t),p(t),t) - H(q(0),p(0),0)$ for Hamiltonian dynamics (see chapter 5.3.1 in [10]) and (2.11), the Jarzynski Identity is given by

$$\Big\langle \exp\big(-\beta(H(q(t),p(t),t) - H(q(0),p(0),0))\big)\Big\rangle \approx \frac{d_i}{d_j}. \tag{4.8}$$

The equation (4.8) is formulated for time-dependent Hamiltonian (4.6) and the weights d_i, d_j. For s_{ij} there is an analogous equation when using the time-dependent Hamiltonian (4.7). In an algorithmic realization of (4.8), an ensemble of states distributed according to $H(q,p,0)$ is generated first. In a second step the algorithm performs biased molecular dynamics trajectories on each sampling point. This simulation propagates the density from one to other restricted regions of Ω. In order to compute equation (4.8) numerically, the exponential difference between the Hamiltonian in the starting points and the Hamiltonian in the end points of the molecular dynamics trajectory have to be averaged.

Density estimation. It can be very difficult to design a suitable algorithm for the biased movement of densities in Jarzynki's approach with a reasonable small amount of work W and a suitable simulation time t (especially for the voronoi cell surface weights). Instead of performing a biased dynamics simulation for the determination of free energy differences one can also estimate the free energy differences on the basis of the generated restricted HMC sampling data within the cells and on the surfaces (see chapter 4.1). This approach will be formulated for the computation of free energy differences between the Voronoi cells (for the determination of d_i), but it can also be applied for the estimation of s_{ij}. Assume, there is a small $\delta_{\tilde{q}}$-environment of a given configurational state $\tilde{q} \in \Phi_i$ in which the value $\exp(-\beta\, V(q))$ can be considered as constant. In this case, the following equation holds:

$$\int_{\Phi_i} \delta_{\tilde{q}}\, dq \cdot \exp(-\beta\, V(\tilde{q})) = \underbrace{\int_{\Omega} \delta_{\tilde{q}} \frac{\Phi_i(q)\, \exp(-\beta\, V(q))}{\int_{\Phi_i} \exp(-\beta\, V(\bar{q}))\, d\bar{q}} dq}_{=:\langle\delta\rangle_{\pi_i}} \cdot \underbrace{\int_{\Phi_i} \exp(-\beta\, V(\bar{q}))\, d\bar{q}}_{\propto d_i}. \tag{4.9}$$

If configurational states $q_i \in \Phi_i$ and $q_j \in \Phi_j$ are selected with an identical δ-environment, then equation (4.9) can be reformulated as

$$\frac{d_i}{d_j} = \frac{\exp(-\beta\, V(q_i))\, \langle\delta\rangle_{\pi_j}}{\exp(-\beta\, V(q_j))\, \langle\delta\rangle_{\pi_i}}. \tag{4.10}$$

In (4.10), the expression $\langle\delta\rangle_{\pi_i}$ can be approximated by the averaged number of states which are element of the δ-environment of q_i. The averaged value is taken over a set of states which are distributed according to the restricted Boltzmann sampling of Φ_i. This is comparable to a density estimation of this restricted sampling at q_i. In order to use (4.10) in a numerical routine, one has to determine a representative q_i for each Voronoi cell Φ_i and one has to count the averaged number of sampling points which are element of a small environment of q_i. It is important to take a δ-environment with a constant volume for each Voronoi cell Φ_i. The representative q_i can be a local minimum. The computational complexity estimation of the algorithmic approach is based on the density estimation method for the determination of d_i and for the computation of the surface integrals (see Table 4.1 on page 58). Although the density estimation approach is theoretically applicable for the Voronoi cell surfaces, the sampling of the Voronoi cell surfaces as described in chapter 4.1 can suffer from the problem of additional energy barriers. Furthermore, it can be difficult to figure out a description of small iso-volume δ-environments on different Voronoi cell suraces, if the Voronoi cell is determined by using internal coordinates. In this case, the determinant of the Jacobian of the transformation into internal coordinats is needed. Instead of using a density estimation for the surface integrals, a combined approach is recommended.

Combination of density estimation and Jarzynski's Identity. Assume the density estimation method is used for the estimation of the weights d_i of the Voronoi cells. Thus, HMC samplings restricted to the Voronoi cells have been generated and evaluated according to (4.10). In this case, the absolute values of d_i are known using the sum-1-normalization. For the estimation of the transition pattern, one can (approximately) transport these sampling points to the Voronoi cell surfaces. The corresponding time-dependent Hamiltonian H_{ij} for the transport $\Phi_i \to \partial\Phi_i\Phi_j$ is

$$H_{ij}(q,p,\tau) = -\frac{1}{\beta}\ln(\Phi_i^{[\alpha]}(q)) - \frac{1}{\beta}\ln(\Phi_j^{[(\tau\alpha)/t]}(q)) + V(q) + K(p).$$

In this case,

$$\frac{\left\langle \exp\big(-\beta(H_{ij}(q(t),p(t),t) - H_{ij}(q(0),p(0),0))\big)\right\rangle}{\left\langle \exp\big(-\beta(H_{ik}(q(t),p(t),t) - H_{ik}(q(0),p(0),0))\big)\right\rangle} \approx \frac{s_{ij}}{s_{ik}}. \tag{4.11}$$

A justification for this approach is given in chapter 4.4. If the method (4.11) is used, then the ratios of the surface integrals per Voronoi cell are known. Thus, exept for unknown row-wise scaling factors $r_1,\ldots,r_m > 0$, the matrix Q is determined. The unknown scalig factors can be estimated in the following way. By using the row-wise ratios s_{ij}/s_{ik} of the surface integrals and a row-wise sum-1-normalization, one can determine a stochastic $m \times m$-matrix K. The desired matrix Q can be written as $Q = R(K-I)$, where $R = \mathrm{diag}(r_1,\ldots,r_m)$ is the diagonal matrix of the unknown scaling factors and I is the $m \times m$-unity matrix. Since $d^\top Q = 0$, we can determine the ratios by solving $d^\top R(K-I) = r^\top D(K-I) = 0$, where $r = (r_1,\ldots,r_m)$ and $D = \mathrm{diag}(d_1,\ldots,d_m)$. The recommended approach is a combination of Jarzynski's Identity based on an adaptive discretization of Ω (requirement R2) and the direct computation of statistical weights using the density estimation approach (requirement R1).

4.4 Approximation of the surface flux

In the algorithmic realization of the discretization of $\mathcal{Q}$ it is often assumed, that the α-regularization is applicable:

$$\frac{\oint_{\partial\Phi_i\Phi_j} f(q)\,\exp(-\beta\,V(q))\,dS}{\oint_{\partial\Phi_i\Phi_k} f(q)\,\exp(-\beta\,V(q))\,dS} \approx \frac{\int_\Omega f(q)\,\Phi_i^{[\alpha]}(q)\,\Phi_j^{[\alpha]}(q)\,\exp(-\beta\,V(q))\,dq}{\int_\Omega f(q)\,\Phi_i^{[\alpha]}(q)\,\Phi_k^{[\alpha]}(q)\,\exp(-\beta\,V(q))\,dq}. \tag{4.12}$$

Instead of comparing the $(3N-1)$-dimensional surface integrals, a full-dimensional approximation is used. Because of the special choice of $\Phi_i^{[\alpha]}$ the denominator and the numerator of the right expression converge against zero for $\alpha \to \infty$. They additionally converge against zero for every differential $\partial/\partial\alpha$.

Justification of the α-approximation. In order to show (4.12), the following considerations are important. If the Voronoi surface $\partial\Phi_i\Phi_j$ is a part of a plane[2] in Ω, then the Riemann sum of (congruent) intervals of these surfaces can be extendet to a Riemann sum of (congruent) boxes in Ω, see Figure 4.1. The volume integrals can be expressed (separately) by a limit over a decreasing

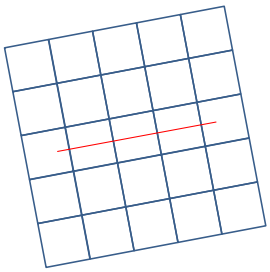

Figure 4.1: Expansion of an interval-based decomposition to a volume-based decomposition. By defining $f(q) = 0$ for all $q \notin \partial\Phi_i\Phi_j$ for the surface integrals, the interval-based decomposition need not "fit" exactly.

volume of the boxes. The volume per box is c^{3N}:

$$\begin{aligned} &\frac{\int_\Omega f(q)\,\Phi_i^{[\alpha]}(q)\,\Phi_j^{[\alpha]}(q)\,\exp(-\beta\,V(q))\,dq}{\int_\Omega f(q)\,\Phi_i^{[\alpha]}(q)\,\Phi_k^{[\alpha]}(q)\,\exp(-\beta\,V(q))\,dq} \\ = \quad &\frac{\lim\limits_{c\to 0}\sum\limits_{l=1}^{L(c)} c^{3N}\,f(q_l)\,\Phi_i^{[\alpha]}(q_l)\,\Phi_j^{[\alpha]}(q_l)\,\exp(-\beta\,V(q_l))}{\lim\limits_{c\to 0}\sum\limits_{l=1}^{L(c)} c^{3N}\,f(q_l)\,\Phi_i^{[\alpha]}(q_l)\,\Phi_k^{[\alpha]}(q_l)\,\exp(-\beta\,V(q_l))} \end{aligned}$$

[2]This is the case, if the Voronoi cells are defined by the Euclidean distance in Ω. A neglection of atoms as described on page 47 is possible.

$$= \frac{\lim\limits_{c\to 0} \sum\limits_{l=1}^{L(c)} c^{(3N-1)}\, f(q_l)\, \Phi_i^{[\alpha]}(q_l)\, \Phi_j^{[\alpha]}(q_l)\, \exp(-\beta\, V(q_l))}{\lim\limits_{c\to 0} \sum\limits_{l=1}^{L(c)} c^{(3N-1)}\, f(q_l)\, \Phi_i^{[\alpha]}(q_l)\, \Phi_k^{[\alpha]}(q_l)\, \exp(-\beta\, V(q_l))},$$

where the states q_l are taken from the midpoints of the boxes, which are also the midpoints of the intervals (for the boxes which include the surface). Note that the partial sums are almost zero for states q_l which are not part of the corresponding surfaces for $\alpha \gg 0$. Furthermore, the Φ-product is constantly almost 0.25 for the surface points and can be canceled out. This shows that the last equation approximately is the Riemann sum of the surface integrals. Thus, the regularization is a good approximation of the ratio of the surface integrals. The considerations show that in equation (4.4) the ratio $Q^{[\alpha]}(i,j)/Q^{[\alpha]}(i,k)$ approximates the ratio $Q(i,j)/Q(i,k)$. Furthermore, for Q the equation $d^\top Q = 0$ holds, which determines the row-wise scaling. In order to also meet $\tilde{d}^\top Q^{[\alpha]} = 0$, with $\tilde{d}_i = \langle \Phi_i^{[\alpha]}, e\rangle_{\pi_q}$, the matrix $Q^{[\alpha]}$ has to be defined according to the row-wise scaling in (4.4). At this stage, the unknown scaling factor μ occurs. $\tilde{d}^\top Q^{[\alpha]} = 0$ is independent from μ.

