A coarse graining method for the identification of transition rates between molecular conformations

Susanna Kube and Marcus Weber

Zuse Institute Berlin ZIB, Takustraße 7, D-14195 Berlin, Germany

(Dated: October 19, 2006)

Abstract

The coarse graining method to be advocated in this paper consists of two main steps. First, the propagation of an ensemble of molecular states is described as a Markov chain by a transition probability matrix in a finite state space. Second, we obtain metastable conformations by an aggregation of variables via Robust Perron Cluster Analysis (PCCA+). Up to now it is an open question how this coarse graining in space can be transformed to a coarse graining of the Markov chain while preserving the essential dynamic information. In the following article we construct a coarse matrix which is the correct propagator in the space of conformations. This coarse graining procedure carries over to rate matrices and allows to extract transition rates between molecular conformations. The new idea of this approach is based on the fact that PCCA+ computes molecular conformations as linear combinations of the dominant eigenvectors of the transition matrix.

Keywords: Markov chains, soft clustering, coarse graining, master equation, molecular kinetics

I. INTRODUCTION

The main challenge in molecular modeling is to gap the different scales of length and time between computational methods and real life phenomena. Computational methods can cover the dynamic of all atoms in detail but they are restricted to short simulation times in the range of femtoseconds (fs) and to medium-sized molecules due to limited computer power. A possible way to circumvent these problems is to do coarse graining.

There are different coarse graining approaches. One can be summarized as coarse grained molecular dynamics (MD). In this approach clusters of atoms are comprised into coarse beads. The main task is then to define a force field that acts on these beads. Thus, short-term interactions are eliminated from the force field which allows MD simulations over longer periods of time. However, these models often depend on a number of parameters, and it is not clear how accurate these models are. Another coarse graining approach is based on an examination of the molecular free energy. The object of interest are conformations. They are defined from a dynamic point of view. A conformation is characterized by a specific set of geometric structures which is preserved over a long period of time before the molecule switches to another conformation. A simple example is the n-butane molecule which exhibits three conformations - the anti conformation and two gauche conformations. Since the conformations are only stable over a small period of time, they are called metastable. Conformations can often be identified with minima of the potential energy surface which leads to two different techniques. Model-based coarse graining uses numerical optimization and search methods to identify basins of attraction and transition states for the construction of transition networks[1]. Transition state theory (TST)[2–5], transition path sampling (TPS)[6–8], and transition interface sampling (TIS)[9] belong to this class of methods. However, some of these methods often suffer from the curse of dimensionality because the number of local minima is in general much larger than the number of conformations. On the other hand, data-based coarse graining uses timeseries analysis and model-independent cluster algorithms in order to detect metastabilities in the

data which are then identified with the conformations. This approach has been investigated in[10,11].

Our method is a mixed deterministic/stochastic approach which combines data-based coarse graining with efficient sampling techniques. From our point of view it is impossible to explore all physically relevant parts of the molecular state space with one single trajectory because it would get trapped in basins of attraction of the potential energy surface. Therefore, the state space is decomposed such that different regions can be sampled independently. The dynamic process is then modeled as a Markov chain in this finite state space. Since the states are situated in different regions of the sampling space, including transition states, the trapping problem does not occur. Hence the resulting Markov chain does not suffer from the problem of broken ergodicity. The Markov chain is defined by a propagator – a transition probability matrix or a transition rate matrix – which is the basis for the identification of conformations.

The coarse graining method introduced in this paper deals with a coarse graining of the propagator matrix. First, we identify metastable conformations by an aggregation of variables via Robust Perron Cluster Analysis (PCCA+)[12]. Second, we construct a coarse matrix which is used as propagator in the space of conformations. The question we are concerned with is the following: How does the coarse matrix represent the dynamic behavior described by the original matrix? We will show that the coarse matrix is the correct propagator in the space of conformations.

The coarse graining procedure is first derived for transition probabilities but it carries over to rate matrices, which can be used to extract transition rates between molecular conformations. Before clustering, the propagator contains transition probabilities within the time-scale of fs, but the coarse grained propagator describes transition probabilities and lifetimes of conformations on a timescale of picoseconds (ps) or even nanoseconds (ns) depending on the molecule to be examined. Thus, the coarse graining procedure bridges the time-lag between the simulation time and the time scale of life times of conformations. However, we cannot expect the coarse matrix to meet all

properties of a stochastic probability matrix or a rate matrix because the coarse grained process will in general not be Markovian.

Our algorithms are general enough to be applicable to systems with many degrees of freedom. Moreover, the method is not restricted to molecular dynamics. It can be applied to any dynamical system that exhibits metastability.

The paper is structured in the following way. Section II contains the fundamentals of our algorithms such as the physical model, the selection of discrete states, and the construction of a Markov chain. The clustering algorithm PCCA+ is described in a separate section because it is essential for the coarse graining. Since it has already been the subject of several previous publications, it is only shortly revisited. The main part is covered by section IV. This section is about our proposed coarse graining method which delivers a propagator acting on the conformations. The algorithms are exemplified by the application to n-pentane, a trivial but elucidating example.

II. ALGORITHMIC DETAILS

In this section we shortly describe the way how to obtain a propagator matrix in some discretized state space. This will be the starting point for the next section. A theoretical validation of the methods can be found in[11,13–16].

A. Physical Model

Throughout the paper we consider an (n, V, T)-ensemble (constant number of particles, constant volume, constant temperature) in equilibrium, i. e., the positions q and momenta p of all atoms are distributed according to the Boltzmann distribution,

$$\pi(q, p) \propto \exp(-\beta H(q, p)),$$

H denotes the Hamiltonian function which is the sum of potential energy $V(q)$ and kinetic energy $T(p)$, and $\beta = 1/k_B T$ denotes the inverse temperature with Boltzmann constant k_B.

If the spatial density $\pi(q) = \exp(-\beta V(q))$ is projected onto appropriate essential degrees of freedom, it decomposes into nearly non-overlapping partial densities which determine the metastable conformations. For example, the conformations of n-pentane, Figure 1, are characterized by two dihedral (or torsion) angles. This will be explained in detail in the following subsection. The Boltzmann density in the dihedral space looks similar to the plot in Figure 2.

