Conformation Dynamics*

Christof Schütte[1,†], Frank Noe[1], Eike Meerbach[1],
Philipp Metzner[1,2], and Carsten Hartmann[1,3]

[1] *Institut für Mathematik, Freie Universität Berlin*
Arnimallee 6, 14195 Berlin, Germany

[2] *Courant Institute of Mathematical Sciences, New York University*
251 Mercer Street, New York, NY 10012, USA

[3] *Mathematics Institute, University of Warwick*
Coventry CV4 7AL, UK

Abstract

This article surveys the present state of the transfer operator approach to the effective dynamics of metastable dynamical systems and the variety of algorithms associated with it.

Keywords: Metastable states, Molecular dynamics, Biomolecular conformations, Hidden Markov model, Maximum-likelihood principle, Protein folding, Spectroscopy.

Introduction

Conformational transitions are critical to the function of proteins and nucleic acids. These transitions span large ranges of length and time scales and include ligand binding [1], complex conformational rearrangements between native protein substates [2, 3], and folding [4, 5]. Understanding these processes is challenging as they often involve various pathways via many intermediate conformations. A particular feature of biomolecular systems is *metastability* which denotes their property of being localized in a certain region of phase space for rather long period of times before rapidly moving to another region in which the dynamics then, again, stays for a very long time [6, 7, 17]. Here "long" is meant with respect to the characteristic time scale of the system, e.g., the typical duration of a molecular bond oscillation.

The purpose of this article is to survey and extend available methods by which one can identify metastable states in biomolecular time series and estimate transition probabilities between them. Both the identification of metastable states and the analysis of the transitions rely on the mathematical concept of the *transfer operator* that is associated with the Markovian dynamics and which,

*Supported by the DFG research center "Mathematics for key technologies" MATHEON (FZT86) in Berlin.

†To whom correspondence should be addressed: *schuette@mi.fu-berlin.de*

after a suitable coarse-graining, encodes the desired information (as it was first shown in [23, 19]). We describe the transfer operator approach following a three-step procedure. First of all, we introduce metastability as a hierarchical concept in which the appropriate number of metastable states depends upon the degree of spatial and temporal resolution that is to be achieved (Section 1). Secondly, we will give precise mathematical statements about how to cluster a time series optimally into metastable states; this involves the problem of how to identify metastable sets at all. For moderate state space dimensions the transfer operator can be computed explicitly by clever discretization and the number of metastable states is determined by the number of distinguished eigenvalues close to one (Perron cluster). The metastable subsets are then determined by the corresponding eigenfunctions. In actual applications a discrete version of the transfer operator is estimated from a molecular dynamics trajectory (Section 2). In addition to the error due to discretization of state space, the transfer operator carries a statistical uncertainty due to finite data which renders the corresponding eigenvalues and eigenfunctions to be inaccurate. We propose a Monte-Carlo method that allows for sampling the variance of the estimated transfer operator thereby providing estimators for the statistical error of the eigenvalues and -functions. Knowledge of the statistical error may moreover be exploited to optimally launch further simulations as to reduce the uncertainties in the observables of interest [8, 9, 10, 11]. If the problem's dimension is high, the transfer operator can no longer be directly computed, which leads over to the third topic (Section 3): if not the full state space can be discretized, we can employ Hidden Markov Models (HMM) that, to some extend, account for the missing information due to neglected degrees of freedom. The HMM method turns out to be extremely powerful in identifying metastable states and computing transition probabilities, as we combine it with dynamical output that comes in form of stochastic differential equations (HMMSDE). The discretization of the transfer operator based on incomplete information amounts to a very coarse discretization; hence HMM assumes that the dynamics between the (hidden) coarse-grained states is still Markovian, i.e., its transition probabilities depend only on the current state but not on the system's history.

Once metastable sets have been identified a typical problem consists in the computation of the respective transition rates or transition pathways. To this end we introduce the basic concepts of Transition Path Theory (TPT, Section 4). The objective of TPT is to analyse the ensemble of reactive trajectories between metastable sets thereby allowing, e.g., for a calculation of transition rates. We illustrate all basic theoretical statements throughout this article with a small molecular example and conclude the discussion by studying the folding dynamics of a biophysically relevant protein (Section 5).

1 Metastability

Throughout this article we study *homogeneous Markov processes* $X_t = \{X_t\}_{t\in\mathbf{T}}$ on a state space $\mathbf{X} \subset \mathbf{R}^n$, where $\mathbf{T}$ is either continuous or discrete. The dynamics of X_t is given by the stochastic transition function

$$p(t, x, A) = \mathbf{P}[X_{t+s} \in A \,|X_s = x]\,, \tag{1}$$

for every $t, s \in \mathbf{T}$, $x \in \mathbf{X}$ and $A \subset \mathbf{X}$. We write $X_0 \sim \mu$, if the Markov process X_t is initially distributed according to the probability measure μ, i.e., $\mathbf{P}[X_0 \in A] = \mu(A)$ for all measurable subsets $A \subset \mathbf{X}$. We denote by $\mathbf{P}_\mu[\cdot]$ the probability measure that is induced by X_t, $X_0 \sim \mu$ on $\mathbf{X}$. The transition function satisfies the Chapman-Kolmogoroff equation [12]

$$p(t+s, x, A) = \int_{\mathbf{X}} p(s, y, A)\, p(t, x, \mathrm{d}y)\,,$$

and we say that the process X_t admits an invariant probability measure μ, if

$$\int_{\mathbf{X}} p(t, x, A)\mu(\mathrm{d}x) = \mu(A)\,.$$

In the following we shall always assume that the invariant measure of the process exists and is unique. A Markov process is called reversible with respect to μ, if

$$\int_A p(t, x, B)\mu(\mathrm{d}x) = \int_B p(t, x, A)\mu(\mathrm{d}x)$$

for every $t \in \mathbf{T}$ and $A, B \subset \mathbf{X}$. If moreover $p(t, x, \cdot)$ is absolutely continuous with respect to Lebesgue measure, then we denote by $p(t, x, y)$ the associated *flat-space transition density*, i.e., we have

$$p(t, x, A) = \int_A p(t, x, y)\mathrm{d}y\,.$$

1.1 Transition probabilities and Transfer Operators

Metastability of some subset of the state space is characterized by the property that the dynamics is likely to remain inside this subset for a long period of time before it eventually exits. In the literature, there are various related but yet different definitions of metastability, e.g., [13, 14, 15, 16]); cf. also [17].

In this article we will focus on an ensemble-based concept as will be outlined below and is described in detail in, e.g., [17]. The objective is to find an optimal *decomposition of the state space into metastable subsets* and the "hopping dynamics" between these subsets. Specifically, a decomposition $\mathbf{D} = \{D_1, \ldots, D_m\}$ of the state space $\mathbf{X}$ consists of a collection of subsets $D_k \subset \mathbf{X}$ with the following properties: (1) positivity, i.e., $\mu(D_k) > 0$ for every k, (2) disjointness $D_j \cap D_k = \emptyset$ for all $j \neq k$ up to sets of measure zero, and (3) the covering property $\cup_{k=1}^m \overline{D_k} = \mathbf{X}$.

Given a Markov process X_t with $X_0 \sim \mu$, we define the transition probability $p(t, D_j, D_k)$ from $D_j \subset \mathbf{X}$ to $D_k \subset \mathbf{X}$ within time t as the conditional probability

$$p(t, D_j, D_k) = \mathbf{P}_\mu[X_t \in D_k \,|X_0 \in D_j] = \frac{\mathbf{P}_\mu[X_t \in D_k \;\&\; X_0 \in D_j]}{\mathbf{P}_\mu[X_0 \in D_j]} \qquad (2)$$

which, assuming absolute continuity of $p(t, \cdot, y)$ with respect to μ, equals

$$p(t, D_j, D_k) = \frac{1}{\mu(D_j)} \int_{D_j} p(t, x, D_k)\, \mu(\mathrm{d}x)\,. \qquad (3)$$

In other words, the transition probability measures the dynamical fluctuations within the stationary ensemble μ. Consequently, we may call a subset $D_k \subset \mathbf{X}$ *metastable* on the time scale $\tau > 0$, if

$$p(\tau, D_k, D_k^c) \approx 0 \quad \text{or, equivalently,} \quad p(\tau, D_k, D_k) \approx 1,$$

where $D_k^c = \mathbf{X} \setminus D_k$ denotes the complement of D_k in $\mathbf{X}$.