Using internal coordinates. In a strict sense, in equation (4.4), (4.7), and (4.11) the surfaces of the Voronoi cells have to be assumed to be parts of planes[3] in Ω in order to meet the requirements of the aforementioned considerations. Thus, using internal coordinates for the simplification of the configurational space as described in chapter 4.1 is not included. For high-dimensional problems, however, one may want to restrict the point of view to a few internal coordinates which describe conformational changes and the leading time-scales sufficiently. Note, that in Theorem 4 the expression $\langle \bar{v}(q,p), \nu\rangle$ in principle accounts for the "derivative" of Φ_j with regard to time. This expression is like applying the chain rule for the derivative of $\Phi_j(\widetilde{q})$, where $\widetilde{q}$ is the result of a realization of the dynamical model. In a strict sense, the crisp membership functions Φ_j do not allow for this derivative. Therefore, in Gauß' Theorem the surface integral construction occurs. We will now use the soft membership functions in (4.3) instead of crisp sets. The eigenequation

$$\mathcal{P}(\tau)\, X_i = \exp(\xi_i \tau)\, X_i$$

for the computation of the desired quantities X_i and ξ_i can be discretized via

$$P^{[\alpha]}(\tau)\, X_i = \exp(\xi_i \tau)\, S^{[\alpha]}\, X_i$$

using two stochastic matrices

$$P^{[\alpha]}(\tau)(i,j) := \frac{\langle \Phi_i^{[\alpha]}, \mathcal{P}(\tau)\Phi_j^{[\alpha]}\rangle_{\pi_q}}{\langle \Phi_i^{[\alpha]}, e\rangle_{\pi_q}}, \quad S^{[\alpha]}(i,j) := \frac{\langle \Phi_i^{[\alpha]}, \Phi_j^{[\alpha]}\rangle_{\pi_q}}{\langle \Phi_i^{[\alpha]}, e\rangle_{\pi_q}}. \tag{4.13}$$

For the derivative of the eigenequation with regard to τ at $\tau = 0$ the resulting eigenequation is

$$\bar{Q}^{[\alpha]}\, X_i = \xi_i\, S^{[\alpha]}\, X_i,$$

[3] A measure-preserving transformation of the states q_l is possible.

where

$$\bar{Q}^{[\alpha]} := \frac{\partial}{\partial \tau}\Big|_{\tau=0} P^{[\alpha]}(\tau). \tag{4.14}$$

Using the regularization $\Phi_j^{[\alpha]}$, a full-dimensional time-derivative of $P^{[\alpha]}(i,j)$ is possible. For the approximation of the contribution of a $\Phi_i^{[\alpha]}\pi_q$-distributed state $q \in \Omega$ to the flux from Φ_i to Φ_j, expressions of the type $\langle \nabla_q \Phi_j^{[\alpha]}, \dot{q} \rangle$ have to be computed. Let $d_j(q) := \mathrm{dist}^2(q, q_j)$ denote the squared distance between q and the Voronoi center q_i in (4.3), then

$$\nabla_q \Phi_j^{[\alpha]}(q) = \alpha\, \Phi_j^{[\alpha]}(q) \left(\Big(\sum_{l=1}^{m} \Phi_l^{[\alpha]}(q)\, \nabla_q d_l(q) \Big) - \nabla_q d_j(q) \right). \tag{4.15}$$

In order to determine the matrices $S^{[\alpha]}$ in (4.13) and $\bar{Q}^{[\alpha]}$ in (4.14), mean value computations with regard to a $\Phi_i^{[\alpha]}$-restricted Boltzmann sampling are mandatory. Because of the pre-factor $\Phi_j^{[\alpha]}$ in (4.15), the mean value computation of the $(\int_\Omega \Phi_i^{[\alpha]})$-type can be transformed into a $(\int_\Omega \Phi_i^{[\alpha]} \Phi_j^{[\alpha]})$-type mean value computation, i.e. restriction to Voronoi cell "surfaces" is possible.

Reactive flux. For every sampling point q in the aforementioned approach, the expression $\langle \nabla_q \Phi_j^{[\alpha]}, \dot{q} \rangle$ has to be averaged over the Boltzmann distributed set of momenta p. One possible choice for the velocities is $\dot{q} := M^{-1}p$. In this case the averaged expression is zero. Like in Theorem 4, a certain portion of the flux has to be assiged for a transition $\Phi_i \to \Phi_j$. This portion is denoted as the *reactive flux*. According to Theorem 4 the contribution $f_{ij}(q)$ of a $\Phi_i^{[\alpha]}$-sampled state q to the flux from Φ_i to Φ_j can be seen as the p-averaged value of

$$f_{ij}(q) := \overline{\max\{\langle \nabla_q \Phi_j^{[\alpha]}, \dot{q} \rangle, 0\}}.$$

In order to meet the detailed balance property for this choice, for every transition $\Phi_i \to \Phi_j$ with momentum vector p there has to be the corresponding transition $\Phi_j \to \Phi_i$ with momentum vector $-p$. This means, that $\nabla_q \Phi_j^{[\alpha]}(q) = -\nabla_q \Phi_i^{[\alpha]}(q)$. A short calculation shows, that this is valid, if $\Phi_i^{[\alpha]}(q) + \Phi_j^{[\alpha]}(q) = 1$. Equivalently, $\Phi_k^{[\alpha]}(q) = 0$ for all $k \neq i, k \neq j$. This implies $\alpha \gg 0$.

Algorithmic realization. When approximating ratios like $\bar{Q}^{[\alpha]}(i,j)/\bar{Q}^{[\alpha]}(i,k)$, the constant factor α in (4.15) cancels out.

$$\begin{aligned} & \frac{\bar{Q}^{[\alpha]}(i,j)}{\bar{Q}^{[\alpha]}(i,k)} \\ = \quad & \frac{\int_\Omega \overline{\max\{\langle (\sum_{l=1}^m \Phi_l^{[\alpha]}(q)\, \nabla_q d_l(q)) - \nabla_q d_j(q), \dot{q} \rangle, 0\}}\, \pi_{ij}(q)\, dq}{\int_\Omega \overline{\max\{\langle (\sum_{l=1}^m \Phi_l^{[\alpha]}(q)\, \nabla_q d_l(q)) - \nabla_q d_k(q), \dot{q} \rangle, 0\}}\, \pi_{ik}(q)\, dq} \cdot \frac{S^{[\alpha]}(i,j)}{S^{[\alpha]}(i,k)} \end{aligned}$$

$$(*) \quad \approx \quad \frac{\int_\Omega \overline{\max\{\langle \Phi_i^{[\alpha]}(q)(\nabla_q d_i(q) - \nabla_q d_j(q)), \dot{q}\rangle, 0\}}\, \pi_{ij}(q)\, dq}{\int_\Omega \overline{\max\{\langle \Phi_i^{[\alpha]}(q)(\nabla_q d_i(q) - \nabla_q d_k(q)), \dot{q}\rangle, 0\}}\, \pi_{ik}(q)\, dq} \cdot \frac{S^{[\alpha]}(i,j)}{S^{[\alpha]}(i,k)}, \tag{4.16}$$

in $(*)$ the approximation $\Phi_i^{[\alpha]}(q) + \Phi_j^{[\alpha]}(q) = 1$ is used. Furthermore, $\pi_{ij}(q)$ denotes the $\Phi_i^{[\alpha]}\Phi_j^{[\alpha]}$-restricted Boltzmann distribution of states and

$$S^{[\alpha]}(i,j) = \frac{\langle \Phi_i^{[\alpha]}, \Phi_j^{[\alpha]} \rangle_{\pi_q}}{\langle \Phi_i^{[\alpha]}, e \rangle_{\pi_q}}.$$

In this case, a reweighting of the restricted Boltzmann samplings according to $\Phi_i^{[\alpha]}\Phi_j^{[\alpha]}$ and $\Phi_i^{[\alpha]}$ can be done in the full-dimensional space by the methods described in chapter 4.3 (a surface measure does not occur). This reweighting leads to a matrix $S^{[\alpha]}$. The π_{ij}- and π_{ik}-samplings, however, only lead to ratios $\bar{Q}^{[\alpha]}(i,j)/\bar{Q}^{[\alpha]}(i,k)$ in (4.16). Note that $\bar{Q}^{[\alpha]}$ can be determined except for an unknown scaling factor μ by the eigenvector method described in chapter 4.3 on page 53. Thus, this approach needs Voronoi cell samplings for the determination of statistical weights and Voronoi cell "surface" samplings for the determination of transition rates.

4.5 Efficiency of the algorithmic approach

The efficiency of statistical thermodynamics simulation software is determined by two aspects. First, the generation of sampling points for a statistical evaluation should converge in a reasonable time, i.e. the number of sampling points needed for the statistical result should be small. Second, the generation of the sampling points should not need too much computing time, i.e. the CPU time per sampling point should be small.

Number of sampling points. Unbiased molecular dynamics simulation is not able to estimate statistical quantities of conformation dynamics within a small number of sampling points. This has been shown in the Introduction. Rare events are the reason for bad statistics. An order of $O \cdot 1/(1-\lambda_2)$ sampling points is needed for convergence, where λ_2 is the second largest eigenvalue of the underlying transfer operator, see first row in Table 4.1. The term O denotes the computational cost for a statistically sufficient sampling of a rapidly mixing part of the state space. Using a space decomposition approach is a good starting point for the reduction of the computational cost, because this approach circumvents the problem of rare events. When using a space decomposition, the statistical weights of the subsets of Ω and the transition behavior between these sets have to be computed. It is not easy to solve this task efficiently. In many algorithmic approaches, the overlap of different localized (modified) Boltzmann densities is an important quantity. Assume, one wants to estimate the overlap of a localized density A with a localized density B and with a localized density C. A very common approach is given by the following method: The estimation is based on a truncated sampling of the density A and on counting the states of A in the overlap regions 'AB' and 'AC'. One can show that this approach can be very inefficient. In many cases, the ratio of the statistical weight of

approach	complexity	problem
unbiased sampling (e.g. thermostated dynamics)	$O \cdot 1/(1-\lambda_2)$	rare events during the sampling process
restricted sampling using overlap statistics (e.g. umbrella sampling)	$O \cdot m \cdot 1/r$	computation of the overlap of sparsely populated regions
Theorem of Gauß	$O \cdot m \cdot (f_m + 1)$	direct estimation of statistical weights

Table 4.1: The complexity of different algorithmic approaches to conformation dynamics compared with a sampling of transition rates based on Gauß' Theorem.