B. Choice of Essential Degrees of Freedom

Since a description of molecular motion in the phase space (positions and momenta of the atoms) is infeasible for larger molecules, we have to restrict the system to essential degrees of freedom in order to analyze the data. Berendsen et al.[17] have investigated how to choose these degrees of freedom. The large scale geometric structure of molecules mainly depends on the dihedral angles. Given the positions of all atoms of a molecule one can compute the values of the dihedral angles. The reverse is not possible because there are different geometric configurations which have the same values of dihedral angles. From all-atoms MD-simulations at high temperature one can extract time series of such angles and figure out which angles exhibit metastabilities. An angle is said to be metastable if it remains near a few (mostly two or three) values over a long period of time and seldomly switches between these values. For example, n-pentane (Figure 1) can be described by two dihedral angles θ_1 and θ_2, the first one spanned by the carbon atoms $6-7-8-9$, and the second one spanned by the carbon atoms $8-7-6-1$. The angles can take all values in the interval from $-\pi$ to π. The dihedral angles involving hydrogen atoms are not essential for our analysis because they are not metastable. The same holds for bond lengths and bond angles. The space spanned by the dihedral angles will be called dihedral space. If the molecular motion was visualized as a trajectory in this space, it could be seen that it switches between the few regions of the dihedral space which are covered by the partial Boltzmann densities. These areas characterize the metastable conformations.

C. Discretization

The decomposition of the sampling space is achieved by a discretization of the dihedral space into a number of N ansatz functions $\{\phi_i\}$, $i \in \{1, \ldots, N\} = E$. In order to find the locations of these ansatz functions and to assure that our discretization covers the accessible state space we perform a global hybrid Monte Carlo sampling at high temperature which avoids the trapping problem. From the high-temperature sampling points we choose a certain number of points such that they are equally distributed w. r. t. their positions in the dihedral space. These points $\{m_i\}_{i=1}^N$ define the set of ansatz functions, for example Voronoi cells with centers $\{m_i\}_{i=1}^N$ as illustrated in Figure 3. In the case of Voronoi cells the ansatz functions become characteristic functions. Other discretizations like global functions[16] or boxes are possible as well. The ensemble of molecular geometries is described by a probability density in this finite state space E. The goal is to derive transition probabilities between these states in order to obtain information about the dynamic behavior of the molecule. Since the states of our discretization are also localized in transition regions, i. e. near saddle points of the potential energy surface, we are able to collect information about rare events. By this the problem of broken ergodicity is avoided.

D. Computation of the Propagator Matrix

The molecular motion is modeled as Markov process on the finite state space E and described by a transition probability matrix $P(t)$. The probability to be in a state $i \in E$ is given by the weighted mean value over all possible position states q w. r. t. the Boltzmann density $\pi(q)$,

$$w_i = \int_{q \in \Omega} \phi_i(q)\pi(q)\, dq. \tag{1}$$

The transition probability from a state i to a state j within a time t is computed by

$$P(t)(i,j) = \frac{1}{w_i} \int_{q \in \Omega} \phi_j(q)\phi_i(T^t q)\pi(q)\, dq, \tag{2}$$

where T^t denotes the application of an appropriate all-atom dynamic simulation. The choice of the dynamic depends on the model. One can use deterministic Hamiltonian dynamic with randomized initial momenta, which simulates a relaxation of the system after initial interaction with the heat bath, while Smoluchowski dynamic simulates the heat bath as permanent stochastic force on the molecule. In the first part of our simulation we generate states $\{q_i\}$ according to the Boltzmann distribution within each ansatz function by Hybrid Monte Carlo sampling techniques. If there is not any energy barrier within an ansatz function, such a sampling is rapidly mixing. In the other case, slow convergence indicates that the ansatz function should be refined[16]. Afterwards, the sampling points are propagated by short-time molecular dynamics simulations over a time span t (usually 40 fs). By monitoring the values of the ansatz functions at the initial points and at the end points one can construct a transition probability matrix $P(t)$ according to (2).

It is worth to mention that we cannot calculate $P(t)$ exactly due to truncation errors resulting from a finite sampling. However, by using a variance based convergence estimator[16,18] it is possible to keep the error small.

In the following we assume reversibility, i. e., there exists a stationary density $w = (w_1, \ldots, w_N)$, see (1), with

$$w^\top P(t) = w^\top,$$

such that

$$w_i \, P(t)(i,j) = w_j \, P(t)(j,i), \qquad \forall \, i,j \in E.$$

This is the famous *detailed balance equation*. Moreover, we assume that P is irreducible. Then w is the unique stationary density.

The reversibility requirement arises from the modeling point of view. Since we consider molecular dynamics in equilibrium, the system of Hamiltonian differential equations describing molecular motion is reversible. This carries over to the discretization. For detailed explanations see[16]. Furthermore, reversibility is a sufficient assumption for $P(t)$ that it can be transformed to a symmetric matrix and to have a real spectrum which ensures the correctness of the cluster algorithm PCCA+.

The diagonal matrix $D = \mathrm{diag}(w)$ can be used to transform $P(t)$ to a symmetric matrix $\overline{P}(t)$,

$$\overline{P}(t) = DP(t).$$

$\overline{P}(t)$ is the matrix of relative transition frequencies. The term "relative" refers to the scaling $\sum_{i,j=1}^{N} \overline{P}(t)(i,j) = 1$. The matrix $\overline{P}(t)$ will be the starting point for our coarse graining approach in section IV A. The reason is the following. If states are aggregated into macro-states, one can simply add the corresponding transition frequencies. This is not possible for transition probabilities because these are conditioned with the probability to be in a certain state.

Remark II.1. *If the matrix $P(t)$ stemming from the simulation does not meet the detailed balance condition, we use the modified reversible matrix*

$$P^{mod}(t) := (P(t) + D^{-1}P^{\top}(t)D)/2.$$

E. Propagation of Densities

The dynamic behavior of the molecule is described by a propagation of a probability density $x(0) \in \mathbb{R}^N$ over a time span t,

$$x(t) = P^{\top}(t)x(0).$$

If the Markov chain meets the Chapman-Kolmogorov equation, $P(t+s) = P(t)P(s)$, this can be extended to time spans nt,

$$x(nt) = P^n(t)x(0), \quad n \in \mathbb{N}, \tag{3}$$

which allows predictions about the long-time dynamic behavior.

F. From Transition Probabilities to Transition Rates

Alternatively, under certain conditions the dynamic can be characterized by transition rates, summarized in the transition rate matrix $Q \in \mathbb{R}^{N \times N}$. Both

matrices $P(t)$ and Q describe the same dynamic but from different points of view. While $P(t)$ describes the propagation of a density over a time span t, Q describes the infinitesimal changes of a given density,

$$\dot{x}(t) = Q^\top x(t) \qquad \text{(master equation)}. \tag{4}$$

If the state space is finite the transition probability matrix $P(t)$ and the rate matrix Q are related by

$$P(t) = \exp(tQ) = \sum_{n=0}^{\infty} \frac{(tQ)^n}{n!}, \quad t \geq 0. \tag{5}$$

Given the rate matrix Q one can derive transition probabilities for arbitrary times t. Q is also called the generator of the Markov chain or the generator of $P(t)$. Moreover, w is the stationary density of $P(t)$ iff $Q^\top w = 0$, see[19]. Since $\sum_j Q(i,j) = 0$ (conservation law), this is also equivalent to the balance equation

$$\sum_i w_i\, Q(i,j) = \sum_i w_j\, Q(j,i), \quad \forall\, j \in E.$$

However, in the following we even assume more. Q is supposed to be reversible, i. e., it meets the detailed balance equation

$$w_i\, Q(i,j) = w_j\, Q(j,i), \qquad \forall\, i,j \in E.$$

Then Q is diagonalizable and the transition probability matrix $P(t)$ possesses the same eigenvectors as Q. This will be important for the coarse graining process in Section IV B.