Transfer Operator. We define the *semigroup of Markov propagators* or forward transfer operators $P^t : L^r(\mu) \to L^r(\mu)$ with $t \in \mathbf{T}$ and $1 \leq r < \infty$ by

$$\int_A P^t v(y)\mu(\mathrm{d}y) \;=\; \int_{\mathbf{X}} v(x)p(t, x, A)\mu(\mathrm{d}x)$$

for any measurable $A \subset \mathbf{X}$. If μ is invariant under the dynamics X_t, then it is easy to see that the characteristic function $\mathbf{1}_{\mathbf{X}} \in L^1(\mu)$ of the entire state space is an invariant density of P^t, i.e., we have $P^t\mathbf{1}_{\mathbf{X}} = \mathbf{1}_{\mathbf{X}}$. As following from its definition, P^t conserves norm, $\|P^t v\|_1 = \|v\|_1$ and positivity, i.e., $P^t v \geq 0$ whenever $v \geq 0$. Hence, P^t is a Markov operator.

If we furthermore suppose that both μ and $p(t, x, \cdot)$ are absolutely continuous with respect to Lebesgue measure, the expression for the propagator P^t becomes

$$P^t u(y) \;=\; \int_{\mathbf{X}} k_t(y, x)u(x)\mu(\mathrm{d}x)\,, \tag{4}$$

where $\mu(\mathrm{d}x) =: \mu(x)\mathrm{d}x$, and we have introduced the *transition kernel*

$$k_t(y, x)\mu(x) = p(t, x, y) \tag{5}$$

that is defined for all x, y for which $\mu > 0$. Obviously, the transition kernel satisfies

$$\int_{\mathbf{X}} k_t(y, x)\mu(y)\mathrm{d}y = 1\,, \quad \forall (x, t) \in \mathbf{X} \times \mathbf{T}\,. \tag{6}$$

For a reversible process the transition kernel is symmetric, i.e.,

$$k_t(x, y) = k_t(y, x).$$

Key idea of the transfer operator approach. The identification of a metastable decomposition is based on the following scheme.

> Given $\tau > 0$, the *number* of metastable states is given by the number of eigenvalues of the propagator P^τ close to its maximum eigenvalue one including itself and counting multiplicity. The metastable *sets* can then be computed from the corresponding eigenfunctions.

This strategy that is outlined in more detail below was first proposed by Dellnitz and Junge [18] for discrete dynamical systems with weak random perturbations. It has been successfully applied to molecular dynamics in various contexts, e.g., [17, 19, 20]. The key idea requires that two conditions on the spectrum of the transfer operator P^τ hold true, namely,

(C1) The essential spectral radius of P^τ is strictly less than one.

(C2) The eigenvalue $\lambda = 1$ is simple and dominant, i.e., $\eta \in \sigma(P^\tau)$ with $|\eta| = 1$ implies that $\eta = 1$.

In this work, we confine our attention to two types of Markov process: (1) overdamped Langevin processes (i.e., diffusion processes), and (2) constant-temperature molecular dynamics (e.g., Nosé-Hoover heat baths). For either case the dynamics is reversible and the transfer operator is self-adjoint. For type (1) systems, conditions (C1)–(C2) are known to be met under relatively weak growths conditions on the potential (see [17]). For systems of type (2), it is unknown whether conditions (C1)–(C2) hold. Nonetheless we include this class of systems here, for they are prevalently used and it is typically assumed that they fulfil the requirements for all practical purposes (i.e., for sufficiently high-dimensional molecules in solution).

We now come to define the *metastability of a decomposition* $\mathbf{D}$ as the sum of the metastabilities of its subsets: Suppose, we fix $\tau > 0$. Then, given an arbitrary decomposition $\mathbf{D}_m = \{A_1, \ldots, A_m\}$ of $\mathbf{X}$ into m distinct sets, we define the metastability of $\mathbf{D}_m$ as

$$\mathrm{meta}(\mathbf{D}_m) = \sum_{j=1}^{m} p(\tau, A_j, A_j)/m.$$

Hence, for each m, the optimal metastable decomposition $\mathbf{D}_m$ can then be defined as the decomposition that maximizes meta$(\cdot)$. The next result is due to [21] and provides the rationale behind the key idea of the transfer operator approach.

Theorem 1.1. *Let $P^\tau : L^2(\mu) \to L^2(\mu)$ be a reversible propagator that satisfies (C1) and (C2). Then P^τ is self-adjoint, and its spectrum is of the form*

$$\sigma(P^\tau) \subset [a, b] \cup \{\lambda_m\} \cup \ldots \cup \{\lambda_2\} \cup \{1\},$$

where $-1 < a \leq b < \lambda_m \leq \ldots \leq \lambda_1 = 1$. The metastability of an arbitrary decomposition $\mathbf{D}_m = \{A_1, \ldots, A_m\}$ of $\mathbf{X}$ is bounded from above by

$$p(\tau, A_1, A_1) + \ldots + p(\tau, A_m, A_m) \leq 1 + \lambda_2 + \ldots + \lambda_m,$$

where the isolated eigenvalues $\lambda_1, \ldots, \lambda_m$ are counted according to their multiplicity. Let further $v_1, \ldots, v_m$ be the corresponding normalized eigenfunctions, and let Q denote the orthogonal projection of $L^2(\mu)$ onto $\mathrm{span}\{\mathbf{1}_{A_1}, \ldots, \mathbf{1}_{A_m}\}$. *The lower metastability bound of the decomposition $\mathbf{D}$ then is*

$$1 + \kappa_2\lambda_2 + \ldots + \kappa_m\lambda_m + c \leq p(\tau, A_1, A_1) + \ldots + p(\tau, A_m, A_m),$$

where $\kappa_j = \|Qv_j\|^2_{L^2(\mu)}$ and $c = a\ (1 - \kappa_2 + \ldots + 1 - \kappa_n)$.

Theorem 1.1 establishes a relation between the state space decomposition into metastable subsets and the *Perron cluster* of dominant eigenvalues close to 1. In particular it states that the metastability of an arbitrary decomposition $\mathbf{D}_m$ cannot be larger than the sum of the first m eigenvalues of the transfer operator. The lower metastability bound is close to the upper bound, whenever the dominant eigenfunctions are almost constant on the metastable subsets $A_1, \ldots, A_m$; in this case, as can be seen easily seen, we have $\kappa_j \approx 1$ and $c \approx 0$. Moreover both lower and upper bound are sharp and asymptotically exact [21].

1.2 Metastability Analysis is Hierarchical

The fundamental finding of Theorem 1.1 is that *metastability analysis is a hierarchical endeavour*; for instance, if we have found the optimal metastable decomposition into, say, $\mathbf{D}_2$ using 2 metastable subsets, it might still be possible that one of the subsets can be decomposed further into metastable sets; this would give rise to a decomposition $\mathbf{D}_3$ into three sets that would yield almost the same metastability, $\mathrm{meta}(\mathbf{D}_3) \approx \mathrm{meta}(\mathbf{D}_2)$.

If the spectrum of the transfer operator has a pronounced gap after the m-th dominant eigenvalue, then the results in, e.g., [22, 13] show that any decomposition into more than m sets will have a significantly reduced metastability in terms of the function $\mathrm{meta}(\cdot)$. In the context of molecular dynamics applications, however, one should always be aware that particular aspects of interest may make it desirable to explore the hierarchy of metastable decompositions up to a certain level that is not necessarily optimal in the sense of maximizing metastability.

2 Discretization

Let $\chi = \{\chi_1, \ldots, \chi_n\} \subset L^2(\mu)$ denote a set of non-negative functions that are a partition of unity, i.e., $\sum_{k=1}^n \chi_k = \mathbf{1}_{\mathbf{X}}$. We define the *Galerkin projection* $\mathcal{G}_n : L^2(\mu) \to \mathcal{S}_n$ onto the finite-dimensional space $\mathcal{S}_n = \mathrm{span}\{\chi_1, \ldots, \chi_n\}$ as

$$\mathcal{G}_n v \;=\; \sum_{k=1}^{n} \frac{\langle v, \chi_k \rangle_\mu}{\langle \chi_k, \chi_k \rangle_\mu}\, \chi_k.$$

If we apply the Galerkin projection to the infinite-dimensional eigenvalue problem $P^\tau v = \lambda v$ we obtain an eigenvalue problem for the discretized propagator $P_n^\tau = \mathcal{G}_n P^\tau \mathcal{G}_n$ acting on the finite-dimensional space $\mathcal{S}_n$. The matrix representation of the finite-rank operator P_n^τ is an $n \times n$ transition matrix $T = (T_{kl})$ with entries

$$T_{kl} \;=\; \frac{\langle P^\tau \chi_k, \chi_l \rangle_\mu}{\langle \chi_k, \chi_k \rangle_\mu}. \tag{7}$$

The finite-rank operator P_n^τ inherits basic properties of the transfer operator P^τ: Its matrix T is a stochastic matrix with invariant measure that is given by the projection invariant measure μ of P^τ to $\mathcal{S}_n$. Moreover, T is reversible, if P^τ is self-adjoint, and, assuming the discretization is fine enough, it also exhibits a Perron cluster of eigenvalues that approximates the corresponding Perron cluster of P^τ with eigenvectors that approximate the dominant eigenvectors of the original transfer operator [17]. Hence the transition matrix T allows for computing metastable sets by computing the dominant eigenvectors and employing an aggregation technique that is known by the name of "Perron Cluster Cluster Analysis" (PCCA) and which is based on the identification strategy described on page 4; we refer to [23, 24] for details.