AB compared to the weight of AC is relevant. In other words, if $b = AB/A$ denotes the statistical weight of AB with regard to A and $c = AC/A$ denotes the statistical weight of AC with regard to A, then the ratio b/c is important. If b and c are small numbers, then a small truncation error in b or c leads to a large error in determining b/c. For this reason, a lot of sampling data is needed. The corresponding complexity is shown in the second row of Table 4.1. For the decomposition approach, m denotes the number of Voronoi cells. The ratio $1/r$ with $r = \exp(-\beta\,\Delta E)$ is taken from the considerations in the beginning of chapter 3.6 on page 43. $1/r$ can be very large. The solution to this problem is given by a direct sampling of the overlap. In our case, the solution to an efficient simulation is given by a restricted sampling of the edges of the Voronoi cells justified by Gauß' Theorem. In the third row of Table 4.1, f_m denotes the averaged number of neighbors of the Voronoi cells. In the case of a metastable dynamical system with a moderate number of neighbors per Voronoi cell, the presented algorithmic approach based on Gauß' Theorem possesses the smallest order of complexity.

The most CPU-consuming part of molecular simulation software. The main complexity problem of classical molecular dynamics simulation is given by the evaluation of the potential energy function $V(q)$. Regarding all atomic interactions, the computational cost of evaluating $V(q)$ increases quadratically with the number N of atoms of the system. Therefore, the most time-consuming step in classical molecular dynamics simulations is the solution of the Hamiltonian dynamics equations. The aim of our work at the Zuse Institute in Berlin was to get rid of Hamiltonian dynamics simulations for the analysis of transition networks of molecular systems. In 1998, Peter Deuflhard already stressed that long-term Hamiltonian dynamics simulations cannot solve an ill-conditioned problem. Thus, this kind of simulation approach for molecular systems cannot be the basis for the analysis of molecular systems. A good starting point to invent something new. Reading this thesis one may come to the conclusion that Hamiltonian dynamics occurs at many different "places" of the software code, but this is not obligatory.

HMC without Hamiltonian dynamics. The first place, where Hamiltonian dynamics may occur, is the Hybrid Monte-Carlo sampling routine. HMC is based on a molecular dynamics simulation used in order to perform the pro-

posal step. From a mathematical point of view, HMC only needs a reversible, volume-preserving mapping as proposal step [22]. HMC does not need molecular dynamics. Convergence against the Boltzmann distribution of states is assured, if the proposal step leads to an ergodic Markov chain. Although ZIBgridfree uses Hamiltonian dynamics, it does not perform molecular dynamics simulations on the full molecular system in each HMC step. It does not need a computation of all potential energy terms of $V(q)$ during the proposal step. Instead of moving all atoms of a molecular system one can restrict the proposal step to a (randomly chosen) subset of atoms[4]. During the proposal step, a short-time molecular dynamics simulation of the molecule is performed by fixing the positions of the non-selected atoms. Since all interaction energies between the fixed atoms are constant during the MD simulation, the corresponding terms need not be computed. Instead of computing N^2 interactions, where N is the total number of atoms, only $(n \cdot N)$ interactions have to be considered, where n is the number of selected atoms. Using this approach, the complexity reduction of HMC is based on the following consideration. The interaction energy between non-bonded parts of the molecular system decreases rapidly, such that separated parts of the system have an almost independent distribution of conformational states. The word "almost" is important this context. If separated parts really had independent distributions, then using a cutoff value for the computation of long-range interactions during the MD proposal step would also lead to the same complexity reduction. Neglecting long-range interactions may lead to incorrect results [52, 85]. In contrast of defining a cutoff, HMC with randomly chosen subsets of moving atoms accounts for all long-range interactions.

Jarzynski's Identity without Hamiltonian dynamics. The second place where Hamiltonian dynamics may be important for the algorithmic realization is given by the propagation of HMC sampling points from the surface of Voronoi cells into the center of them. As mentioned in chapter 4.3, Hamiltonian dynamics is selected in this context for the presentation of theoretical results because it simplifies the computation of the work W along the propagation. Note that an arbitrary non-reversible propagation can be applied, see also [24]. This is the key contribution of Jarzynski's Identity in statistical thermodynamics. For arbitrary non-reversible propagations, the work can be computed as

$$W = \int_0^t \frac{\partial H(q(\tau), p(\tau), \tau)}{\partial \tau}\, d\tau.$$

But how can one get rid of molecular dynamics simulations, although transition rates have to be computed (which are based on a dynamical model). Isn't there the theorem that says, that complexity of a problem can not be reduced by reformulation? The intrinsic reduction of the complexity of the conformation dynamics approach is given by Gauß' Theorem and the insight that knowing the exact transition processes *within* the conformations is not necessary for the computation of transition rates *between* the conformations. Note, however,

[4]The atoms which are selected for the proposal step of HMC should not be chosen purely by random. One should build groups of atoms in advance and select one of these groups for a short-time molecular dynamics simulation. Some (important) groups can be selected with a higher probability, in order to improve the statistics of this group. As long as the probability distribuion for the selection of the groups is identical in each proposal step, the detailed-balance condition of HMC is assured, see also [74].

that being blind for the transition processes within the conformations is not the solution. Some information is needed which is given by the eigenfunctions of the generator $\mathcal{Q}$. Whatever is the answer to the question of complexity, one aspect is clear: Molecular simulation can definitely get by without *long-term* Hamiltonian dynamics trajectories! Furthermore, long-term Hamiltonian dynamics trajectories focus on the simulation of the transition processes *within* the conformations which is the most uninteresting information for conformation dynamics.

Chapter 5

Numerical Examples

In this chapter two different numerical examples of the subspace approach to molecular Markov State Models are presented. Example 5.1 is based on an artificial 2-dimensional potential energy function, whereas Example 5.2.2 is taken from a real-word simulation of a binding process of the Aminopeptidase N inhibitor bestatin (APN).

5.1 Illustrative Example

In order to illustrate the computation of Q, we start with a simplified 2-dimensional example of a potential energy function [55] and $\beta = 3.34$. In this example, a system of two atoms moving in one dimension is analyzed. The potential energy function is $V : R^2 \to R$ with

$$\begin{aligned} V(q) &= 3\,\exp\big(-q_1^2-(q_2-\frac{1}{3})^2\big) - 3\,\exp\big(-q_1^2-(q_2-\frac{5}{3})^2\big) \\ &\quad -5\,\exp(-(q_1-1)^2-q_2^2) - 5\,\exp(-(q_1+1)^2-q_2^2) \\ &\quad +0.2\,q_1^4 + 0.2\,(q_2-\frac{1}{3})^4. \end{aligned} \tag{5.1}$$

A contour plot of this q_1-symmetric function is shown in Fig. 5.1, see also [50]. The potential function has three local minima. A direct thermostated dynamics simulation on the basis of the potential energy function does not provide correct statistics, see Figure 5.2. The matrix Q will be evaluated for a set of indicator functions Φ stemming from a regular grid of $9 \times 9 = 81$ discretization boxes in $\Omega = [-2;2] \times [-2;2]$. Up to this stage, the dynamical model which is in the "heart" of $\mathcal{Q}$ is not yet defined. The computation of $\bar{v}(q,p)$ is based on further assumptions. For the computation of Q we will assume that the mean velocity vector $\bar{v}(q,p)$ is independent from q (like in Hamiltonian dynamics) in this simple example. We will also assume that $z(q)$ is independent[1] from the direction of the normal vector ν. In this case, $z(q) = z$ is a constant number only leading to a rescaling of Q. Q is only computed up to an unknown scaling factor $\mu > 0$. z can be set to 1. The normalization of the densities is also part of the unknown

[1] For Hamiltonian dynamics, the velocity vector is $\bar{v}(q,p) = \dot{q} = M^{-1}p$. For this special choice z is independent from the direction of ν if all masses of the atoms are equal.

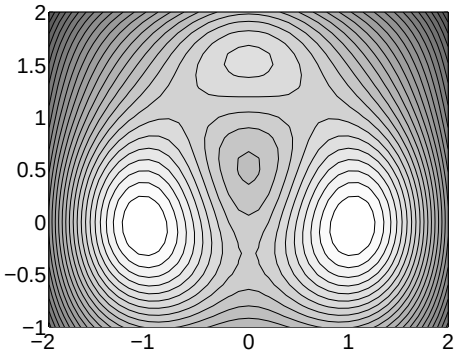

Figure 5.1: (Example 5.1) Potential energy function. The function has three local minima. Two deeper ones at about $(-1;0),(1;0)$, and one minimum at about $(0;5/3)$.

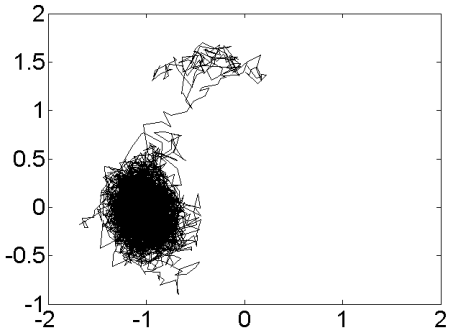

Figure 5.2: (Example 5.1) Direct thermostated simulation of the potential energy function (10000 steps). Starting in the upper lower minimum, the trajectory crosses the barrier and is trapped in one of the deeper local minima.

scaling factor μ. Thus, instead of the normalized expression $\pi_q(q)$, the integrals in (4.1) are based on the Boltzmann expression $\exp(-\beta\, V(q))$. Because of $z \equiv 1$ the computation of the elements of $Q(i,j), i \neq j$, is only based on the ratio s_{ij}/d_i, i.e. only on the weights of intersecting surfaces and on the weights of the boxes. The diagonal elements of Q are determined such that the row sums of Q are zero. The surface weights $s_{ij} = \int_{\partial\Phi_i\Phi_j} \exp(-\beta\, V(q))\, dS$ can be computed numerically with a one-dimensional Gauß-Legendre quadrature rule based on 5 nodes per surface [67]. The weights of the boxes $d_i = \int_{\Phi_i} \exp(-\beta\, V(q))\, dq$ are computed by a two-dimensional Gauß-Legendre quadrature rule based on 25 nodes per box. After computation of the matrix Q, the matrix Q_c is approximated analogously to P_c in chapter 3.4 by $\widetilde{Q}_c = \mathcal{A}_{disc}^{-1}\, \Xi_{disc}\, \mathcal{A}_{disc}$. The matrices $\mathcal{A}_{disc}$ and Ξ_{disc} are 3×3 matrices in our example. These matrices are approximations of the matrices $\mathcal{A}$ and Ξ for the continuous operator $\mathcal{Q}$. Note that the matrices $\mathcal{A}$ and Ξ are also 3×3-dimensional in the continuous case. Without loss of generality, the scaling of Q is chosen in a particular way: Q defines an 81×81-matrix $M := Q + I$ which is a stochastic matrix with minimal diagonal element 0.