III. IDENTIFICATION OF METASTABLE CONFORMATIONS BY CLUSTERING

One might ask for the relation between the propagator $P(t)$ and the conformations we are looking for. As it was demonstrated in earlier publications[11,20], the problem of identifying metastable conformations can be reduced to an analysis of the transition matrix $P(t)$ of the corresponding Markov chain. The idea of metastability in a Markov chain is similar to the concepts of

nearly completely decomposable (NCD) or *nearly reducible* Markov chains[21,22]. States which belong to a metastable conformation are characterized by large transition probabilities among each other but small transition probabilities to other states. In the case of completely decoupled stable conformations the matrix could be rearranged to block-diagonal structure. Hence, the identification of metastable conformations corresponds to the locating of a hidden block structure in the matrix[15]. Such algorithms exist for NCD- or nearly reducible Markov chains[23–25] but they all focus on a crisp clustering. That means, the output of such cluster algorithms is a reordering of the rows and columns of the transition matrix such that states which belong to the same conformation appear in consecutive order. The resulting matrix will be nearly block diagonal.

Our approach is different. We know that there are so-called transition states, i. e. states which are located in transition zones (near saddle points) of the potential energy surface. In the dihedral space, transition states are states of the discretization from where the dynamical process can proceed to different metastable regions with equal probability. Such states cannot be assigned uniquely to one of the conformations. Therefore, we use the Robust Perron Cluster Analysis PCCA+[12] which delivers a *soft* clustering instead of a *crisp* or hard clustering, i. e., states are assigned to clusters with certain weights.

The number of clusters, n_C, corresponds to the number of eigenvalues θ_i of $P(t)$ near the Perron root $\theta = 1$,

$$P(t)X = X\Theta, \quad \Theta = \mathrm{diag}(\theta_i), \quad \theta_i \approx 1 \,\forall\, i = 1, \ldots, n_C, \quad X^\top D X = id, \quad (6)$$

where id denotes the identity matrix, and X is the eigenvector matrix $X \in \mathbb{R}^{N \times n_C}$. Note that $n_C \ll N$! The assignment of states to clusters is expressed by membership vectors $\{\chi_k\}_{k=1}^{n_C} \in \mathbb{R}^N$. The entry $\chi(i,k)$ is a number between zero and one and gives the weight with which the state i belongs to cluster k. Given the eigenvectors, PCCA+ computes a non-singular transformation matrix $A \in \mathbb{R}^{n_C \times n_C}$ such that the transformed vectors $\chi = [\chi_1, \ldots, \chi_k]$,

$$\chi = XA, \quad A \text{ non-singular}, \quad (7)$$

are non-negative and sum to one,

$$\sum_{k=1}^{n_C} \chi_k(i) = 1, \quad i = 1, \ldots, N. \tag{8}$$

A detailed description of the PCCA+ approach and the algorithm for the determination of A by solving an optimization problem can be found in[12,16].

As already mentioned, in the case of existence of a reversible generator Q, the transition probability matrix $P(t)$ possesses the same eigenvectors as Q. The eigenvalue cluster Θ of $P(t)$ at 1 transfers to an eigenvalue cluster Λ of Q at 0,

$$\Theta(t) = \exp(t\Lambda).$$

Thus, we can follow the same idea as for $P(t)$ and find χ as a linear combination of eigenvectors of Q corresponding to eigenvalues $\lambda \approx 0$, i.e.

$$\chi = XA, \quad QX = X\Lambda, \quad \Lambda = \mathrm{diag}(\lambda_i)_{i=1}^{n_C},\ \lambda_i \approx 0. \tag{9}$$

To sum up, metastable sets are convex linear combinations of all states from our discretization where the linear factors are given by the values of the membership vectors χ_k. These values can be considered as probabilities that a certain state belongs to a certain cluster. The properties of the transformation matrix A will not be essential for the coarse graining (besides the non-singularity). Just keep in mind that the membership vectors χ span the same space as the eigenvectors X.

Our coarse graining method reduces the full Markov chain with $P(t) \in \mathbb{R}^{N\times N}$ to the n_C metastable sets. The coarse grained process will be described by a propagator $P_c(t) \in \mathbb{R}^{n_C \times n_C}$.

IV. CONSTRUCTION OF COARSE-GRAINED PROPAGATORS

We are not interested in the dynamic within a metastable region which takes place on small time-scales (fs) but we ask for average life times and transition pathways to other conformations. These processes take place on larger time scales (ps). One possible way would be to propagate the density

in the space of all ansatz functions and to project the end point to the conformational space. In contrast we aim to propagate a density directly in the conformational space to obtain the same end point. The idea is illustrated in Figure 4. This requires a coarse graining of the transition matrices such that the diagram commutes. In the case of a crisp clustering the coarse graining can be done straight forward but the resulting matrix cannot be used for a correct propagation of densities. Coarse graining in the case of a soft clustering becomes more complicated but it is the soft clustering which makes it possible for the first time that the reduced matrix is a correct propagator. As far as we know this has never been examined before. Simon and Ando[21] have shown how to construct coarse matrices in the case of nearly decomposable matrices but their approach does not take into account transition states. In our case, the propagation error would be too large because the deviation of our matrices from the nearest completely decomposable matrix is not small enough.

In section IV A we demonstrate a coarse graining scheme for the transition probability matrix $P(t)$ based on a soft clustering, and we explain the use of the coarse matrix for the propagation of densities in section IV A 2. It turns out that the proposed clustering can be extended to an arbitrary number of time steps. This is the key for an extension of the coarse graining approach to transition rates in section IV B. Moreover, it will be demonstrated that our coarse-graining method covers the ideal case of an embeddable matrix $P(t)$, i. e., if Q is the generator of $P(t)$, Q_c will be the generator of $P_c(t)$. Section IV B 2 deals with the extraction of average life times and transition rates for metastable conformations from the coarse grained transition rate matrix.