The entries of T can be computed from realizations of the underlying Markov process X_t. Letting $\mathbf{E}_x$ denote the expectation of X_t started at $X_0 = x$, and using that $p(\tau, x, D_l) = \mathbf{E}_x(\chi_l(X_\tau))$, where $D_l = \mathrm{supp}(\chi_l)$, we have

$$T_{kl} = \frac{1}{\langle \chi_k, \chi_k \rangle_\mu} \int_{\mathbf{X}} \chi_k(x) \mathbf{E}_x[\chi_l(X_\tau)]\, \mu(\mathrm{d}x).$$

If $x_0, \ldots, x_N$ denote a time series obtained from a sufficiently long realization of the Markov process with time step τ, then the entries of T can be approximated by the relative transition frequencies

$$T_{kl} \approx T_{kl}^{(N)} \;=\; \frac{\sum_{j=1}^{N} \chi_k(x_j) \cdot \chi_l(x_{j+1})}{\sum_{j=1}^{N} \chi_k(x_j)^2} \,. \tag{8}$$

The rightmost expression in the last equation is the Maximum-Likelihood estimator of the spatially discretized transfer operator T; see formula (9) below, where also the *sampling error* of the transition matrix is discussed. The calculation of the relative frequencies may become problematic, even though the time series is very long. Let alone the dimension of state space, we often face what is called the *trapping problem*. The rate of convergence of $T_{kl}^{(N)} \to T_{kl}$ as $N \to \infty$ depends on the smoothness of the partitioning functions χ_k as well as on the mixing properties of the Markov process [25]. Especially mixing is crucial as convergence is geometric with a rate constant $\lambda_1 - \lambda_2 = 1 - \lambda_2$, where λ_2 denotes the second largest eigenvalue. If the system is metastable, we have $\lambda_2 \approx 1$, hence convergence is extremely poor. For realistic biomolecular system this will typically be the case, and, in fact, there is a huge amount of articles in the literature that deals with the question of how to overcome the *trapping problem* in molecular simulations. We will not take up the discussion about the sampling problem that is beyond the scope of this article and refer the interested reader to the relevant literature, e.g., [26, 27].

Henceforth we shall suppose that we have already a "sufficiently long" time series in the sense that it contains enough statistical information about some — but not necessarily all — revelant metastable states of the system. We denote this time series by $\{X_t\}_{t=t_0,\ldots,t_N}$ with $X_{t_i} \in \mathbf{X} \subseteq \mathbf{R}^n$ being either the atomic positions and/or momenta or some lower-dimensional observable, e.g., certain dihedral angles or base-pair parameters. Let us further assume that $\{X_{t_0}, X_{t_1}, \ldots\}$ comes with a uniform time step $\tau = t_{k-1} - t_k$. Setting $t_0 = 0$ we thus have $t_k = k\tau$ and $T = t_N = N\tau$ which, to simplify notation, will be often written as $t = 0, ..., N$.

A remark is in order: it can be proved [17] that the finite-dimensional Galerkin basis $\chi = \{\chi_1, \ldots, \chi_n\}$ yields a transition matrix (T_{ij}) that converges to the continuous operator as $n \to \infty$ and $\mathrm{diam}(\mathrm{supp}(\chi_k)) \to 0$. That is, if the partition defined by χ is sufficiently fine we can approximate the continuous transfer operator by simply counting transitions between the sets on which the χ_k are supported, provided the time series is sufficiently long. However avoiding a combinatorial explosion of discretization "boxes" as the dimension of state space increases, we will often employ a very coarse partition, e.g., by considering only parts of the variables that are assumed to be significant for the conformation dynamics (in this case the χ_k are supported on non-compact cylindrical sets). We shall suppose that the dynamics between the coarse sets is Markovian, i.e., the transition probabilities to go from one set to another depends only the current state but not on the process' history; this will be, for example, the case if the unresolved part of the dynamics is sufficiently fast mixing. In general, the appropriate choice of the discretization boxes so as to ensure the Markov property is non-trivial, and we refer to, e.g., the articles [28, 29, 30] for a discussion of the subject.

Furthermore, as we will see in Sec. 2.3.5, that it may be of interest to use

Galerkin subspaces that are *not* spanned by characteristic functions belonging to some kind of discretization boxes but subspaces spanned by smooth functions with overlapping support (so-called fuzzy Galerkin discretizations of the transfer operator, cf. [58]).

2.1 Error Estimation

Let $\chi = \{\chi_1, \ldots, \chi_n\}$ be characteristic functions, i.e.: $\chi_k(x) \in \{0, 1\}\, \forall x$, such that the transitions between discrete states are unambiguously identifiable and countable. We furthermore assume that jump process between discrete states is Markovian.

Let then the frequency matrix $C = (c_{ij})$ count the number of observed transitions between states, i.e., c_{ij} is the number of observed transitions from state i at time t to state j at time $t + \tau$, summed over all times t. In the limit of an infinitely long trajectory, the elements of the true transition matrix are given by the trivial estimator:

$$\hat{T}_{ij}(\tau) = \frac{c_{ij}}{\sum_k c_{ik}} = \frac{c_{ij}}{c_i}, \tag{9}$$

where $c_i := \sum_{k=1}^{m} c_{ik}$ is the total number of observed transitions leaving state i. For a trajectory of limited length, the underlying transition matrix $T(\tau)$ cannot be uniquely determined. The probability that a particular $T(\tau)$ would generate the observed trajectory is given by:

$$\mathbf{P}[C|T] = \prod_{i,j=1}^{m} T_{ij}^{c_{ij}}$$

Conversely, the probability that the observed data was generated by a particular transition matrix $T(\tau)$ is

$$\mathbf{P}[T|C] \propto \mathbf{P}[T]\, \mathbf{P}[C|T] = \mathbf{P}[T] \prod_{i,j \in S} T_{ij}^{c_{ij}}, \tag{10}$$

where $\mathbf{P}[T]$ is the prior probability of transition matrices before observing any data. It turns out that $\hat{T}(\tau)$, as provided by (9), maximizes $\mathbf{P}[C|T]$ and therefore also $\mathbf{P}[T|C]$ on condition that the transition matrices are uniformly distributed *a priori*. In the limit of infinite sampling, $\mathbf{P}[T|C]$ converges towards a delta distribution with its peak at $\hat{T}(\tau)$. When sampling is finite, the uncertainties of the entries of $\hat{T}(\tau)$ may be estimated by the element-wise standard deviations of $\mathbf{P}[T|C]$.

In general, one is interested in computing a particular property, $f(T(\tau))$, from the transition matrix. The symbol f may represent any smooth function, decomposition or algorithm, such as the eigenvalues- or eigenvectors. One is then interested how the uncertainty of the transition matrix, induced by the distribution $\mathbf{P}(T|C)$, carries over to uncertainties in the target function. In other words, *for a given observation C, what is the distribution of target functions, $\mathbf{P}[f(T)|C]$ and its standard deviation*?

Approaches to estimate the standard deviation based on first-order perturbation theory and Dirichlet sampling have been proposed in [8]. While being computationally efficient, these approaches do not allow for conserving of a number of physically meaningful constraints. In particular, only stochastic matrices

should be considered ($T_{ij} \geq 0$ and $\sum_j T_{ij} = 1 \ \forall i, j$), and for molecular transitions that are in equilibrium (not driven by an external force), detailed balance is expected to hold ($\pi_i T_{ij} = \pi_j T_{ij} \ \forall i, j$), where π_i is the stationary probability of state i. A general method to sample transition matrices according to these constraints can be based on Markov Chain Monte Carlo (MCMC): Here, one generates a series of matrices, $T^{(k)}$, $k = 1...N$, distributed according to $\mathbf{P}[T|C]$, which can be used to compute a distribution of $f(T)$. MCMC iterates proposal and acceptance steps. Given a current matrix $T^{(k)}$, a new matrix T' is proposed based in some stochastic manner. There are many possible choices of proposal steps. For correctness, it is only required that the probabilities for the forward and backward proposal, $\mathbf{P}[T^{(k)} \to T']$ and $\mathbf{P}[T' \to T^{(k)}]$, can be evaluated, and that any matrix of the distribution can be generated from any other matrix within a finite series of proposal steps. Then, the proposed matrix is accepted with probability:

$$p_{accept} = \frac{\mathbf{P}[T' \to T^{(k)}]}{\mathbf{P}[T^{(k)} \to T']} \frac{\mathbf{P}[T'|C]}{\mathbf{P}[T^{(k)}|C]}. \tag{11}$$

Upon acceptance, the proposed matrix becomes a member of the sample, $T^{(k+1)} := T'$, while upon rejection, the previous matrix is accounted for again: $T^{(k+1)} := T^{(k)}$. Efficient approaches to generate transition matrices according to such a scheme are described in [57].