The dominant eigenvectors X_{disc} of M and the dominant eigenvectors of Q are identical. Since M is a stochastic matrix, the standard PCCA+ algorithm can be applied in order to find χ_{disc} and the corresponding transformation matrix $\mathcal{A}_{disc}$. If $\lambda_{disc} \approx 1$ is an eigenvalue of M, then $\xi_{disc} = \lambda_{disc} - 1$ is an eigenvalue of Q. For the regular 9×9-discretization, the results are

$$\Xi_{disc} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -0.000065 & 0 \\ 0 & 0 & -0.012396 \end{pmatrix}$$

and

$$\mathcal{A}_{disc} = \begin{pmatrix} 0.002413 & 0.498793 & 0.498794 \\ 0.000000 & -0.480285 & 0.480285 \\ -0.039562 & 0.019781 & 0.019781 \end{pmatrix}.$$

Note, that $\mathcal{A}_{disc}$ meets the theoretical results from Lemma 2 and Corollary 1. The statistical weights w of the conformations are given by the first row of $\mathcal{A}_{disc}$. Because of the symmetric discretization, the statistical weights represent the symmetry of the potential energy function. Furthermore, the approximated infinitesimal generator has the desired symmetry[2]

$$\widetilde{Q}_c = \begin{pmatrix} -0.012366 & 0.006183 & 0.006183 \\ 0.000030 & -0.000047 & 0.000017 \\ 0.000030 & 0.000017 & -0.000047 \end{pmatrix}. \tag{5.2}$$

Note, that $\widetilde{Q}_c$ is computed up to an unknown scaling factor μ. Therefore, it is computed without giving the time unit. If the time unit is unknown the absolute values of $\widetilde{Q}_c$ do not make sense. Only ratios of elements of $\widetilde{Q}_c$ can be interpreted physically. In order to give an example for a possible physical interpretation: The transition rate from conformation two (one of the deeper minima in Figure 5.1) to the upper minimum is about 2 times larger than the transition rate between the deeper minima, $\widetilde{Q}_c(2,1) \approx 2 \cdot \widetilde{Q}_c(2,3)$.

Further theoretical results. The optimal value of the objective function $I_R[\mathcal{A}_{disc}]$ is 2.575 according to (3.6). A Markov state model $\widetilde{P}_c(\tau) = \exp(\tau \widetilde{Q}_c)$ based on the approximation $\widetilde{Q}_c$ of Q_c only insufficiently allows for an interpretation as a transition matrix, because the membership functions χ are not crisp, they are only nearly indicator functions of sets. The value of $I_W[\mathcal{A}_{disc}]$ is 2.867. The feasible set $\mathcal{F}$ has several vertices and allows for more than one possible transformation matrix $\mathcal{A}_{disc}$ depending on the choice of the convex objective function. If $I_W[\mathcal{A}_{disc}] = 3$, then there would be (up to permutation of cluster indices) only one vertex of $\mathcal{F}$, which would lead to an optimal solution $\mathcal{A}$ independent from the choice of a convex objective function [75]. The I_R-optimal membership functions according to the 9×9-grid based expansion are plotted in the left column of Figure 5.5.

Adaptivity. As mentioned in chapter 3.4, the approximation quality of the eigenfunctions and of the eigenvalues of $\mathcal{P}$ determine the discretization error.

[2] In the case of an x-symmetric potential energy function, the following elements of Q should be identical $Q(1,2) = Q(1,3)$, $Q(2,1) = Q(3,1)$, and $Q(2,3) = Q(3,2)$.

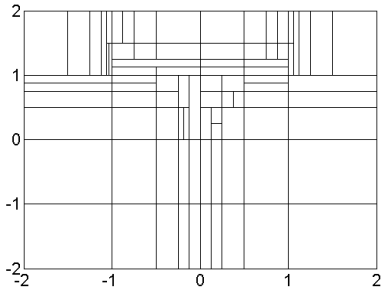

Figure 5.3: (Example 5.1) Final discretization of an adaptive refinement approach. Starting with 16 discretization boxes the algorithm terminates after a decomposition of Ω into 86 boxes.

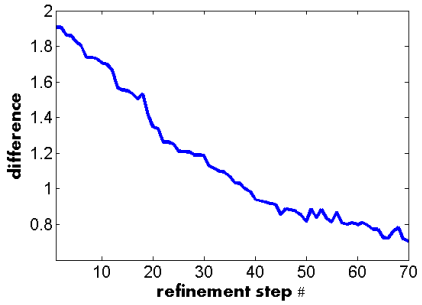

Figure 5.4: (Example 5.1) The decreasing maximal χ_{disc}-difference between neighboring boxes in the adaptive refinement approach.

Assuming the existence of an infinitesimal generator[3], the eigenfunctions of $\mathcal{P}$ are given by the eigenfunctions of $\mathcal{Q}$. The space spanned by the eigenfunctions of $\mathcal{Q}$ is identical to the space spanned by the conformations $\chi_1, \ldots, \chi_n$. This leads to a very simple refinement strategy. For a given discretization Φ, the 1-norm difference between the n-vectors $\chi_{disc}(i,:)$ and $\chi_{disc}(j,:)$ (rows of the matrix χ_{disc}) is determined if Φ_i and Φ_j are neighboring discretization boxes. The basis set Φ_i is hierarchically refined in the direction where the set Φ_j is located in Ω if the χ_{disc}-*difference* is maximal for the pair (i,j) of discretization sets. The maximal χ_{disc}-difference is a number between 0 and 2 independent of the dimension of the space Ω and independent of the number n of conformations. A corresponding adaptive decomposition of Ω for the illustrative example of the three-minima potential in Figure 5.1 is shown in Figure 5.3. Starting with $4\times4 = 16$ discretization boxes and a maximal χ_{disc}-difference of 1.9, the refinement has been terminated for a maximal χ_{disc}-difference of 0.7 and 86 discretization boxes. The decreasing maximal difference during the refinement algorithm can be seen in Figure 5.4. The eigenfunctions are approximated adaptively. The discretization is also appropriate for the approximation of the eigenvalues of the

[3]One can also assume Option 3 in chapter 3.5. In this case the eigenfunctions of $\mathcal{P}$ are the p-independent eigenfunctions of $\mathcal{L}$. They can be approximated by an adaptive decomposition of Ω.

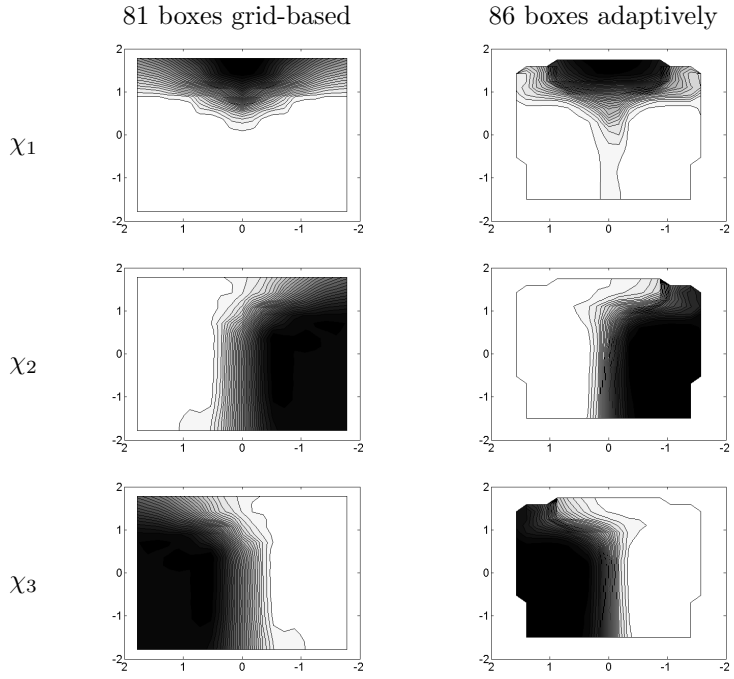

Figure 5.5: (Example 5.1) Final membership functions of the three conformations of the potential in Fig. 5.1 (white= 0, black= 1). Left column: Membership functions for a regular 9x9-grid-based decomposition of Ω. Right column: Membership functions for an adaptive decomposition of Ω. The membership functions are nearly constant inside the metastable parts. High gradients can be found in transition regions where a refinement of the basis functions is needed.

transfer operator $\mathcal{P}$, because the boxes Φ_i either have a small weight d_i or they represent a rapidly mixing part of the potential V (as described at the end of chapter 3.4). If the approximation $\widetilde{Q}_c$ is based on Gauß-Legendre quadrature, the result is

$$\widetilde{Q}_c = \begin{pmatrix} -0.002962 & 0.001408 & 0.001554 \\ 0.000007 & -0.000009 & 0.000002 \\ 0.000007 & 0.000002 & -0.000009 \end{pmatrix}. \tag{5.3}$$

The corresponding membership functions are shown in the right column of Figure 5.5. As said before, only ratios of the entries of Q_c can be interpreted physically as long as the time unit of Q_c is unknown. The ratios of the type $\widetilde{Q}_c(i,j)/\widetilde{Q}_c(j,i)$ of the elements of $\widetilde{Q}_c$ are almost identical in (5.2) and (5.3). The reason is that the statistical weights of the conformations are almost identical in (5.2) and (5.3). The ratios $\widetilde{Q}_c(2,1)/\widetilde{Q}_c(2,3)$ are different in (5.2) and (5.3), because the transition regions are better resolved in the adaptive approach. The good approximation of the transition regions in the adaptive algorithm also leads to membership functions which are more crisp than the grid-based membership functions. This fact can be seen in the optimal value of the objective function $I_R[\mathcal{A}_{disc}] = 2.811$. This value is larger than 2.575 in the grid-based case. The function I_R is maximized for an optimal transformation matrix $\mathcal{A}_{disc}$

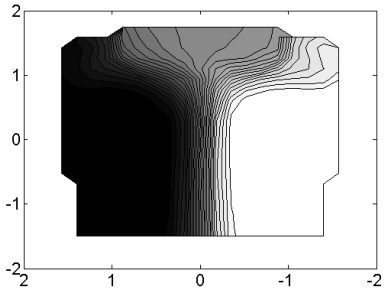

Figure 5.6: (Example 5.1) A membership function according to the PCCA+ algorithm for the case of $n = 2$ conformations and for the discretization shown in Figure 5.3.