A. Coarse Graining of a Transition Probability Matrix

1. Definition of Coarse Graining Operators

After having identified metastable sets we are interested in transition probabilities between these sets. For this purpose we first consider the transition

frequencies between two clusters C_k and C_l, $k, l = 1, \ldots, n_C$. For a crisp clustering, these frequencies are obtained by just adding the rows and columns of the transition frequency matrix $\overline{P}(t)$ which belong to the same cluster. This results in a coarse frequency matrix $\overline{P}_c(t) \in \mathbb{R}^{n_C \times n_C}$,

$$\overline{P}_c(t)(k,l) = \sum_{i \in C_k, j \in C_l} \overline{P}(t)(i,j).$$

In terms of membership vectors this reads

$$\overline{P}_c(t) = \chi^\top \overline{P}(t)\chi = \chi^\top D P(t) \chi.$$

The coarse stochastic matrix $P_c(t)$ is obtained by rescaling $\overline{P}_c(t)$ to row sum one,

$$P_c(t) = \tilde{D}^{-1}\overline{P}_c(t) = \tilde{D}^{-1}\chi^\top D P(t)\chi. \tag{10}$$

The coarse stochastic matrix $P_c(t)$ has the stationary density $\tilde{w} = \mathrm{diag}(\tilde{D})$ which is related to the density w via

$$\tilde{w} = \chi^\top w.$$

In the above derivation we only used the fact that the rows of χ and P sum to one. Therefore, the same method can be applied for a soft clustering. Before doing so we want to generalize the concept by introducing the following operators.

Definition IV.1. *Let $P(t) \in \mathbb{R}^{N \times N}$ be a stochastic transition probability matrix with stationary density vector $w \in \mathbb{R}^N$. Let $\chi = [\chi_1, \ldots, \chi_{n_C}] \in \mathbb{R}^{N \times n_C}$ denote the matrix of soft membership vectors and $D = diag(w)$ the matrix with w on its diagonal. With $\tilde{D} = diag(\chi^\top diag(D))$ we define the* **restriction** *and* **interpolation** *operators:*

$$R : \mathbb{R}^N \mapsto \mathbb{R}^{n_C}, \quad R = \chi^\top.$$

$$I : \mathbb{R}^{n_C} \mapsto \mathbb{R}^N, \quad I = D\chi\tilde{D}^{-1}.$$

2. Propagation of Densities

The above defined operators are used to transform the density vectors between the discretized dihedral space (fine densities) and the conformational space (coarse densities). Multiplication with R transforms the fine density $x \in \mathbb{R}^N$ into a coarse density $x_c \in \mathbb{R}^{n_C}$,

$$x_c = Rx.$$

This corresponds to an aggregation of variables, or a coarse graining of the density histogram as visualized in Figure 5. The operation preserves positivity and the sum of the vector elements. Furthermore, $\tilde{w} = R\,w$. Vice versa, the fine density can be obtained from a coarse density by interpolation,

$$x = Ix_c.$$

To make this transformation feasible, note that $I\tilde{w} = w$. That means, the interpolation operator I reconstructs the fine density as a weighted sum of partial stationary densities where the weights are given by the coarse density x_c, see Figure 5.

Now we try to answer the question that was posed in the introduction: How does the coarse stochastic matrix $P_c(t)$ represent the dynamic behavior described by the fine stochastic matrix $P(t)$?

Ideally, the diagram in Figure 4 commutes in that

$$RP^\top(t)x(0) = P_c^\top(t)Rx(0). \tag{11}$$

In other words, it would be desirable to interchange restriction and propagation. Thus, we could save a considerable amount of work by applying the propagator in the coarse space.

However, (11) is not satisfied with the coarse grained operator $P_c(t)$ as defined in (10). It only holds for special initial densities[26]. On the other hand, note that $RP^\top(t) = A^\top \Theta A^{-\top} R$ due to (6) and (7). This leads to the idea of redefining the coarse transition probability matrix as

$$\hat{P}_c(t) := (RI)^{-\top} I^\top P(t) R^\top. \tag{12}$$

It can be shown that the matrix $\hat{P}_c(t)$ defined in (12) is diagonalizable, $\hat{P}_c(t) = A^{-1}\Theta A$. Moreover, it has row sum 1 and the invariant density $\tilde{w}$. However, besides the case of a crisp clustering, $\hat{P}_c(t)$ is not a stochastic matrix. Some entries might be negative. This is due to the definition of metastable conformations because they are not disjoint sets but overlapping functions. Negative entries in the transition matrix are necessary to correct the global behavior. Nevertheless, the matrix can be used to calculate the propagation of densities by a coarse equation because it satisfies

$$RP^\top(t)x(0) = \hat{P}_c^\top(t)Rx(0). \tag{13}$$

Thus, the diagram in Figure 4 commutes.

The advantage of (13) becomes clear if we consider the extension to time steps nt according to (3). Instead of computing powers of a high-dimensional matrix $P(t)$ we can reduce the computational effort to the calculation of $\hat{P}_c^n(t) = A^{-1}\Theta^n A$ and use the equality

$$RP^{n\top}(t)x(0) = A^\top\Theta^n A^{-\top}Rx(0) = \hat{P}_c^{n\top}(t)Rx(0).$$

Thus, $\hat{P}_c(t)$ reflects the desired long-term dynamic behavior w. r. t. the propagation of densities as illustrated in Figure 6.

B. Coarse Graining of Transition Rates

The observation that propagation and restriction are interchangeable for times nt motivated us to look for a generalization of the coarse graining to arbitrary times $t > 0$. In the continuous-time case the behavior of the dynamical system is governed by its generator Q. In the following section we explain this concept and examine how the coarse graining process carries over to the rate matrix.

1. A Coarse Master Equation

We are not interested in the evolution of the densities $x(t) \in \mathbb{R}^N$ but in the evolution of the densities $x_c(t) \in \mathbb{R}^{n_C}$ w. r. t. metastable sets. Our goal is to

define a **coarse master equation**

$$\dot{x}_c(t) = Q_c^\top x_c(t), \quad Q_c \in \mathbb{R}^{n_C \times n_C}, \tag{14}$$

such that $||Rx(t) - x_c(t)||$ is small for all $t > 0$. Similar to (12), we define a coarse matrix

$$\hat{Q}_c^\top = RQ^\top I(RI)^{-1}. \tag{15}$$

It can be shown[26] that the matrix $\hat{Q}_c$ defined in (12) is diagonalizable

$$\hat{Q}_c = A^{-1}\Lambda A, \tag{16}$$

it has row sum 0 and the stationary density $\tilde{w}$.

Note that $\hat{Q}_c$ can have some negative off-diagonal elements. The interpretation is the same as for $\hat{P}_c$ in section IV A. However, in the following we still speak of a coarse rate matrix.