2.2 Illustrative Example and PCCA

In the following we present results for the analyis of the dynamical behaviour of trialanine, a small peptide composed of three alanine amino acid residues.

For the molecular dynamics simulation of trialanine we have used the Gromos96 vacuum force field [31] in which trialanine is represented by 21extended atoms. The structural and dynamical properties of this molecule are mainly determined by two central peptide backbone angles Φ and Ψ. In addition, at very high temperatures, the otherwise planar peptide bond angle Ω may also undergo some conformational transition (see Figure 1).

The time series of 50000 steps has been generated by means of Hybrid Monte Carlo (HMC) at a temperature of 700 K [32]. The deterministic proposals for HMC are generated by running a 500fs trajectory employing the Verlet integration scheme with a time step of 1fs, yielding an acceptance rate of about 93 percent.

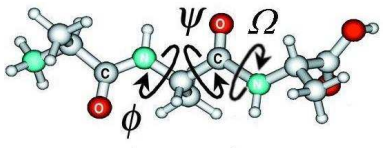

Figure 1: The trialanine molecule shown in ball-and-stick representation. At room temperature the overall structure of trialanine is sufficiently described by the two torsion angles Φ and Ψ, whereas at higher temperature also the dynamics of the peptide bond angle Ω becomes nontrivial.

From the thus obtained time series we compute the observation sequences in Φ, Ψ, and Ω. The space spanned by these three torsion angles is the compact 3-torus $\mathbf{T}^3 = S^1 \times S^1 \times S^1$ and will be called the *torsion space* in the following. The empirical distribution on torsion space is shown in Figure 2, where we have shifted the periodic intervals so as to avoid cut-offs.

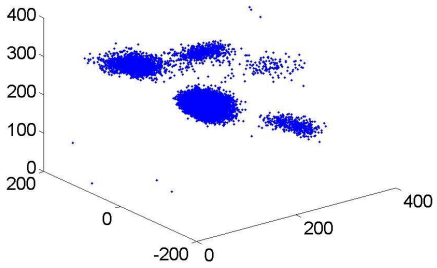

Figure 2: Observation time series from MD simulation of trialanine. Empirical distribution on torsion space.

We find that the molecular dynamics torsion space does not explore the 3-torus uniformly; rather we see five clearly pronounced clusters in left panel of Figure 3. The different colours for the clusters have been assigned by clustering the entire data set using the K-means algorithm, whereby each data point is assigned to exactly one cluster. As the K-means algorithm clusters data according to geometric distance in torsion space, we call the resulting clusters *geometric clusters*. Other algorithms for geometric clustering result in almost the same cluster assignment in this case.

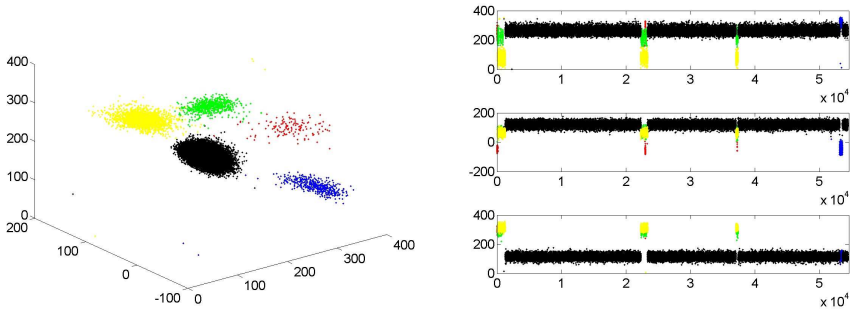

Figure 3: Left: Clustering of sampling distribution in torsion space into 5 geometric clusters (see text). Right: Colouring of the original time series according to geometric cluster assignment. Remark: In the subsequent the following numbering of these geometric clusters is used: number 1 corresponds to the red cluster, number 2 to the blue one, 3 to green, 4 to black, and 5 to yellow.

Uniform box discretization. The available time series is now used to discretize the transfer operator according to the procedure described on page 4. We choose a uniform box discretization of torsion space into 30^3 boxes, i.e., each

dimension is uniformly discretized into 30 boxes. This results in a transition matrix with many null columns and rows, for not all boxes are visited by the dynamics. By ignoring such boxes 1451 boxes remain, and we end up with a reversible 1451×1451 transition matrix T. Figure 4 illustrates the sparsity pattern of T; the numerically computed first six eigenvalues are shown in Table 1. Assuming that the data set contains indeed five metastable sets and clustering according to the five first eigenfunctions, it turns out that the metastable sets are almost identical with the geometric clusters obtained by K-means (see also the discussion below Figure 6).

k	1	2	3	4	5
λ_k	1.0000	0.9993	0.9992	0.9937	0.9773

Table 1: Five dominant eigenvalues of the transition matrix as resulting from direct discretization of the transfer operator by uniform discretization of torsion space into 30^3 boxes. All following eigenvalues are considerably smaller.

Of course, we might as well try to identify, say, the first $M = 3$ dominant metastable sets. To this end we carry out the robust version of the Perron-Cluster Cluster Analysis (PCCA) described on page 4. A detailed description of this particular variant, PCCA+, can be found in [24]. PCCA+ proceeds by plotting the entries of the second eigenvector v_2 to the dominant eigenvalue λ_2 against the the entries of the third eigenvector, v_3, corresponding to λ_3: the first eigenvector, v_1, is constant and can be ignored. The entries of v_2 and v_3 represent the values of the eigenvectors in the respective discretization box in torsion space, i.e., each box is mapped to a point in v_2-v_3 plane. It has been shown in [24] that the resulting points lie in the edges of a triangle as is illustrated in the right panel of Figure 4 below. PCCA+ eventually identifies the three dominant metastable sets as the aggregation of boxes belonging to the same cluster in the triangle.

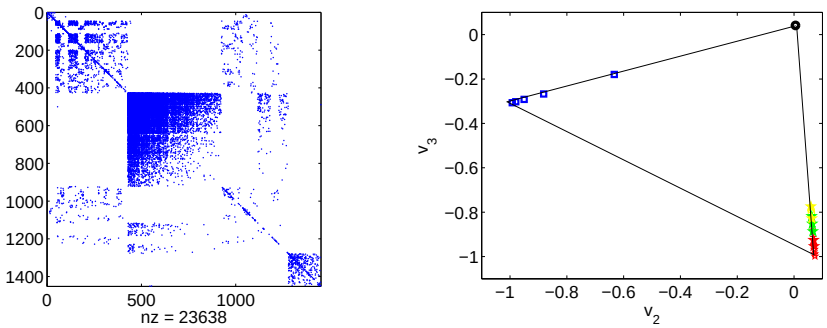

Figure 4: Results of uniform discretization of the transfer operator. Left: Sparsity pattern of the resulting dicretization matrix after uniform discretization of torsion space into 30^3 boxes. Right: Plot of second versus third eigenvector; colouring according to the scheme introduced in the caption of Figure 2. Obviously, all boxes from the black cluster get mapped to almost the same point, while the outmost blue squares also represents thousands of images of boxes from the blue cluster.

If each point in the right panel of Figure 4 is coloured according to the

scheme that is described in the caption of Figure 2, it turns out that the first two metastable sets found by PCCA+ are identical with the black and blue geometrical clusters, respectively, while the the third metastable set is the aggregation of the three remaining geometric clusters (see Figure 5).

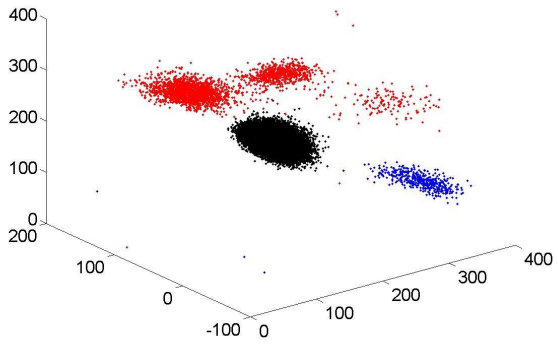

Figure 5: Aggregation of distribution in torsion space into 3 metastable sets according to PCCA+ based on uniform box discretization of the transfer operator.