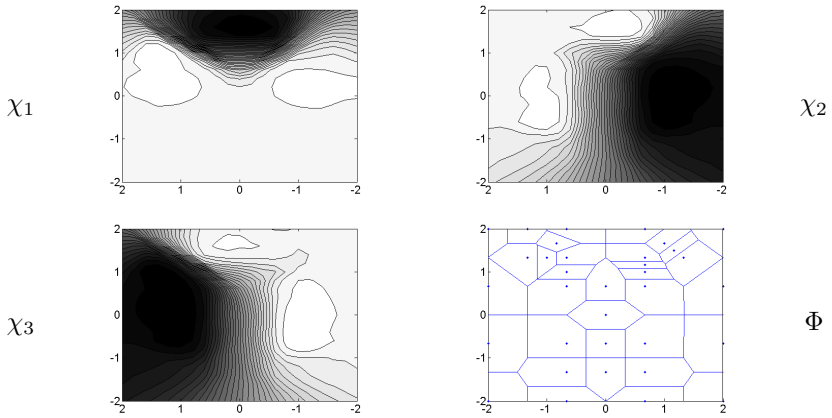

Figure 5.7: (Example 5.1) Final membership functions of the three conformations of the potential in Fig. 5.1 (white= 0, black= 1). Final Voronoi discretization Φ.

which is also optimal for the objective function $I_W[\mathcal{A}_{disc}] = 2.959 \approx 3$. The interpretation of χ_1, χ_2, and χ_3 as committor functions is possible. In chapter 2.7, committor functions are only defined for the case of $n = 2$ conformations. A resulting membership function of the PCCA+ algorithm for the given adaptive discretization and for $n = 2$ conformations is shown in Figure 5.6. A comparison of this membership function with the committor function of the potential V computed by Metzner [49] clearly shows that the two concepts (the concept of membership functions and the concept of committor functions) are strongly related.

Avoiding degeneration of Q_c. Note that the elements of the matrix $\widetilde{Q}_c$ in (5.3) are smaller than the corresponding elements of the approximated infinitesimal generator (5.2) in the grid-based approach. The reason is given by the special scaling of Q such that the matrix $M = Q + I$ is a stochastic matrix with minimal diagonal element 0. The problem is that in (4.1) the matrix elements of Q are antiproportional to the weights of the basis functions which leads

to large entries for small weighted discretization boxes. This can be a numerical problem. Note, however, that all numerical computations can be done using a matrix $\overline{Q} := DQ$ instead of Q, where D is the diagonal matrix of the weights d_i of the basis functions Φ_i. $\overline{Q}$ is a symmetric matrix with row-sum zero. Instead of using the matrix $M = Q + I$ for the scaling of Q, the matrix $\overline{M} := (\overline{Q} + I)$ can be used. The eigenvector computation can be done by solving a generalized symmetric eigenvalue problem

$$\overline{M}\,\chi_{disc} = D\,\chi_{disc}\,\Lambda.$$

In this case, the minor influence of discretization boxes with small weights can be used numerically.

Estimation of the time unit. The scaling of Q has a further consequence. The appropriate scaling factor μ for Q_c is missing. In other words, the time unit of Q_c is unknown. The aforementioned illustrative example has been constructed with 3 conformations. The reason is that a two-conformations example would be trivial. For two conformations, the matrix Q_c is always given by

$$Q_c = \mu \begin{pmatrix} -w_1^{-1} & w_1^{-1} \\ w_2^{-1} & -w_2^{-1} \end{pmatrix}.$$

Discarding μ it only depends on the weights of the conformations. The weights of the conformations, however, can be estimated by computing free energy differences. Dynamical information is not necessary. *Except for the identification of the conformations χ_1 and χ_2 in Ω, a numerical simulation of a two-conformations molecular system cannot contribute to the theoretical understanding of its kinetics, because (without knowing μ) the free energy difference between the conformations is the only information included in Q_c.* One may try to interpret the simulation data of a two-conformations system in terms of local transition rates. Note, however, that the discretized matrix Q is never appropriate for an interpretation in terms of rates according to chapter 2.6! Starting with $n = 3$ conformations, the matrix Q_c can provide interesting kinetic information without knowing the scaling factor μ, e.g. by determining ratios of outer diagonal elements of Q_c which are not of the type $Q_c(i,j)/Q_c(j,i)$. In the illustrative example it can be answered whether transitions from conformation 2 to 1 or transitions from 2 to 3 are preferred. This information cannot be derived from the statistical weights of the conformations. However, researchers may be interested in the absolute values of Q_c. In this case an estimation of μ is needed. If a theoretical understanding of experimental data is intended, then the time unit of Q_c can simply be adjusted by the experiment, see also [79]. Can μ be determined solely with theoretical methods? As shown in chapter 2.1, the transition pattern of molecular kinetics is not valid for single trajectories stemming from dynamics simulations. The transition pattern can approximately be derived from an ensemble of molecular subsystems distributed according to the χ-restricted Boltzmann distribution of states (see Option 1 in chapter 3.5). Simulating many trajectories in parallel like in the algorithm of Voter [73] can represent this ensemble of subsystems. First-exit times stemming from this simulations can be used to rescale the diagonal elements of Q_c. The relation between the diagonal elements of Q_c and the holding probabilities is used for

this rescaling. For the computation of the holding probabilities, the conformations have to be transformed into indicator functions of sets. The approach is an approximation. Furthermore, waiting for a first transition of an arbitrary trajectory in a canonical ensemble can take a lot of time. Another approach is to estimate the holding probabilities for the discretization sets $\Phi_1, \ldots, \Phi_m$ which should be much smaller than the holding probabilities of $\chi_1, \ldots, \chi_n$ for small times $t > 0$, see also [76] and [46]. This approach conceals the fact that Q is not a transition rate matrix of a Markov process. The holding probability of Φ_i is not independent from the special choice of the initial state $q \in \Phi_i$. This independence may be regarded as valid for subsets of Ω which include(!) a rapidly mixing part of the configurational space, i.e. it is valid for a union of discretization sets Φ_i with $\chi_{disc}(i, l) \approx 1$ for a given $l \in \{1, \ldots, n\}$. This ends up in a similar approach like Voter's algorithm. Nevertheless, an estimation of the time unit using a theoretical approach is always based on the realization of a molecular dynamics simulation which needs to specify the dynamical model. Furthermore, the estimation of a time unit by computing holding probabilities is always based on further assumptions (interpretation of χ as indicator functions or regarding Q as a transition rate matrix).

Density estimation approach. Using a Voronoi tessellation and an adaptive refinement in the sense of chapter 4.2 by inserting new nodes of Voronoi cells, the final discretization of the configurational space Ω and the conformations can be seen in Figure 5.7. The conformations have been computed on the basis of the density estimation approach in chapter 4.3. The statistical weights of the conformations are $w_1 = 0.0021$, $w_2 = 0.4905$, and $w_3 = 0.5075$, in good agreement with the statistical weights (first row of $\mathcal{A}_{disc}$) on page 63. The optimal value of $I_W[\mathcal{A}_{disc}] = 2.963$ allows for an interpretation of the membership functions as committor functions. The density estimation method is suitable for determining the statistical weights of the conformations. The complexity of this sampling approach is shown in the last row of Table 5.1. The Voronoi cells have an averaged number of $f_m \approx 5.37$ neighbors. If the statistical data for the overlap of neighboring Voronoi cells is not based on Gauß' Theorem but based on a restricted sampling of the Voronoi cells only, then the averaged ratio $1/r \approx 4500$ shows that the intersecting surfaces of the Voronoi cells with their neighbors have very different statistical weights. This would leed to large computational costs. Finally, a direct thermostated sampling of the potential energy function is not suitable for the computation of transition probabilities at all, the trajectory is trapped in the deeper local minima, see Figure 5.2.

5.2 Real-World Examples

Before analyzing the binding path of an APN inhibitor in chapter 5.2.2, a short overview of state of the art molecular simulations will illustrate that there is a need for a reduction of complexity in molecular simulation and that there is a chance to apply the aforementioned theoretical results in existing algorithmic approaches.

approach	complexity
unbiased sampling	no sufficient statistics after 10000 steps (Figure 5.2)
restricted sampling using overlap statistics	mean $1/r \approx 4500$
Theorem of Gauß	$f_m \approx 5.37$

Table 5.1: (Example 5.1) The complexity of different algorithmic approaches to conformation dynamics compared with a sampling of transition rates based on Gauß' Theorem according to Table 4.1 on page 58. For the decomposition of the conformational space see Figure 5.7.

5.2.1 State of the art molecular simulations

There is a growing interest in accelerated molecular dynamics simulations in the field of biomolecular simulation. In this context, the temperature-based replica exchange algorithm, also called parallel tempering, has to be mentioned [68, 25]. Replica exchange can be used to sample from the canonical ensemble of molecular systems. Since standard thermostated molecular dynamics simulations are not suitable for crossing high-energy barriers, replica exchange has been introduced for decreasing these barriers by high-temperature simulations. The decrease of free energy barriers by including high-temperature simulations mainly corresponds to a decrease of inner energy barriers of the molecular system. Entropical energy barriers are not addressed sufficiently in this context. Using this argument, Zuckermann and Lyman [86] have shown that replica exchange is not as efficient as it is supposed to be. It can (in the worst case) be less efficient than a single thermostated molecular dynamics simulation. The argument of Zuckermann and Lyman can be interpreted in the following sense. The energy barriers along the lowest energy transition paths of real-world molecular systems are, in general, not very high, but an unbiased molecular dynamics simulation does not "find" these paths (even at high temperature).