Theorem IV.2. *Let be given a reversible rate matrix Q with master equation $\dot{x}(t) = Q^\top x(t)$, $x(0) = x_0 \in \mathbb{R}^N$, and the corresponding restriction and interpolation operators defined in IV.1. The result $x_c(t)$ of the coarse master equation (14) with $x_c(0) = Rx_0$ and $\hat{Q}_c^\top$ defined by (15) satisfies*

$$||Rx(t) - x_c(t)|| = 0, \, \forall \, t \in [0, \infty).$$

Proof. Note that $x(t) = \exp(tQ^\top)x(0)$ and $x_c(t) = \exp(t\hat{Q}_c)x_c(0)$. Consequently,

$$||Rx(t) - x_c(t)|| = ||(R\exp(tQ^\top) - \exp(t\hat{Q}_c^\top)R)x_0||.$$

Moreover, the following equation is satisfied,

$$\begin{aligned} R\exp(tQ^\top) &= A^\top X^\top \exp(tQ^\top) \\ &= A^\top \exp(t\Lambda)X^\top \qquad \text{(eq. (9))} \\ &= A^\top \exp(t\Lambda)A^{-\top}R. \end{aligned}$$

From (16) we obtain

$$\exp(t\hat{Q}_c^\top) = \exp(tA^\top \Lambda A^{-\top}) = A^\top \exp(t\Lambda) A^{-\top}. \tag{17}$$

This yields the proposition. □

The meaning of the theorem becomes clear from Figure 7. Since

$$R \exp(tQ^\top) x(0) = \exp(t\hat{Q}_c^\top) R x_0,$$

the diagram commutes, and the coarse matrix $\hat{Q}_c$ is the correct propagator in the conformational space. Thus, for the examination of the long-term dynamic behavior it is not necessary to compute the exponential of a – possibly large – matrix Q but it suffices to compute the exponential of the small matrix $\hat{Q}_c$ which is given by (17).

There is another improvement of the proposed coarse graining scheme. If the coarse matrices were defined analogous to the crisp clustering as in (10), it would hold in general

$$P_c(t) \neq \exp(tQ_c).$$

Thus, it is not clear how to derive a coarse rate matrix in this case. Furthermore, the solution of the coarse equation cannot be related directly to the solution of the corresponding fine equation.

On the other hand, the modified coarse graining method is consistent. Comparing (12) and (15) it can be seen that

$$\hat{P}_c(t) = \exp(t\hat{Q}_c).$$

Note that even if the matrix $P(t)$ has no unique logarithm due to negative or complex eigenvalues in the lower part of the spectrum, the logarithm of the coarse matrix $\hat{P}_c(t)$ is unique, if the dominant eigenvalues are real and non-defective. Thus, it may be possible to derive a generator $\hat{Q}_c = \log(\hat{P}_c(t))/t$ for the coarse grained dynamic process although the generator for the fine process might not exist.

2. Extraction of Kinetic Information

We are not necessarily interested in solving (14) but in the information we gain from $\hat{Q}_c$. In the following assume that the sample paths of $\{X(t)\}$ are right-continuous step functions. Such processes are also called Markov jump processes[19]. Then the entries of the rate matrix Q can be related to the mean holding times within the states and transition rates between different states. The holding time T_i in state i for a stochastic process $\{X(t)\}$ is defined as the random variable

$$T_i(t) = \inf(s \geq 0 : X(t+s) \neq i, X(t) = i).$$

If $X(0) = i$, then $X(T_i)$ with $T_i = T_i(0)$ represents the state the Markov chain visits just after leaving state i. It can be shown[27] that T_i decays exponentially,

$$\mathbb{P}[T_i > s] = \exp(-h(i)s), \qquad \forall\, s > 0.$$

$h(i)$ is called the *jump rate* associated with state $i \in E$. The average life time of state i is given by

$$\mathbb{E}[T_i] = \frac{1}{h(i)}.$$

As shown in[27,28], the entries of a rate matrix Q can be related to some conditioned transition probabilities and the jump rates via

$$-\frac{Q(i,j)}{Q(i,i)} = \mathbb{P}[X(T_i) = j | X(0) = i], \quad \text{and} \quad Q(i,i) = -h(i). \tag{18}$$

Thus, the rate matrix Q can be represented by the inverse average life times $H = \mathrm{diag}\{h(i)\}$ and some transition matrix K,

$$Q = H(K - id),$$

where $K(i,j) = Q(i,j)/h(i)$ is the conditional probability of a transition from state i to state j given that the process starts in i. K describes the *embedded* Markov chain. This characterization forms a basis for a numerical simulation of the Markov jump process. Assume we have already generated points according to the Boltzmann distribution within each cell of our discretization by hybrid Monte Carlo sampling. Now, instead of propagating each point over

a predefined time step t we propagate each point until it leaves the starting cell. We collect the information about the time this process takes and the cell the point goes to in order to estimate the average life time $\mathbb{E}[T_i]$ and the conditioned transition probabilities $K(i,:)$. This gives the matrix Q via (18).

The other way round, from $\hat{Q}_c$ obtained by coarse graining one can reconstruct $\hat{H}_c$ and $\hat{K}_c$ which contain the information about mean holding times and conditioned transition probabilities for the coarse states or conformations.

V. APPLICATION TO N-PENTANE

We present the application to the n-pentane molecule $CH_3(CH_2)_3CH_3$ which was modeled with Merck Molecular Force Field[29,30] at a temperature of $300K$. The rate matrix Q resulted from a conformation dynamics simulation with ZIBgridfree[16,31], a program package based on mesh-free methods which was developed at Zuse Institute Berlin. The dynamic behavior of pentane is described by the two torsion angles θ_1 and θ_2. The 2-dimensional dihedral space was discretized by 100 Voronoi cells, see Figure 3. The center of these cells were obtained by a selection of 100 sampling points from a high temperature hybrid Monte-Carlo pre-sampling at $1000K$. Within each cell we generated nearly 3000 points according to the Boltzmann distribution by hybrid Monte-Carlo sampling with umbrella strategies[16,32] and Gelman-Rubin convergence indicator[18]. These points were propagated by molecular dynamics simulations until they left their starting cells. Average life times and conditioned transition probabilities were computed to set up the rate matrix Q via (18).