Finally, we can aggregate the transition matrix into a 3×3 matrix coupling matrix between the three dominant metastable sets as resulting from PCCA+ after discretization of the torsion space transfer operator into 30^3 boxes:

$$T_{PCCA,discr} = \begin{pmatrix} 0.9999 & 0.0001 & 0.0000 \\ 0.0024 & 0.9975 & 0.0001 \\ 0.0051 & 0.0022 & 0.9927 \end{pmatrix}.$$

Next, we let PCCA+ find $M = 4$ dominant metastable sets, in the course of which the hierarchical aspect of metastability analysis should become clear. The analysis is now based on the eigenvectors v_2, v_3 and v_4, and we have to consider the projections onto the v_1-v_2-v_3 space as is shown in Figure 6. By comparison with the right panel of Figure 4 this demonstrates that the first three of the four dominant metastable sets are given by the black, blue, and red geometric cluster, while the fourth is the aggregation of the yellow and the green geometric cluster. That is, the four dominant metastable sets results from decomposition of one of the three dominant metastable sets into two different subsets.

Discretization based on geometric clustering. Several articles advocate taking geometric clusters as discretization boxes for the transfer operator, e.g., [29]. Following this route yields the discretization matrix

$$T = \begin{pmatrix} 0.9328 & 0 & 0.0522 & 0 & 0.0149 \\ 0 & 0.9952 & 0 & 0.0048 & 0 \\ 0.0091 & 0 & 0.9486 & 0 & 0.0423 \\ 0 & 0.0000 & 0 & 0.9999 & 0.0001 \\ 0.0011 & 0 & 0.0143 & 0.0021 & 0.9825 \end{pmatrix},$$

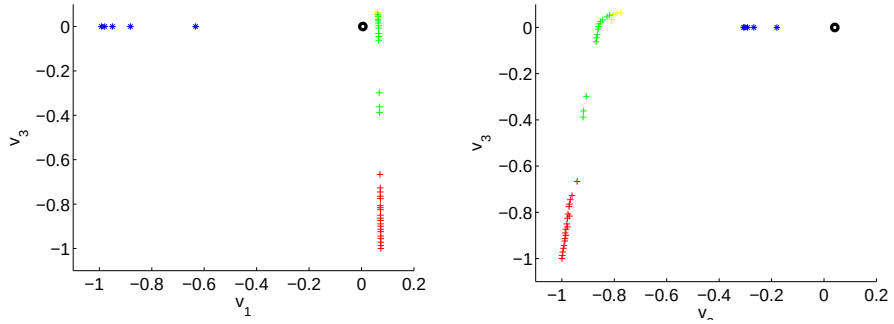

Figure 6: Results of uniform discretization of the transfer operator. Two different projections of the v_1, v_2, v_3-map. Left: onto the v_1-v_3 plane. Right: onto the v_2-v_3-plane.

k	1	2	3	4	5
λ_k	1.0000	0.9986	0.9952	0.9415	0.9173

Table 2: Complete set of eigenvalues as resulting from discretization of the transfer operator by discretization of torsion space into 5 boxes resulting from K-means clustering.

where the boxes are numbered according to numbering given in the caption of Figure 3. Its eigenvalues are given in Table 2 below.

We observe that all 5 clusters are metastable sets (as expected). However, we also observe significant deviation between the fourth and fifth of these eigenvalues and the respective eigenvalues resulting from uniform box discretization (see Table 1). The deviations may be explained by the coarseness of the geometric clusters as discretization boxes, but it also raises the question of the statistical reliability of the corresponding eigenvalues. Indeed, Figure 7 shows the distribution of the eigenvalues that is obtained from the distribution of the clusters' transition matrices which indicates a large variance of fourth and fifth eigenvalue (cf. Section 2.1).

In order to identify the three dominant metastable sets based on the K-means discretization we again apply PCCA+ to the eigenvectors associated with the three dominant eigenvalues. The result is visibly indistinguishable from the one displayed in Figure 5. The similarity of the result to the result of the uniform discretization demonstrates that in this particular case the geometric clusters give the appropriate discretization boxes.

A warning. We have to keep in mind that the geometric clusters are based on Euclidean distances, and hence do not incorporate kinetic information. Accordingly geometric clustering cannot distinguish between geometrically similar configurations that are kinetically separated. This problem is expected to become more severe as the dimensionality of the system increases, for situations may easily occur in which a strong change in one important degree of freedom leads into a new metastable set while a slight changes in many variables does not lead into a new metastable set; nonetheless, these two changes may be similar in terms of Euclidean distances. This results in general inappropriateness

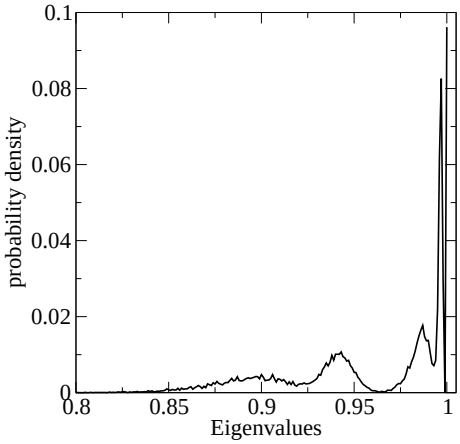

Figure 7: Distribution of eigenvalues as computed from the distribution of reversible transition matrices.

of geometric clusters for discretization of the transfer operator. In contrast to "blind" geometric clustering, the adaptive refinement of geometric clusters in such a way as to maximize the metastability between sets may be a useful approach for high-dimensional systems [29]. In typical cases the resulting cluster discretization can contain quite large numbers of geometric clusters that resolve rather fine details of the underlying distribution in state space (see Section 5).

2.3 Kernel Approximation

The discretization of the transfer operator becomes a tedious issue, or even impossible, if the dimension of state space is high. On the other hand, geometric clustering methods may provide a loophole from the curse of dimensions, but are in danger of ignoring the kinetic separation of geometrically close conformations. Consequently, we shall ask: is there a way to apply geometric clustering without ignoring the kinetic separation of states? Fortunately, we can answer this question in the affirmative, and we devote the next section to this problem. We start with some preliminary considerations.

2.3.1 Ornstein-Uhlenbeck kernels

Consider an Ornstein-Uhlenbeck (OU) process

$$dX_t = -F(X_t - \bar{x}) + \Sigma dW_t\,, \quad X_0 = x_0 \tag{12}$$

with $W(t)$ denoting Brownian motion in $\mathbf{X} \subseteq \mathbf{R}^n$ $\Sigma \in \mathbf{R}^{n\times n}$, and $F \in \mathbf{R}^{n\times n}$ being symmetric and positive definite. Its solution X_t is a time-homogeneous Markov process with transition function that is absolutely continuous with respect to Lebesgue measure on $\mathbf{R}^n$. If we set $B = \Sigma\Sigma^T$, the flat-space transition density at time t assumes the form

$$p(t, x_0, x) = Z(t) \exp\left(-\frac{1}{2}(x - \xi(t))^T C(t)^{-1}(x - \xi(t))\right)\,,$$

where we have used the shorthands $\xi = \bar{x} + \exp(-tF)(x_0 - \bar{x})$ and $Z = (2\pi)^{-n/2}(\det C)^{-1/2}$. The symmetric, positive definite matrix C can be shown to be the unique solution of the Lyapunov equation

$$C(t)F^T + FC(t) = B - \exp(-tF)B\exp(-tF^T)\,.$$

The corresponding invariant measure is absolutely continuous with respect to Lebesgue measure. Its density reads

$$\mu(x) = Z_\infty \exp\left(-\frac{1}{2}(x-\bar{x})^T C_\infty^{-1}(x-\bar{x})\right),$$

with $Z_\infty = (2\pi)^{-n/2}(\det C_\infty)^{-1/2}$ and C_∞ being the unique positive-definite solution of the Lyapunov equation $C_\infty F^T + FC_\infty = B$. The associated (unweighthed, i.e., flat-space) Markov propagator $P^t : L^p \to L^p$ then is

$$P^t f(x) = \int p(t, x_0, x) f(x_0) \mathrm{d}x_0\,.$$

The object of interest in this section is the *sampling kernel*

$$\kappa_t(x_0, x) = p(t, x_0, x)\mu(x_0)\,, \tag{13}$$

that can be directly computed from numerical trajectories of the OU process as we will show below. The μ-weighted version of the *transition kernel*, equation (5), has been introduced in Section 1. In terms of the sampling kernel the weighted transition kernel reads

$$k_t(x, x_0) = \frac{1}{\mu(x)} p(t, x_0, x) = \frac{1}{\mu(x)} \kappa_t(x_0, x) \frac{1}{\mu(x_0)}.$$

Using the expression for μ above, the sampling kernel can be expressed as

$$\kappa_t(x_0, x) = Z(t) Z_\infty \exp\left(-\frac{1}{2}((x-\bar{x})^T, (x_0-\bar{x})^T)\mathcal{C}(t)^{-1} \begin{pmatrix} x-\bar{x} \\ x_0-\bar{x} \end{pmatrix}\right) \tag{14}$$

with

$$\mathcal{C}^{-1} = \begin{pmatrix} C^{-1} & C^{-1}\exp(-tF) \\ \exp(-tF^T)C^{-1} & \exp(-tF^T)C^{-1}\exp(-tF) + C_\infty^{-1} \end{pmatrix}.$$