In this context, a tremendous improvement in efficiency has been achieved by "steering" a molecular dynamics simulation through the rough energy landscape. Steered molecular dynamics [39, 40] developed in the group of Schulten applies non-equilibrium molecular dynamics simulations in order to force a system to undergo a conformational change from A to B. This approach has been shown to be very efficient for bridging time-scales of molecular dynamics simulation by using biasing forces. An important theoretical basis for using steered molecular dynamics simulation in the context of equilibrium investigations is Jarzynski's identity. If the steered molecular dynamics simulation is done "step by step" as described in chapter 4.3 on page 53, then steered molecular dynamics also allows for the computation of transition rates. An adaptive approach for the construction of the corresponding Voronoi cells has been shown in this thesis. An adaptive approach is desirable: In some applications, the time-scales of binding processes are far above the time-scales of simulation, such that additional steering forces have to be artificially high in order to accelerate molecular dynamics simulations in a propriate way [39].

As described in chapter 2.3, the common basis for all molecular dynam-

ics simulations is the construction of a dynamical model. From a mathematical point of view, methods based on a trajectory (its discretized version is a Markov chain) explore a high-dimensional space by using a "ray", an one-dimensional object. Thus, the statistical results in a high-dimensional space are obtained by combining many one-dimensional explorations. In this thesis, Gauß' Theorem has been used in order to change molecular simulation from observing trajectories (simulation *along* the process) to the sampling of Voronoi cell surfaces (simulation *perpendicular* to the process). In principle, Voronoi cell surfaces are as high-dimensional as the conformational space minus 1. Thus, it should be more efficient than MD to combine statistical information of many Voronoi cells surfaces in order to explore the high-dimensional space (if numerical quadrature on Voronoi cell surfaces was trivial). Furthermore, energy barriers that occur along the MD trajectories need not occur perpendicular to it.

It has been shown for real-world examples that a discretization approach can be used in order to accelerate the sampling of the transition pattern of larger molecular systems. Chodera et al. [12] analyzed short-time molecular dynamics simulations in order to successfully derive long-time molecular processes. In this thesis it has been shown by Gauß' Theorem, that long-time processes do not need the simulation of any dynamical model. The molecular simulation problem can be transformed to a pure numerical quadrature problem combined with a function approximation problem. The latter one is solved by the discretization approach. There is a further advantage of discretization approaches which has not been mentioned yet. A common problem of all algorithms, which use a Markov chain in order to sample from the state space, is the lack of a criterion that assures the completeness of the sampling. Are there any neglected transition pathways? In the hierarchical, adaptive scheme of this thesis, in principle, there is no missing pathway. There is only a sufficient or insufficient discretization of the conformational space. The discretization is always a partition-of-unity and, therefore, includes the total space at each step of the algorithm. Thus, a rational initial discretization of the state space can assure a good resolution of the transition behavior (i.e. it assures the completeness of the approach). In practice, an initial discretization can be based on methods which do not suffer from energetic barriers of the potential energy landscape. One famous example is given by the software ConCoord [15] developed in the group of de Groot. Using the aforementioned theory combined with a sufficient initial discretization of the state space can tackle the time-scales problem of molecular simulation.

5.2.2 Binding path of an APN inhibitor

In this section, a binding path of the inhibitor bestatin of aminopeptidase N (APN) from *Escherichia coli* (ePepN, EC 3.4.11.2) will be simulated in the context of the aforementioned theoretical investigations. The results and figures are taken from Bujotzek and Weber [9]. An inhibition of human APN is a possible drug target in order to treat cancer. APN especially plays an important role in tumor progression in several human malignancies [70].

Algorithmic approach. For the simulation of a binding path of the inhibitor candidate, the software code ZIBgridfree has been used in order to generate restricted HMC samplings of Voronoi cells located in the binding pocket of APN.

The molecular structure has been taken from the Protein Data Base [7] with PDB code 2DQ6. Proteine structures were reduced to the domain harboring the binding pocket [9] and the flexibility of the protein backbone has been restricted (boundary dimensions: (42, 52, 50)Å). We applied an implicit water model with distance dependent dielectric constant. 335 soft basis functions have been sampled. For the restricted sampling, the concept of equation (4.4) for the computation of a Galerkin discretization Q of $\mathcal{Q}$ has been applied. We sampled 3 HMC chains per basis function with a maximal number of 2500 steps per chain. The Gelman-Rubin convergence check of the sampling is set to 1.2, which is the proposed value of the original paper [32]. For the proposal step a modified molecular dynamics simulation of 60 steps with a time-step of 1fs is applied. In order to derive the surface flux in (4.4), an averaged potential energy value has been computed for the overlap of the soft basis functions.

Visualization of the results. By using PCCA+ for the matrix Q, the conformations χ of the binding process can be identified. For the visualization, the sampling points of each Voronoi cell Φ_i of the sampling is weighted with the corresponding weight d_i and with the corresponding membership value $\chi_{disc}(i, l)$, where l is the index of the conformation to be visualized. Then, a volume rendering method is applied [62] in order to show the restricted Boltzmann distribution of the conformations. The binding process of the inhibitor candidate can be decomposed into several steps represented by these conformations. The steps of the binding path are visualized in Figure 5.8 by selecting one configurational state per conformation out of the sampling data. On the basis of Q one can also quantify the percentage of the flux taking place between the different conformational barriers and the statistical weights of the conformations. This is visualized in Figure 5.9.

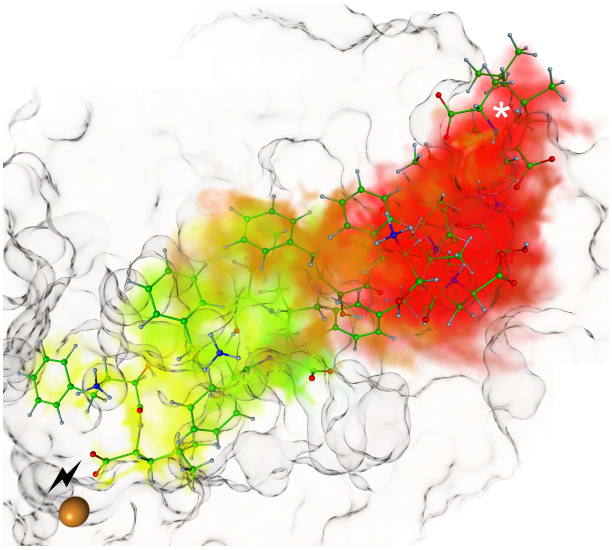

Figure 5.8: (Example 5.2.2) A binding path of the APN inhibitor bestatin from the surface of the protein into the active site. This binding path does not include the low weighted conformations in Fig. 5.9.

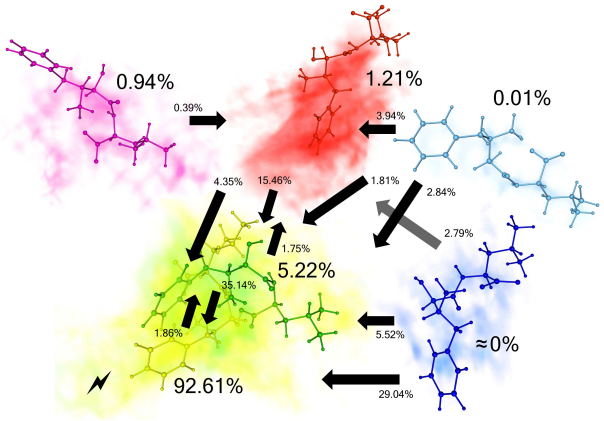

Figure 5.9: (Example 5.2.2) Percentage of the flux between the conformations and their statistical weights.

Performance of the sampling method. The sampling of the whole binding process of the APN inhibitor candidate in Figure 5.8 has been performed by ZIBgridfree on an AMD Opteron 885 computer (2×2.6 GHz) within 10 days [9]. This performance cannot be achieved with long-term molecular dynamics simulations, because there are high potential energy barriers within the binding path of the inhibitor candidate such that molecular dynamics can not lead to good statistical results (see Introduction). In other words, for the femtosecond time-scale of molecular dynamics simulations the relation $\lambda_2 \approx 1$ holds. There is an improvement in terms of computational costs for the APN simulation bewteen "restricted sampling using overlap statistics" (B) and the "Theorem of Gauß" (A). This improvement can be seen by the complexity results in Table 4.1 on page 58. In this table the expression O denotes the computationally costs of the sampling of a rapidly mixing part of the conformational space, assuming that for each sampling a comparable number of sampling points have to be generated in order to achieve a good statistics for the relevant observables. Ad (A): For the APN example in the ZIBgridfree framework, 335 Voronoi cells have been generated and corresponding restricted samplings have been performed. Using Gauß' Theorem, 335 samplings have to be performed for the estimation of the statistical weights using equation (4.10). There are 7251 neighborhood relations between the Voronoi cells (averaged number of neighbors, $f_m = 21.6$). For the simulation of fluxes between the Voronoi cells, 7251 restricted samplings have to be performed, additionally. This means, that a total number of 7586 rapidly mixing, local samplings are necessary in order to compute Q. Ad (B): If one only takes into account the statistical data from the 335 local samplings and (on this basis) tries to estimate the transition behavior by counting sampled states in overlap regions, then the ratio $1/r$ plays an important role for the determination of the computational costs. Here is a "bad" example. Using the ZIBgridfree approach, the estimated flux between Voronoi cell number 106 and Voronoi cell number 286 is almost 10^7 times larger than the flux between Voronoi cell number 106 and Voronoi cell number 9. In order to generate 1 sampling point in the overlap region between 106 and 9 one has to generate an expected number of about 10^7 sampling points for Voronoi cell number 106,

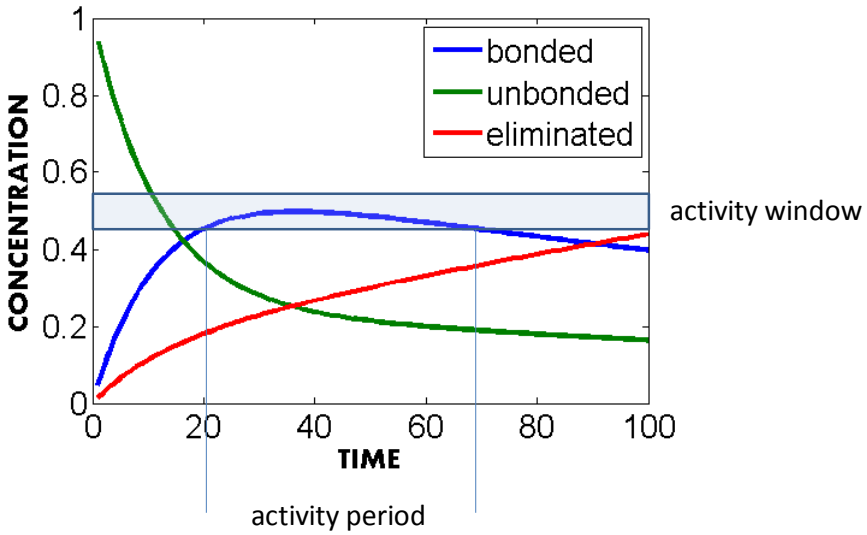

Figure 5.10: The injection of an inhibitor molecule at time zero (green curve) starts two processes in the human body: The binding process to the target molecule (blue curve) and the elimination process of the unbonded inhibitor molecule (red curve). Good drug candidates have an extended activity period. Accelerating the binding process by injecting more drug molecules is often impossible because it can reach a toxic dose.

i.e. $1/r = 10^7$. Comparing $1/r$ and f_m in Table 5.2, one can see that Gauß' Theorem can reduce the computational costs significantly. Note, that the f_m-value depends on the dimensionality of the molecular system. In the worst case, $f_m = m$.