To obtain the conformations we applied PCCA+. There are 9 eigenvalues of Q close to 0,

$$\begin{aligned}\{\lambda_i\}_{i=1}^{9} \;=\; \{&2.6e-9,\ -2.2e-3,\ -3.0e-3,\ -3.7e-3,\ -4.7e-3,\\ &-5.7e-3,\ -6.8e-3,\ -8.9e-2,\ -9.8e-2\},\end{aligned}$$

followed by a gap to the 10th eigenvalue $\lambda_{10} = -0.70$. This corresponds to 9 metastable conformations which can be distinguished according to the

orientation of one of the two dihedral angles (g± and t denote the ± gauche and trans orientations), see Figure 8:

$$\text{conf.} = \{g+/g+, g-/g-, g+/g-, g-/g+, g-/t, t/g-, t/g+, g+/t, t/t\}$$

with weights

$$\{w_i\}_{i=1}^{9} = \{0.0036, 0.0032, 0.0640, 0.0680, 0.1248, 0.1232, 0.1140, 0.1567, 0.3426\}.$$

In order to compute the membership functions $\{\chi_i\}_{i=1}^{n_C}$ we calculated the eigenvectors of the transition probability matrix $P(t)$ and a transformation matrix A. The transformation matrix A was computed with the Inner Simplex Algorithm[33,34] which generates an initial guess for the optimization problem. The solution satisfies the non-negativity condition only approximately (smallest entry of χ is -0.0808) but it still meets (8) which is essential for the coarse-graining process. From the coarse rate matrix computed according to (15) we obtained the approximated mean holding times

$$\{h_i^{-1}\} = \{10.39, 11.38, 361.68, 334.35, 198.91, 227.31, 172.40, 235.74, 283.64\}ps,$$

and the transition probabilities of the embedded Markov chain

$$\hat{K}_c = \begin{pmatrix} 0 & 0.0000 & 0.0097 & 0.0105 & 0.0002 & -0.0002 & 0.5690 & 0.4045 & 0.0062 \\ 0.0000 & 0 & 0.0149 & 0.0098 & 0.4288 & 0.5403 & -0.0003 & 0.0009 & 0.0055 \\ 0.0172 & 0.0206 & 0 & 0.0000 & 0.0003 & 0.4502 & 0.0020 & 0.5089 & 0.0010 \\ 0.0148 & 0.0137 & -0.0001 & 0 & 0.5042 & 0.0020 & 0.4591 & 0.0048 & 0.0012 \\ 0.0003 & 0.1687 & -0.0001 & 0.1634 & 0 & -0.0062 & 0.0017 & 0.1881 & 0.4842 \\ 0.0000 & 0.3049 & 0.1462 & 0.0007 & -0.0337 & 0 & 0.1161 & 0.0023 & 0.4634 \\ 0.2898 & 0.0001 & 0.0003 & 0.1410 & 0.0011 & 0.0951 & 0 & -0.0164 & 0.4889 \\ 0.2184 & 0.0002 & 0.1355 & 0.0010 & 0.1778 & 0.0018 & -0.0243 & 0 & 0.4899 \\ 0.0018 & 0.0015 & 0.0000 & 0.0002 & 0.2518 & 0.2076 & 0.2675 & 0.2695 & 0 \end{pmatrix}.$$

Observe that conformation 3 and 4 have large mean holding times even though they have small weights. This hints to the fact that there are large energy barriers to the other conformations.

The rate matrix was used for a kinetics simulation with predefined start and end conformation. In general, every initial density would converge to the stationary density by (14). However, by permanently eliminating a desired product out of the ensemble, one can push the process into the direction of

this product. This corresponds to *Le Chatelier's Principle*. Mathematically this can be done by a projection of the density onto the orthogonal complement of the desired end point. For details see[26]. The results for a $(g+/t) \rightharpoonup (t/g+)$ transition of pentane are shown in Figure 9. It can be interpreted as follows. During the conformational change from (g+/t) to (t/g+)-pentane other conformations are formed which can be seen as transition states. The transition is visualized in Figure 10. The left picture shows the start conformation (g+/t), the right one the end conformation (t/g+). At each time step of the kinetics simulation a similar density plot can be computed. The picture in the middle shows the transition state at 1000ps simulation length. It is very similar to the t/t-conformation and can be considered as the intermediate distribution of states at this particular time. Thus, the method can be used to obtain possible transition pathways for other conformational changes and gives information about the duration of such transformations.

VI. CONCLUSION

The complexity of molecular kinetics can be reduced significantly by a restriction to metastable conformations which are almost invariant sets of molecular dynamical systems. The main goal is to describe the dynamic as a Markov process on these conformations. Ideally, the coarse graining process results in a small transition matrix which has the correct stationary density and reflects the essential dynamic behavior. However, due to discretization errors or the lack of metastabilities in the dynamical system it is not possible to meet all requirements at the same time. While previous articles aimed to construct coarse matrices with certain structural properties, we take a different point of view and construct the coarse matrices such that they reflect the correct dynamic behavior. This is possible because the clustering method PCCA+ characterizes the conformations as linear combinations of the eigenvectors of the unreduced transition matrix. On this basis we defined a restriction and interpolation operator which are used for density transformations. However, our coarse matrices do not always admit a physical interpretation due to negative

entries. But if there really exists a hidden low-dimensional Markov process in the model, these entries are small. Thus, it is possible to derive mean holding times and conditioned transition probabilities between the clusters from the coarse grained matrix. Furthermore, we have shown how to generate and interprete ensembles of transition pathways by the propagation of densities in the the conformational space.

Acknowledgment The authors like to thank P. Deuflhard for his suggestions concerning mesh free methods, as well as W. Huisinga for his illustrative lecture on Markov processes. Furthermore, thanks to J. Schmidt-Ehrenberg and A. Riemer for their assistance in programming in *amira*.

---

[1] M. T. Wolfinger, W. A. Svrcek-Seiler, C. Flamm, I. L. Hofacker, and P. F. Stadler, J. Phys. A: Math. Gen. **37**, 4731 (2004).

[2] H. Eyring, J. Chem. Phys. **3**, 107 (1935).

[3] E. Wigner, Trans. Faraday Soc. **34**, 29 (1938).

[4] D. G. Truhlar and B. C. Garrett, Ann. Rev. Phys. Chem. **35**, 159 (1984).

[5] E. Vanden-Eijnden and F. A. Tal, J. Chem. Phys. **123**, 184103 (2005).

[6] D. Chandler, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. Berne, G. Ciccotti, and D. Coker (Singapure: World Scientific, 1998), pp. 51–66.

[7] C. Dellago, P. G. Bolhuis, and P. L. Geissler, in *Advances in Chemical Physics*, edited by I. Prigogine and S. A. Rice (Wiley, 2002), vol. 123, chap. 1, pp. 1–78.

[8] G. Hummer, J. Chem. Phys. **120**, 516 (2004).

[9] T. S. van Erp, D. Moroni, and P. G. Bolhuis, J. Chem. Phys. **118**, 7762 (2003).

[10] P. Deuflhard, in *Trends in Nonlinear Analysis*, edited by M. Kirkilionis, S. Krömker, R. Rannacher, and F. Tomi (Springer, 2003), pp. 269–287.

[11] C. Schütte and W. Huisinga, in *Handbook of Numerical Analysis*, edited by P. G. Ciarlet and C. L. Bris (North–Holland, 2003), vol. X. Special Volume Computational Chemistry, pp. 699–744.