The sampling kernel is in one-to-one correspondence with the parameters of the OU process, and can therefore be used to estimate the unknown parameters of a stochastic process. If the transition kernel is given in terms of its covariance matrix $\mathcal{C}$, the respective parameter matrices are given by

$$\begin{aligned} \exp(-t\hat{F}) =& M_{11}^{-1} M_{12} \\ \hat{C}_\infty^{-1} =& M_{22} - M_{12}^T M_{22}^{-1} M_{12} \\ \hat{B} =& \hat{C}_\infty \hat{F}^T + \hat{F}\hat{C}_\infty\,, \end{aligned} \tag{15}$$

where we used the notation

$$\mathcal{C}^{-1} = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{pmatrix}.$$

2.3.2 Invariant Measure of Transition Kernels

Whenever the sampling kernel κ_t is known, the associated invariant measure is obtained upon integration, viz.,

$$\mu(x) = \int p(t, x_0, x)\mu(x_0)\mathrm{d}x_0 = \int \kappa_t(x_0, x)\mathrm{d}x_0 \,.$$

Here $p(t, \cdot, \cdot)$ denotes the associated flat-space transition function. For Gaussian transition kernels with stationary covariance matrix $\mathcal{C}$ we get

$$\mu(x) \propto \exp\left(-\frac{1}{2}x^T \hat{C}_\infty^{-1} x\right) \,, \quad \hat{C}_\infty = M_{22} - M_{12}^T M_{11}^{-1} M_{12}.$$

2.3.3 Gaussian approximation of sampling kernels

Let $\kappa_t(x, y)$ denote the sampling kernel of some Markov process. Our algorithmic strategy will be to approximate κ_t by a superposition of Gaussian sampling kernels: Let some not loo large integer M be given. We are interested in finding the optimal approximation of κ_t by a superposition of M Gaussian sampling kernels, i.e., we intend to solve the optimization problem

$$\left\| \kappa_t(x, y) - \sum_{k=1}^{M} \alpha_i \kappa_{i,t}(x, y) \right\| \to \min_{\alpha_i, \kappa_{i,t}} \,, \tag{16}$$

where the α_i are positive weights, $\{\kappa_{i,t}\}_i$ is a collection of Gaussian sampling kernels and $\|\cdot\|$ is some appropriate norm. There are several algorithms for solving this optimization problem even in higher dimensions, e.g., [33, 34], where the appropriate algorithm clearly depends on the specific choice of $\|\cdot\|$.

Apart from the question of how to solve the optimization problem it is important to notice that the kernel approximation requires to solve only a purely static problem that, nevertheless, incorporates the complete dynamics via the sampling kernel κ_t, thereby respecting the kinetics of the problem.

2.3.4 Additive kernels and metastability

In order to understand the properties of processes which sampling kernels are superpositions of Gaussian kernels we first have to study additive kernels.

Let $k_{i,t}$, $i = 1, \ldots, M$, be collection of Markovian transition kernels with absolutely continuous invariant probability measures μ_i on the joint state space $\mathbf{X}$. We consider the mixed kernel

$$\mu(x)k_t(x, y)\mu(y) = \sum_{i=1}^{M} \alpha_i \mu_i(x) k_{i,t}(x, y)\mu_i(y) \,,$$

that is a convex combination of the $k_{i,t}$, i.e., the coefficients α_i sum up to one. By linearity it then follows that $\mu = \sum_i \alpha_i \mu_i$ is the invariant measure of the mixed kernel that is absolutely continuous with respect to Lebesgue measure. We need the following definition.

Definition 2.1. *Suppose that $\mu > 0$ almost everywhere. The (mixed) transition kernel k_t is called ϵ-metastable, if and only if*

$$O_{ij} = \int_{\mathbf{X}} \frac{\mu_i(x)\mu_j(x)}{\mu(x)} \, \mathrm{d}x \leq \epsilon.$$

for all $i, j = 1, \ldots, N$ with $i \neq j$.

Almost invariant densities of k. It is convenient to weight the μ_i against the invariant measure μ of the mixed process. The weighted densities

$$\Phi_i(x) = \frac{\mu_i(x)}{\mu(x)}$$

are obviously in $L^1(\mu)$. The condition for ϵ-metastability translates into

$$O_{ij} = \langle \Phi_i, \Phi_j \rangle_\mu \leq \epsilon \,, \tag{17}$$

where $\langle \cdot, \cdot \rangle_\mu$ denotes the inner product in $L^2(\mu)$,

$$\langle u, v \rangle_\mu = \int_{\mathbf{X}} u(x)v(x)\mu(x) \, \mathrm{d}x \,.$$

The following statement can be proved.

Theorem 2.1. *Let the mixed transition kernel $k_t = \sum_{i=1}^{M} k_{i,t}$ be ϵ-metastable with invariant measure $\mu > 0$ (almost everywhere). Assuming that the probability measures μ_i of the local kernels $k_{i,t}$ are absolutely continuous with respect to Lebesgue measure μ_i, we have*

$$\| k_t \circ \Phi_i - \Phi_i \|_{1,\mu} \leq 2(1 - \alpha_i) \, \epsilon.$$

for all $i = 1, \ldots, M$, where

$$\left(k_t \circ \Phi_i \right)(y) = \int_{\mathbf{X}} k_t(y, x)\Phi_i(x)\mu(x)\mathrm{d}x \,.$$

Key observation: Too much metastability. When one considers additive kernels that result from the optimal approximation of the sampling kernel of a given process then one typically observes the following (cf. [35]): The almost invariance of the Φ_i that is estimated via Theorem 2.1 is far more pronounced than the metastability in the original process. A moment of reflection tells us why: for unweighted norms the optimal Gaussian approximation of the sampling kernel in the sense of (16) does not lead to an accurate approximation in the improbable transition regions between the main centers of the metastable sets. In fact, the localized Gaussian kernels decay exponentially fast in the overlap regions, whereas the transition regions of the full transition kernel are significantly larger, though, still small enough such that they do not contribute to the total approximation error. Hence, the approximate kernel $\sum_i \alpha_i \kappa_i$ is much smaller in the transition regions than the original sampling kernel, thus the frequency of transitions is much smaller and the metastability is much more pronounced. In order to correct for this problem we have to add an appropriate amount of transitions to our description.

2.3.5 Assignment to metastable and transition states

In order to estimate the number of transitions correctly we count the transitions in an available molecular dynamics time series $X = \{X_k\}_{k=1,\ldots,N+1}$. From this we obtain the time series $Z = \{Z_k\}_{k=1,\ldots,T}$ with $Z_k = (X_k, X_{k+1})$ that is underlying the sampling kernel.

Let us suppose that we decompose the dynamics into M metastable sets. Then each point in the time series X can be assigned to either state by means of the almost invariant densities Φ_i, $i = 1, \ldots, M$, i.e., we define the core sets

$$\mathcal{M}_i = \{X_k : 1 \leq k \leq N+1,\, \Phi_i(X_k) > \theta \|\Phi_i\|_\infty\}, \quad i = 1, \ldots, M\,.$$

where $\theta > 0.5$ is some appropriate user-selected threshold (e.g., $\theta = 0.95$). These core sets may thought of the regions surrounding the deepest energy minima of the system, which were defined manually in Ref. [36] If the overlaps between the local densities Φ_i are small enough, we may assume that $\mathcal{M}_i \cap \mathcal{M}_j = \emptyset$ for $i \neq j$. All other data points X_k will be assigned to the *transition set*

$$\mathcal{M}_0 = \{X_k : 1 \leq k \leq N+1,\; \Phi_i(X_k) \leq \theta \|\Phi_i\|_\infty,\, i = 1, \ldots, M\}\,.$$

Transitions are events (k, X_k) with $X_k \in M_j$ for any $j = 0, \ldots, M$ and $X_{k+1} \notin M_j$. Accordingly, we can classify transitions in terms of the time series Z. To this end we define

$$M_{ij} = \{Z_k = (X_k, X_{k+1}) : 1 \leq k \leq N,\, X_k \in \mathcal{M}_i \text{ and } X_{k+1} \in \mathcal{M}_j\}\,.$$

If we let $\#A$ denote the number of elements in the set A, then

$$\#\mathcal{M}_i = \sum_j \#M_{ij}\,, \quad i = 0, \ldots, M\,,$$

counts the transitions to the coarse-grained sets $\mathcal{M}_0, \ldots, \mathcal{M}_M$ within a single time step. The corresponding optimal Maximum-Likelihood transition matrix under the observation $\{X_k\}_{k=0,\ldots,T}$ has the entries

$$p(i,j) = \frac{\#M_{ij}}{\#\mathcal{M}_i}. \tag{18}$$

If the transition set $\mathcal{M}_0$ is further subdivided to optimize metastability this way of deriving a transition matrix between metastable sets is quite similar to the approach suggested in [36], where the trajectory was cut into pieces, each of which connected two different core regions, counting a transition for each such piece. This similarity become very close, if we use the Φ_i as Galerkin ansatz functions for the discretization of the transfer operator.