From simulation to inhibitor design. We have to recall an important theoretical result. In contrast to a molecular dynamics simulation of the binding process of an APN-inhibitor candidate, the results of a molecular kinetics simulation do not hold for the transition pattern of single molecules. The binding process of an *ensemble* of APN-inhibitor candidates is observed. A transition is happening if an *arbitrary* inhibitor molecule in the canonical ensemble leaves a conformation A and enters the next step of the binding path. If we want to wait for the transition of a *special* inhibitor candidate of the ensemble inside a molecular dynamics simulation, this may take much more time. Thus, the barriers identified between the conformations of a molecular kinetics simulation are a forteriori barriers for the single inhibitor molecules during their own" binding process. The identification of conformations and especially the characterization of barriers during the binding process can contribute to the design of inhibitor molecules for pharmaceutical industry. It is important to accelerate the bind-

approach	complexity
unbiased sampling	no sufficient statistics
restricted sampling using overlap statistics	max $1/r \approx 10^7$
Theorem of Gauß	$f_m \approx 21.6$

Table 5.2: (Example 5.2.2) The complexity of different algorithmic approaches to conformation dynamics compared with a sampling of transition rates based on Gauß' Theorem according to Table 4.1 on page 58.

ing process of inhibitor molecules, because this can extend the activity period of the drug, see Figure 5.10. The inhibitor molecule should bind to the APN target and reach the activity window much faster than the elimination process (by metabolism) takes place. The molecular kinetics framework fits exactly to the framework of pharmacokinetic investigations and can contribute to the simulation of possible drug candidates.

The overestimation of transition rates. It has been shown that the transition rates of molecular kinetics overestimate the "transition rates" of molecular dynamics. Besides this conceptual overestimation, there is also a numerical overestimation of rates. The Galerkin discretization in (3.4) can be seen as the Rayleigh-Ritz principle of molecular kinetics. The determinant or the trace of $P_c(\tau)$ is maximized (metastability is maximized) only for the special choice of χ as a linear combination of the dominant eigenfunctions of $\mathcal{P}(\tau)$ in Theorem 2. For all conformations χ which are different from this choice (in particular for set-based conformations) the metastability is underestimated, i.e. rates are overestimated. In the numerical approach, only an *approximation* of the eigenfunctions of $\mathcal{P}(\tau)$ is computed. Thus, rates are overestimated in all numerical approaches. The optimal rates are given by Theorem 2.

Conclusion

Instead of selecting a special dynamical model, this thesis suggests to start molecular simulation with defining desirable properties of a new transfer operator. One desirable property is the existence of an infinitesimal generator. Using this assumption, the time-extrapolation error of molecular simulation vanishes. This new concept has some algorithmic consequences. Instead of generating molecular dynamics trajectories, it is only important to approximate the eigenvalues and eigenfunctions of the new infinitesimal generator, because these objects include all relevant information of the kinetics of the molecular system. Their approximation can be done within an hierarchical, adaptive, and meshless algorithmic approach. Using Gauß' Theorem, it has been shown that the simulation problem can be reformulated into solving high-dimensional quadrature problems (the generation of trajectories corresponding to a special dynamical model is not necessary). Some important algorithmic details have been derived in this thesis which lead to a very efficient algorithmic framework for molecular simulation.

Outlook. This thesis has introduced a new concept for the simulation of molecular processes. It has been shown that this new concept allows for a correct time-extrapolation and it is computationally more efficient than molecular dynamics simulations. Some problems have not been solved in this thesis. The first problem is given by the fact, that an unknown time unit μ occurs when applying Gauß' Theorem. Using molecular dynamics simulations in order to determine μ has been suggested as a workaround. This aspect has to be analyzed more precisely in future. The second problem is given by solving high-dimensional integrals efficiently. Markov Chain Monte Carlo methods have been applied in this thesis. In general, these methods need a lot of sampling points in order to solve the quadrature problem. Additionally, deterministic error estimation is not possible within their framework. The third problem is given by the complexity estimation of the new approach. The new approach is only efficient, if the averaged number of Voronoi cell neighbors (f_m) is small. It is not clear, whether it is always more efficient to estimate the statistical weights of all Voronoi cell surfaces than to simulate transitions between the cells. All these aspects have to be investigated in the future.

Bibliography

[1] E. Akhmatskaya and S. Reich. The Targeted Shadowing Hybrid Monte Carlo (TSHMC) Method. In B. Leimkuhler, C. Chipot, R. Elber, A. Laaksonen, A. Mark, T. Schlick, Ch. Schütte, and R. Skeel, editors, *New Algorithms for Macromolecular Simulation, Lecture Notes in Computational Science and Engineering*, volume 49, pages 145–158. 2006.

[2] H.C. Andersen. A "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations. *Journal of Computational Physics*, 52:24–34, 1983.

[3] I. Andricioaei, J. Straub, and A. Voter. Smart darting Monte Carlo. *J. Chem. Phys.*, 114(16):6994–7000, 2001.

[4] R. B. Bapat and T. E. S.Raghavan. *Nonnegative Matrices and Applications*. Cambridge University Press, 1997.

[5] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics*, 81(8):3684–3690, 1984.

[6] A. Berezhkovskii and A. Szabo. Ensemble of transition states for two-state protein folding from the eigenvectors of rate matrices. *J. Chem. Phys.*, 121(18):9186–9187, 2004.

[7] H.M. Berman, K. Henrick, and H. Nakamura. Announcing the worldwide Protein Data Bank. *Nature Structural Biology*, 10(12):980, 2003.

[8] B. Berne and J. Straub. Novel methods of sampling phase space in the simulation of biological systems. *Current Opinion in Structural Biology*, 7:181–189, 1997.

[9] A. Bujotzek and M. Weber. Efficient Simulation of Ligand-Receptor Binding Processes Using the Conformation Dynamics Approach. *Journal of Bioinformatics and Computational Biology*, 7(5):811–831, 2009.

[10] Ch. Chipot, M. Scott Shell, and A. Pohorille. Introduction. In Ch. Chipot and A. Pohorille, editors, *Free Energy Calculations – Theory and Applications*, pages 1–31. Springer, 2007.

[11] G. E. Cho and C. D. Meyer. Comparison of perturbation bounds for the stationary distribution of a Markov chain. *Lin. Alg. App.*, 335(1–3):137–150, 2001.

[12] J. D. Chodera, W. C. Swope, J. W. Pitera, and K. A. Dill. Long-time protein folding dynamics from short-time molecular dynamics simulations. *Multiscale Model. Simul.*, 5(4):1214–1226, 2006.

[13] L. J. Clancy. *Aerodynamics*. Pitman Publishing Limited, London, 1975.

[14] G. E. Crooks. Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *J. Stat. Phys.*, 90:1481–1487, 1998.

[15] B. L. de Groot, D. M. F. van Aalten, R. M. Scheek, A. Amadei, G. Vriend, and H.J.C. Berendsen. Prediction of protein conformational freedom from distance constraints. *Proteins*, 29:240–251, 1997.

[16] C. Dellago, P. G. Bolhuis, and P. L. Geissler. Transition Path Sampling. *Adv. Chem. Phys.*, 123:1–78, 2002.

[17] M. Dellnitz and O. Junge. On the approximation of complicated dynamical behavior. *SIAM J. Num. Anal.*, 36(2):491–515, 1999.

[18] P. Deuflhard. From molecular dynamics to conformational dynamics in drug design. In M. Kirkilionis, S. Krömker, R. Rannacher, and F. Tomi, editors, *Trends in Nonlinear Analysis*, pages 269–287. Springer, 2003.

[19] P. Deuflhard, W. Huisinga, A. Fischer, and Ch. Schütte. Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Lin. Alg. Appl.*, 315:39–59, 2000.

[20] P. Deuflhard and C. Schütte. Molecular conformation dynamics and computational drug design. In J. M. Hill and R. Moore, editors, *Applied Mathematics Entering the 21st Century. Proceedings ICIAM 2003*, pages 91–119, 2004. Invited paper.

[21] P. Deuflhard and M. Weber. Robust Perron Cluster Analysis in Conformation Dynamics. In M. Dellnitz, S. Kirkland, M. Neumann, and C. Schütte, editors, *Lin. Alg. App. – Special Issue on Matrices and Mathematical Biology*, volume 398C, pages 161–184. Elsevier, 2005.

[22] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Phys. Lett. B*, 195(2):216–222, 1987.

[23] W. E and E. Vanden-Eijnden. Towards a theory of transition paths. *J. Statist. Phys.*, 123(3):503–523, 2006.

[24] e. Schöll-Pachinger and C. Dellago. A proof of Jarzynski's nonequilibrium work theorem for dynamical systems that conserve the canonical distribution. *J. Chem. Phys.*, 125(5):054105, 2006.

[25] J. D. Earl and M. W. Deem. Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.*, 7:3910, 2005.

[26] T. S. Erp, D. Moriani, and P. G. Bolhuis. A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.*, 118:7762–7774, 2003.

[27] A. K. Faradjian and R. Elber. Computing time scales from reaction coordinates by milestoning. *J. Chem. Phys.*, 120:10880–10889, 2004.

[28] O. Forster. *Analysis 3. Integralrechnung im R^n mit Anwendungen*. Vieweg-Verlag, 3rd edition, 1996.

[29] P. L. Freddolino, F. Liu, M. Gruebele, and K. Schulten. Ten-Microsecond Molecular Dynamics Simulation of a Fast-Folding WW Domain. *Biophys J.*, 94(10):L75–L77, 2008.