[12] P. Deuflhard and M. Weber, in *Lin. Alg. App. – Special Issue on Matrices and Mathematical Biology*, edited by M. Dellnitz, S. Kirkland, M. Neumann, and C. Schütte (Elsevier Journals, 2005), vol. 398C, pp. 161–184.

[13] C. Schütte, W. Huisinga, and P. Deuflhard, in *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems.*, edited by B. Fiedler (Springer, 2001), pp. 191–223, preprint. Available via http://www.math.fu-berlin.de/~biocomp.

[14] C. Schütte, *Conformational Dynamics: Modelling, Theory, Algorithm, and Application to Biomolecules* (Habilitation Thesis, Department of Mathematics and Computer Science, Free University Berlin, 1999).

[15] P. Deuflhard, W. Huisinga, A. Fischer, and C. Schütte, Lin. Alg. Appl. **315**, 39 (2000).

[16] M. Weber, Doctoral thesis, Department of Mathematics and Computer Science, Freie Universität Berlin (2006), published by Verlag Dr. Hut, München.

[17] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen, Proteins **17**, 412 (1993).

[18] A. Gelman and D. B. Rubin, Statistical Sience **7**, 457 (1992).

[19] P. Brémaud, *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*, no. 31 in Texts in Applied Mathematics (Springer-Verlag New York, 1999).

[20] P. Deuflhard and C. Schütte, in *Applied Mathemetics Entering the 21st Century. Proceedings ICIAM 2003*, edited by J. M. Hill and R. Moore (2004), pp. 91–119.

[21] H. A. Simon and A. Ando, Econometrica **29**, 111 (1961).

[22] C. D. Meyer, SIAM Rev. **31**, 240 (1989).

[23] W. J. Stewart and W. Wu, ORSA J. Comput. **4**, 336 (1992).

[24] R. J. Tarjan, SIAM J. Comp. **1**, 146 (1972).

[25] T. Dayar, Tech. Rep. BU-CEIS-9808, Department of Computer Science and Engineering, Bilkent University (1998).

[26] S. Kube and M. Weber, ZIB-Report 06-35, Zuse Institute Berlin (2006).

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

[28] S. Kube and M. Weber, ZIB-Report 05-43, Zuse Institute Berlin (2005).

[29] T. Halgren, J. Am. Chem. Soc. **114**, 7827 (1992).

[30] T. Halgren, J. Comp. Chem. **17**, 490 (1996).

[31] M. Weber and H. Meyer, ZIB-Report 05-17, Zuse Institute Berlin (2005).

[32] G. Torrie and J. Valleau, J. Chem. Phys. **66**, 1402 (1977).

[33] M. Weber and T. Galliat, ZIB-Report 02-12, Zuse Institute Berlin (2002).

[34] M. Weber, W. Rungsarityotin, and A. Schliep, in *From Data and Information Analysis to Knowledge Engineering*, edited by M. Spiliopoulou, R. Kruse, C. Borgelt, A. Nürnberger, and W. Gaul (Springer-Verlag Berlin · Heidelberg, 2005), Studies in Classification, Data Analysis and Knowledge Organization, pp. 103–110.

[35] J. Schmidt-Ehrenberg, D. Baum, and H.-C. Hege, in *IEEE Visualization 2002*

(IEEE Computer Society Press, 2002), pp. 235–242.

[36] D. Stalling, M. Westerhoff, and H.-C. Hege, in *Visualization Handbook*, edited by C. R. Johnson and C. D. Hansen (Academic Press, 2004).

Figure caption list

FIG. 1. Balls-and-stick-representation of pentane with atom indices. Larger balls correspond to carbon atoms, smaller balls to hydrogen atoms.

FIG. 2. Histogram of the Boltzmann density of n-pentane in the dihedral space obtained from a simulation with ZIBgridfree at 300K. The histogram was obtained by dividing the dihedral space $[-\pi, \pi]^2$ into 72×72 boxes.

FIG. 3. Discretization of the dihedral space of pentane by Voronoi cells.

FIG. 4. Coarse graining scheme for transition probabilities. Instead of propagating an initial density $x(0)$ in a high-dimensional space by $P(t)$ and restricting the result $x(t)$ to the coarse variable $x_c(t)$, we aim to construct a propagator $P_c(t)$ which acts directly in the coarse space on a coarse grained initial density $x_c(0)$.

FIG. 5. Density vectors in different state spaces, stemming from a discretization of n-pentane.

FIG. 6. Extension of the coarse graining scheme to time steps nt, $n \in \mathbb{N}$.

FIG. 7. Coarse graining scheme for transition rates.

FIG. 8. 2d schematic plot of the Boltzmann density of pentane w. r. t. the two dihedral angles θ_1 and θ_2. The conformations are numbered from 1 to 9.

FIG. 9. Conformation kinetics simulation of n-pentane. The 9 lines correspond to the weights of the 9 clusters. The (g+/t)-conformation was selected as start conformation (solid line) and the (t/g+)-conformation as end conformation (dashed line). The density is plotted over an interval of 6000ps at every 100ps.

FIG. 10. Volume rendering of the (g+/t)-conformations of pentane (left), the (t/g+)-conformation (right), and the corresponding transition macro-state (middle) in *amira/amiraMol*[35,36].

Figures

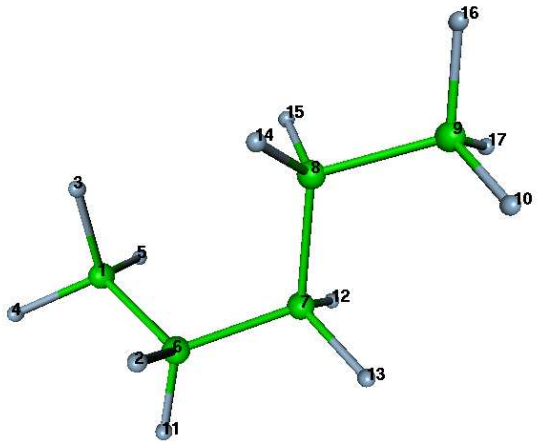


FIG. 1:

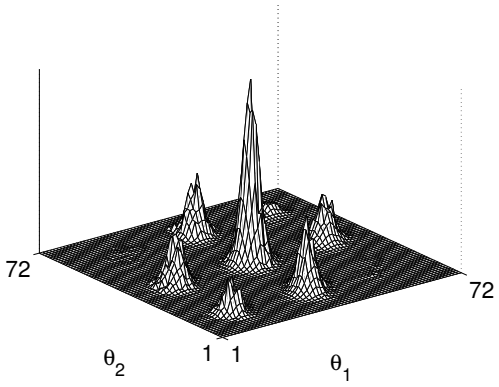


FIG. 2:

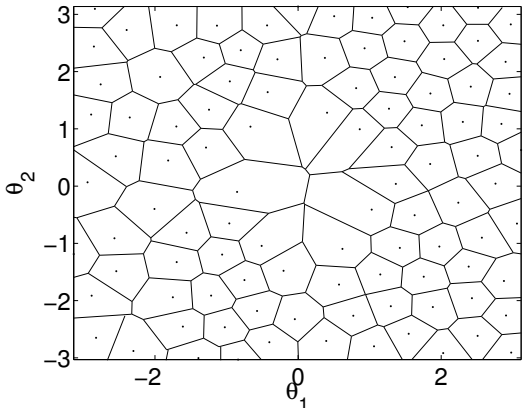


FIG. 3:

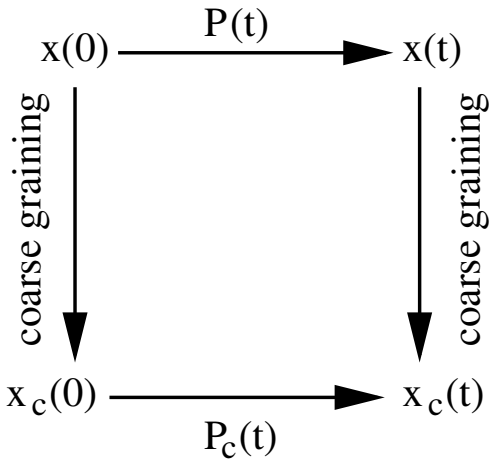
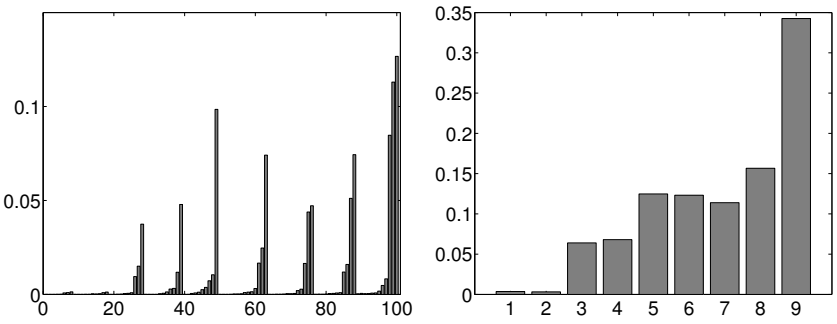


FIG. 4:



(a) Stationary density x in the dihedral space discretized by 100 cells.

(b) Stationary density x_c in the conformational space consisting of 9 conformations.

FIG. 5:

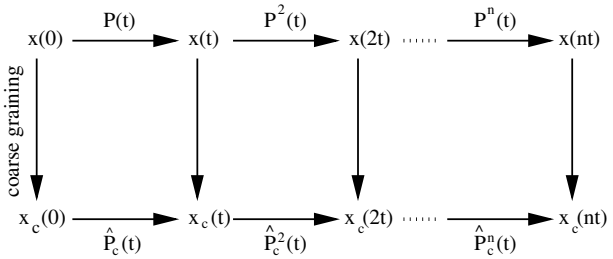

x(0)
P(t)
x(t)
P^2(t)
x(2t)
P^n(t)
x(nt)
coarse graining
x_c(0)
x_c(t)
x_c(2t)
x_c(nt)
P̂_c(t)
P̂_c^2(t)
P̂_c^n(t)

FIG. 6:

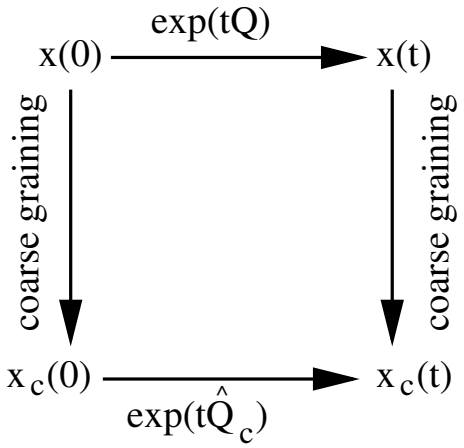


FIG. 7:

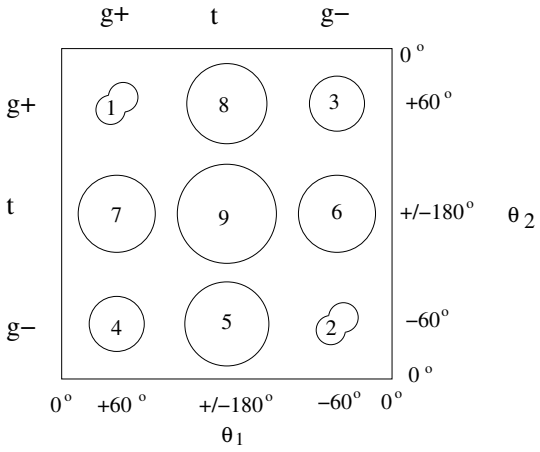


FIG. 8:

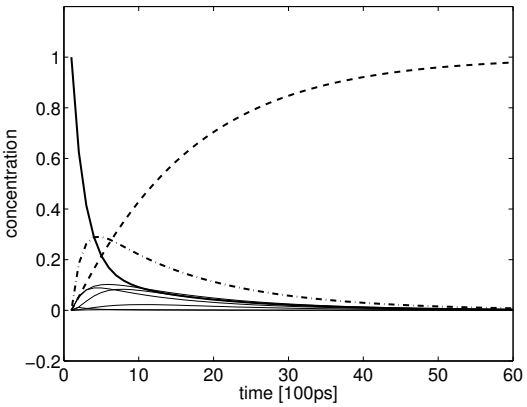


FIG. 9:

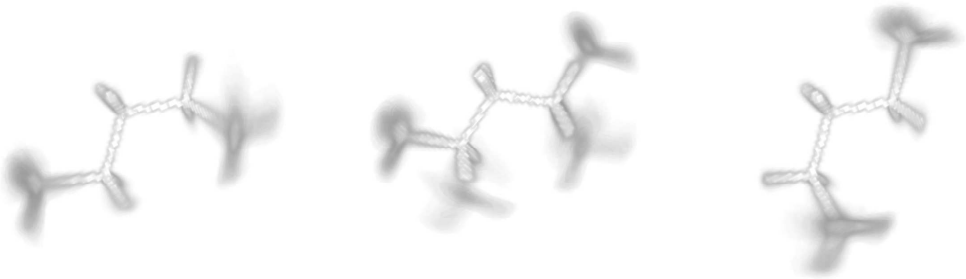


FIG. 10:

*