2.3.6 Illustrative example revisited

We consider the trialanine example from the previous section, and define the sampling sequence $Z = \{Z_k\}_{k=1,\ldots,N}$ with $Z_k = (X_k, X_{k+1})$ from the original time series $X = \{X_k\}_{k=1,\ldots,N+1}$. Next, we solve the discrete analogue of the optimization problem (16): Find $\kappa = \sum_i \alpha_i \kappa_i$ with *stationary* Gaussian kernels κ_i that optimally approximate the empirical distribution generated by Z. One possibility to do this is to maximize the likelihood of observing of Z given κ. This

leads to the following Maximum-Likelihood problem: Let M be fixed, identify κ_i by its mean $\bar{z}_i$ and its covariance matrix C_i, aggregate $\bar{z} = (\bar{z}_1, \ldots, \bar{z}_M)$ and $C = (C_1, \ldots, C_M)$, then maximize

$$\mathbf{P}\left[Z|\alpha, \bar{z}, C\right] = \prod_{j=1}^{N} \sum_{i=1}^{M} \alpha_i \kappa_i(Z_j) \tag{19}$$

over all admissible parameters $\alpha, \bar{z}, C$, where *admissible* means that the C_i are symmetric and positive definite, and $\sum \alpha_i = 1$. Herein, we use the HMMGauss algorithm to compute the optimal parameters $(\alpha, \bar{z}, C)$ for the trialanine data; see the next section for details. Setting $M = 5$, the output of HMMGauss is employed to, firstly, compute the functions Φ_i and the sets $\mathcal{M}_i$ for $i, j = 1, \ldots, 5$ with the resulting M_{ij}, and then, secondly, the transition set $\mathcal{M}_0$ following to the procedure described in the last subsection. The result is illustrated in the top panel of Figure 8.

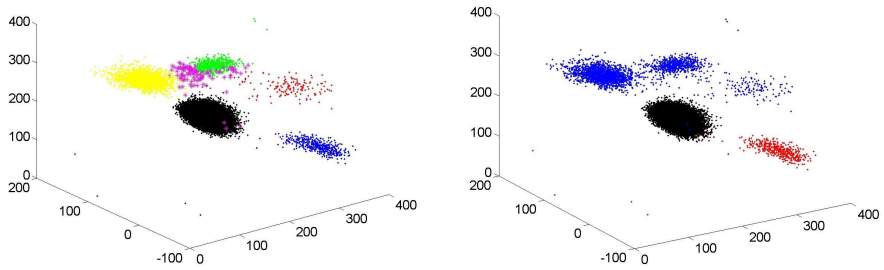

Figure 8: Aggregation of the empirical distribution on torsion space via kernel approximation based on HMMGauss (see text). Right: Aggregation into 6 sets: $\mathcal{M}_i$, $i = 1, \ldots, 5$, and the transition set $\mathcal{M}_0$. The five metastable sets $\mathcal{M}_i$, $i = 1, \ldots, 5$ almost agrees with the geometric clusters (colouring accordingly); points in $\mathcal{M}_0$ are shown as magenta crosses. Left: Further aggregation into 3 metastable sets as resulting from PCCA+ (see text).

We observe that the five metastable sets $\mathcal{M}_i$, $i = 1, \ldots, 5$ are almost identical with the five geometric clusters. Additionally, the transition set $\mathcal{M}_0$ is clearly visible between the green and the yellow set. Taking these six sets as discretization boxes for the discretization of the transfer operator and applying PCCA+ with the aim of finding three dominant metastable sets yields the PCCA+ triangle shown in Figure 9. The PCCA+ clustering again indicates that, on a coarser level, three of the metastable sets (red, green, yellow) and the transition set $\mathcal{M}_0$ form a single metastable set.

The resulting transition matrix of the three dominant metastable sets is

$$T = \begin{pmatrix} 0.9998 & 0.0001 & 0.0001 \\ 0.0030 & 0.9967 & 0.0004 \\ 0.0071 & 0.0024 & 0.9906 \end{pmatrix},$$

which agrees with the previously computed transition matrix within the achievable statistical accuracy.

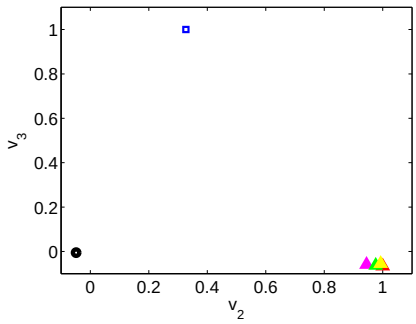

Figure 9: Plot of second versus third eigenvector; colouring according to the assignment of the respective box to the 6 clusters identified via kernel approximation.

2.4 Concluding Remarks

The previous steps of this section have demonstrated that we ought to use algorithms for the Gaussian approximation of the sampling kernel which can still be used for dimensional problems. The local Gaussian sampling kernels κ_i with the corresponding weights α_i give then rise to a coarse-graining of state space into M metastable sets $\mathcal{M}_0, \ldots, \mathcal{M}_M$ (core sets), including a transition set $\mathcal{M}_0$, that can be used to discretize the transfer operator in order to identify the dominant $\hat{M} < M$ metastable sets. We will revisit this procedure in Section 5.

Another perspective on kernel approximation. The procedure results in a specific reduced model for the original molecular dynamics problem. The very flexibility of the approach is due to possible choices of the local kernels as we shall illustrate briefly. To this end let $\kappa_{i,t}$ denote the local sampling kernel of the metastable set with index $i \in \{1, \ldots, \hat{M}\}$, and let $p_i(t, \cdot, \cdot)$ be the (flat-space) transition function associated with $\kappa_{i,t}$; the local transition function approximates the dynamical behaviour of the process while being in set $\mathcal{M}_i$. In addition, we have a jump process switching between the metastable sets according to the transition matrix T. The state space $\hat{S}$ of the reduced model is composed of $\hat{M}$ copies of the original state space S, i.e.,

$$\hat{S} = S \times \{1, \ldots, \hat{M}\}.$$

The overall (flat-space) transition function p that is generated by the local transition functions and the jump process thus has the form

$$p : [0, \infty) \times \hat{S} \times \hat{S} \to \mathbf{R}^+, \quad p(t, x, i, y, j) = p_j(t, x, y) T_{ij}. \tag{20}$$

If, for instance, the local transition Gaussian functions p_i are the transition functions of an OU process, we can find matrices $F^{(i)}$, $\Sigma^{(i)}$ and vectors $\bar{x}^{(i)}$, such that p_i is generated by

$$dX_t = F^{(i)} \left(X_t - \bar{x}^{(i)} \right) dt + \Sigma^{(i)} dW_t \,.$$

In this case, the reduced model that generates the overall transition function (20) has the form of switched stochastic differential equations,

$$\begin{aligned} dX_t &= F^{(q)}\left(X_t - \bar{x}^{(q)}\right) dt + \Sigma^{(q)} dW_t \\ q_t &= \text{Markov jump process with states } 1, \ldots, \hat{M}\,, \end{aligned} \tag{21}$$

where the jump process is governed by the transition matrix T.

3 Hidden Markov Models

In this section we further develop the considerations from the last paragraph and answer the question of how the metastable states can be identified and analysed, if the dimension of state space is too high as to admit a direct discretization of the transfer operator. As before we assume that we are given a sufficiently long time series $\{X_t\}_{t=t_0,\ldots,t_N}$ of micro-states $x \in \mathbf{X} \subseteq \mathbf{R}^n$ (i.e., atomic positions and/or momenta). The time series may also be given in terms of certain distinguished observables $f : \mathbf{R}^n \to \mathbf{R}^m$, $y = f(x)$ that are nonlinear functions of the microscopic states $x \in \mathbf{R}^n$ such as dihedral angles or base-pair parameters.

The approach using Hidden Markov Models (HMM) can be summarized as follows: By analysing a (possibly incomplete) time series we (1) construct a finite-state Markov jump process that models the hopping *between* metastable conformations; then, (2), we parametrize appropriate stochastic models that approximates the dynamics *within* each conformation. The HMM method is used to construct an unobserved (hidden) jump process thereby accounting for lack of information due to incomplete observations. Over the last few years, various algorithms in this direction have been developed combining HMM with Maximum-Likelihood based parametrization of the local stochastic models; see, e.g., [37, 38, 39]. We will review this framework now.