[30] D. Frenkel and B. Smit. *Understanding Molecular Simulation – From Algorithms to Applications*, volume 1 of *Computational Science Series*. Academic Press, 2002.

[31] R. Funderlic and C. Meyer. Sensitivity of the stationary distribution vector for an ergodic Markov chain. *Lin. Alg. Appl.*, 76:1–17, 1986.

[32] A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Stat. Sci.*, 7:457–511, 1992.

[33] S. Geschgorin. Über die Abgrenzung der Eigenwerte einer Matrix. *Izv. Akad. Nauk. UdSSR Otd. Fiz.-Mat. Nauk*, 7:749–754, 1931.

[34] F. Haack. Properties of reduced reversible Markov chains. ZIB Report 08-42, Zuse Institute Berlin, 2008.

[35] N. Singhal Hinrichs and V. S. Pande. Calculation of the distribution of eigenvalues and eigenvectors in Markovian state models for molecular dynamics. *J. Chem. Phys.*, 126(24):244101, 2007.

[36] W. G. Hoover. *Time Reversibility, Computer Simulation and Chaos*. World Scientific, 1999.

[37] W. G. Hoover, A. Kenichiro, C. G. Hoover, and S. V. de Groot. Time-reversible deterministic thermostats. *Physica D: Nonlinear Phenomena*, 187:253–267, 2004.

[38] W. Huisinga. *Metastability of Markovian systems: A transfer operator based approach in application to molecular dynamics*. Doctoral thesis, Freie Universität Berlin, 2001.

[39] S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys J.*, 72(4):1568–1581, 1997.

[40] S. Izrailev, S. Stepaniants, B. Isralewitz, D. Kosztin, H. Lu, F. Molnar, W. Wriggers, and K. Schulten. Steered molecular dynamics. In P. Deuflhard, J. Hermans, B. Leimkuhler, A. E. Mark, S. Reich, and R. D. Skeel, editors, *Computational Molecular Dynamics: Challenges, Methods, Ideas*, volume 4 of *Lecture Notes in Computational Science and Engineering*, pages 39–65. Springer-Verlag, Berlin, 1998.

[41] C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78:2690–2693, 1997.

[42] E. T. Jaynes. Gibbs vs Boltzmann entropies. *Amer. J. Phys.*, 33:391–398, 1965.

[43] W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Cryst.*, A32:922–923, 1976.

[44] W. Kabsch. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Cryst.*, A34:827–828, 1978.

[45] M. Kijima. *Markov Processes for Stochastic Modeling*. Stochastic Modeling Series. Chapman and Hall, 1997.

[46] S. Kube and M. Weber. Conformation Kinetics as a Reduced Model for Transition Pathways. ZIB Report 05-43, Zuse Institute Berlin, 2005.

[47] S. Kube and M. Weber. A coarse graining method for the identification of transition rates between molecular conformations. *J. Chem. Phys.*, 126:024103–024113, 2007.

[48] S. Kube and M. Weber. Computation of equilibrium densities in metastable dynamical systems by domain decomposition. volume 1048 of *Numerical Analysis and Applied Mathematics*, pages 339–342. Int. Conf. on Num. Analy. and Appl. Math. 2008, AIP Conference Proceedings, Kos, 2008.

[49] Ph. Metzner. *Transition Path Theory for Markov Processes*. Doctoral thesis, Department of Mathematics and Computer Science, Freie Universität Berlin, 2008.

[50] Ph. Metzner, Ch. Schütte, and E. Vanden-Eijnden. Illustration of transition path theory on a collection of simple examples. *J. Chem. Phys.*, 125:084110, 2006.

[51] C. D. Meyer. Sensitivity of the stationary distribution of a Markov chain. *SIAM Journal on Matrix Analysis and Applications*, 15:715–728, 1994.

[52] J. Norberg and L. Nilsson. On the truncation of long-range electrostatic interactions in DNA. *Biophys J.*, 79(3):1537–1553, 2000.

[53] L. Onsager. Initial Recombination of Ions. *Phys. Rev.*, 54(8):554–557, 1938.

[54] V. S. Pande. Private communications, 2009.

[55] S. Park, M. K. Sener, D. Lu, and K. Schulten. Reaction paths based on mean first-passage times. *J. Chem. Phys.*, 119(3):1313–1319, 2003.

[56] C. R. Johnson R. A. Horn. *Matrix Analysis*. Cambridge University Press, 1990. (chapter 8).

[57] F. Ritort. Work fluctuations and transient violations of second law: Perspectives in theory and experiment. *Sem. Poincare*, 2:193–227, 2003.

[58] S. Röblitz. *Statistical Error Estimation and Grid-free Hierarchical Refinement in Conformation Dynamics*. Doctoral thesis, Department of Mathematics and Computer Science, Freie Universität Berlin, 2008.

[59] J.-P. Ryckaert, G. Cicotti, and H.J.C. Berendsen. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints; Molecular Dynamics of n-Alkanes. *Journal of Computational Physics*, 23:327–341, 1977.

[60] M. Sarich, F. Noé, and Ch. Schütte. On the Approximation Quality of Markov State Models. *In preparation, FU Berlin, Dept. of Mathematics*, 2009.

[61] T. Schlick. *Molecular Modelling and Simulation*. Springer, 2002.

[62] J. Schmidt-Ehrenberg, D. Baum, and H.-Ch. Hege. Visualizing dynamic molecular conformations. In *IEEE Visualization 2002*, pages 235–242. IEEE Computer Society Press, 2002.

[63] Ch. Schütte. *Conformational Dynamics: Modelling, Theory, Algorithm, and Application to Biomolecules*. Habilitation thesis, Department of Mathematics and Computer Science, Freie Universität Berlin, 1999.

[64] H. Senderowitz, F. Guarnieri, and W. C. Still. A smart Monte Carlo technique for free energy simulations of multiconformational molecules. Direct calculations of the conformational populations of organic molecules. *J. Am. Chem. Soc.*, 117:8211–8219, 1995.

[65] H. Senderowitz and W. C. Still. Simple but smart Monte Carlo algorithm for free energy simulations of multiconformational molecules. *J. Comp. Chem.*, 19(15):1736–1745, 1998.

[66] D. E. Shaw, M. M. Deneroff, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson, K. J. Bowers, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. P. Grossman, C. R. Ho, D. J. Ierardi, I. Kolossvary, J. L. Klepeis, T. Layman, Ch. McLeavey, M. A. Moraes, R. Müller, E. C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, and S. C. Wang. Anton, a special-purpose machine for molecular dynamics simulation. *Proceedings of the 34th annual international symposium on Computer architecture, San Diego, California, USA June 09 - 13*, pages 1–12, 2007.

[67] A. H. Stroud and D. Secrest. *Gaussian Quadrature Formulas*. Prentice-Hall, Englewood Cliffs, NJ, 1966.

[68] R. H. Swendsen and J. S. Wang. Replica Monte Carlo simulation of spin glasses. *Physical Review Letters*, 57:2607–2609, 1986.

[69] W. C. Swope and J. W. Pitera. Describing protein folding kinetics by molecular dynamics simulation. 1. Theory. *J. Phys. Chem. B*, 108:6571–6581, 2004.

[70] H. Tsukamoto, K. Shibata, H. Kajiyama, M. Terauchi, A. Nawa, and F. Kikkawa. Aminopeptidase N (APN)/CD13 inhibitor, Ubenimex, enhances radiation sensitivity in human cervical cancer. *BMC Cancer*, 8:74, 2008.

[71] E. Vanden-Eijnden and M. Venturoli. Markovian milestoning with Voronoi tesselation. *J. Chem. Phys.*, 130:194101, 2009.

[72] E. Vanden-Eijnden, M. Venturoli, G. Ciccotti, and R. Elber. On the assumptions underlying Milestoning. *J Chem. Phys.*, 129:174102, 2008.

[73] A. F. Voter. Parallel Replica Method for Dynamics of Infrequent Events. *Phys. Rev. B*, 57:13985, 1998.

[74] L. Walter and M. Weber. ConfJump: a fast biomolecular sampling method which drills tunnels through high mountains. ZIB-Report 06-26, Zuse Institute Berlin, 2006.

[75] M. Weber. *Meshless Methods in Conformation Dynamics*. Doctoral thesis, Department of Mathematics and Computer Science, Freie Universität Berlin, 2006. Published by Verlag Dr. Hut, München.

[76] M. Weber. Conformation-based transition state theory. ZIB Report 07-18, Zuse Institute Berlin, 2007.

[77] M. Weber. An efficient analysis of rare events in canonical ensemble dynamics. ZIB Report 08-36, Zuse Institute Berlin, 2008.

[78] M. Weber and K. Andrae. A simple method for the estimation of entropy differences. *Accepted for MATCH Communications in Mathematical and in Computer Chemistry*, July 2009.

[79] M. Weber, R. Becker, R. Köppen, and V. Durmaz. Classical hybrid Monte-Carlo simulation of the interconversion of hexabromocyclododecane stereoisomers. *Journal of Molecular Simulation*, 34(7):727–736, 2008.

[80] M. Weber and S. Kube. Preserving the Markov Property of Reduced Reversible Markov Chains. volume 1048 of *Numerical Analysis and Applied Mathematics*, pages 593–596. Int. Conf. on Num. Analy. and Appl. Math. 2008, AIP Conference Proceedings, Kos, 2008.

[81] M. Weber, S. Kube, L. Walter, and P. Deuflhard. Stable computation of propability densities for metastable dynamical systems. *SIAM J. Multisc. Mod.*, 6(2):396–416, 2007.

[82] M. Weber and H. Meyer. ZIBgridfree - adaptive conformation analysis with qualified support of transition states and thermodynamic weights. ZIB-Report 05-17, Zuse Institute Berlin, 2005.

[83] E. Wigner. The transition state method. *Trans. Faraday Soc.*, 34, 1938.

[84] Wikipedia. Entry: ANTON, July 22th 2009. `http://en.wikipedia.org/wiki/Anton_(computer)`.

[85] A. Yamane, F. Shimojo, and K. Hoshino. Effects of long-range interactions on the structure of supercritical fluid mercury: Large-scale molecular-dynamics simulations. *J. Phys. Soc. Jpn.*, 75, 2006.

[86] D. M. Zuckerman and E. Lyman. A Second Look at Canonical Sampling of Biomolecules using Replica Exchange Simulation. *J. Chem. Theory Comput.*, 2(4):12001202, 2006.