The idea of HMM. Roughly speaking, a HMM is a stochastic process with both hidden and observable states; the hidden states of a HMM are described by a Markov jump process, while the observable states are understood as their output that, e.g., follows a certain probability distribution conditional on the hidden state.

Suppose we consider a system admitting a metastable decomposition $\mathbf{D} = \{B_1, \ldots, B_m\}$. Then, at any time t, the system will be in one of the metastable sets $B_q \subset \mathbf{X}$ with $q = 1, \ldots, m$. Hence, for each t the integer value of the metastable state, q, represents a jump process q_t between the metastable states. The task then is, given a series of observations $\{X_t\}_{t=t_0,\ldots,t_N}$, to identify the underlying (hidden) time series of metastable states, $\{q_t\}_{t=t_0,\ldots,t_N}$.

We assume that the observed data $\{X_t\}_{t=t_0,\ldots,t_N}$ comes with a uniform time step $\tau = t_{k-1} - t_k$. Setting $t_0 = 0$ we thus have $t_k = k\tau$ and $T = t_N = N\tau$. In the generic case one assumes that the probability of observation X_k given the hidden state q_k and the previous observation X_{k-1} can be modelled by a certain family ρ_θ of observation distributions, i.e.,

$$\mathbf{P}\left[X_k | X_{k-1}, q_k\right] = \rho_{\theta(q_t)}(X_k | X_{k-1})\,,$$

where θ denotes the parameters of this family of distributions (e.g., mean and covariance matrix of a family of normal distributions), and the dependence of θ on the hidden state q_k indicates that each hidden state gives rise to a different observation distribution. Provided that the hopping dynamics is Markovian, the probability to go from one metastable (i.e., hidden) state $q = i$ to another one, $q = j$, within one time step τ is given by $T_{ij} = p(\tau, B_i, B_j)$. That means, we may regard the sequence $\{q_t\}$ as a, yet unknown, realization of an M-state Markov chain with transition matrix T. Conversely, the observations X_t are considered as *a priori* unknown random functions of the q_t, where the random functions are the local models in each of the conformations. Given a class of local models, e.g., certain linear probability distributions, the Maximum-Likelihood approach consist in finding the most likely sequence of hidden states $q \in \{1, \ldots, M\}$ given a series of observations $\{X_t\}$.

As yet, HMM assumes that the number of metastable conformations, M, is known. However, as we will argue below, it is sufficient to start the data analysis with a sufficiently large M and determine the optimal such $\hat{M} \leq M$ afterwards, for instance, by clustering states according to the eigenvalues of the transition matrix.

The likelihood. The hidden states $q \in \{1, \ldots, M\}$ typically correspond to different observation distributions ρ_{θ_j}, $j = 1, \ldots, M$, where the values θ_j and the M^2 entries of the transition matrix T are *a priori* unknown. We summarize all unknown parameters in one parameter vector

$$\Theta = (\theta_1, \ldots, \theta_M, T).$$

The likelihood of the parameters Θ is a probability density $\mathbf{P}[X, Q|\theta]$ of the possible observations X considered as a function of Θ and the hidden path $Q = \{q_k\}_{k=0,\ldots,N}$, i.e., $\mathcal{L} : (\Theta, Q) \mapsto \mathbf{P}[X, Q|\Theta]$. We have

$$\mathcal{L}(\Theta, Q) = \nu(X_0|q_0) \prod_{k=1}^{N} T_{q_{k-1},q_k}\, \rho_{\theta(q_k)}(X_k|X_{k-1})\,, \tag{22}$$

where ν denotes the system's initial distribution. In the HMM framework the sequence of hidden path Q appears as an unknown parameter that has to be determined. But as Q is hidden, finding it by maximizing the likelihood is not an option; instead, we have to estimate *simultaneously* the hidden path Q as well as the parameters Θ which is done employing the Expectation-Maximization (EM) algorithm [40].

The idea is as follows: integrating over all possible hidden paths, we obtain

$$\mathbf{P}\left[X|\Theta\right] = \sum_{Q} \mathcal{L}(\Theta, Q)\,, \tag{23}$$

by which we can introduce the distribution of the hidden path conditional on the observation, viz.,

$$\mathbf{P}\left[Q|X, \Theta\right] = \frac{\mathbf{P}\left[X, Q|\Theta\right]}{\mathbf{P}\left[X|\Theta\right]}\,. \tag{24}$$

An EM algorithm iteratively improves an initial estimate Θ_0 of the optimal parameters by constructing iterates $\Theta_1, \Theta_2, \ldots$ by the following procedure:

$$\Theta_{n+1} = \operatorname*{argmax}_{\Theta} \mathcal{Q}(\Theta, \Theta_n) \tag{25}$$

with $\mathcal{Q}$ being the expected value of the log-likelihood

$$\mathbf{E}\left[\log \mathcal{L}(\Theta, Q)|X, \Theta_n\right] = \sum_Q \mathbf{P}\left[Q|X, \Theta_n\right] \log \mathbf{P}\left[Q, X|\Theta\right] ,$$

i.e., the expectation of the log-likelihood over the hidden path given the parameters Θ_n of the last step. In each step of the EM algorithm the following two sub-steps have to be performed: computation of the expected log-likelihood (the E-step) and maximization of $\mathcal{Q}$ (the M-step).

Under rather general, non-pathological conditions the EM-iteration converges to a local optimum Θ_* by successively increasing the likelihood. EM is a partially non-Bayesian, Maximum-Likelihood method. Its final result gives a point estimate for Θ together with a probability distribution over the hidden path (see below). After convergence of the EM algorithm the optimal hidden path is eventually determined by

$$Q_* = \operatorname*{argmax}_Q \mathcal{L}(\Theta_*, Q). \tag{26}$$

The thus obtained path is called the *Viterbi path*, and its efficient computation is possible by means of the Viterbi algorithm; for more details see [37]. While Viterbi and EM algorithms are integral parts of each HMM procedure, different choices for the family of observation distributions generate different HMM models. We discuss two different cases.

HMMGauss: Stationary, normally distributed observation. The perhaps easiest observation model is the family of normal distributions

$$\rho_\theta(x) \propto \exp\left(-\frac{1}{2}(x - \bar{x}^{(q)})^T C_{(q)}^{-1}(x - \bar{x}^{(q)})\right)$$

with the unknown parameters

$$\theta = \{(\bar{x}^{(1)}, \ldots, \bar{x}^{(M)}, C_{(1)}, \ldots, C_{(1)}) : \bar{x}^{(i)} \in \mathbf{R}^n,\, C_{(i)} = C_{(i)}^T > 0\}.$$

In this case, the observation of current state X_k does *not* depend on the previous state X_{k-1}, but only on the hidden state (via the dependence of the parameters) as is indicated by the index (q).

HMMSDE: Dynamical SDE Output. As local output of the HMM, we consider Ornstein-Uhlenbeck (OU) processes

$$dX_t = F^{(q)}\left(X - \bar{x}^{(q)}\right) dt + \Sigma^{(q)} dW_t\,,$$

where, again, (q) indicates the dependence of the parameters on the hidden state; see also equation (21) in the last section that was the result of the kernel approximation procedure. From the formal solution

$$X_{t+\tau} = \bar{x} + e^{\tau F}\left(X_t - \bar{x}\right) + \int_0^\tau e^{(\tau - s)F} \Sigma dW_s\,.$$

of the OU process on the time interval $[t, t+\tau]$ we can compute the probability density $\rho_\theta(X_{k+1}|X_k)$ of an observation of X_{k+1} at time $t_{k+1} = t_k + \tau$ given an observation X_k at time t_k, namely,

$$\rho_\theta(X_k|X_{k-1}) = \frac{\exp\left(-\frac{1}{2}\left(X_{k+1}-\mu_k\right)^T R(\tau)^{-1}\left(X_{k+1}-\mu_k\right)\right)}{(2\pi)^{-d/2}\sqrt{\det R(\tau)}} \tag{27}$$

where

$$\mu_k = \bar{x} + e^{\tau F}\left(X_k - \bar{x}\right), \quad R(\tau) = \int_0^\tau e^{sF}\Sigma\Sigma^T e^{sF^T}\,\mathrm{d}s\,. \tag{28}$$

It is easy to see that the conditional observation distribution equals the flat transition function of the OU process, i.e., $\rho_\theta(X_k|X_{k-1}) = p(\tau, X_{k-1}, X_k)$.

Algorithm. Realizations of the EM algorithm for both stationary Gaussian and dynamical OU output are discussed in [37, 38, 39, 41]. In either case the necessary computational effort for one step of the EM algorithm scales linearly with the length of the observation sequence and quadratically with the number of hidden states.

Each E-step together with the initial condition of the hidden state and the transition matrix T yields occupation probabilities $\nu_k(q)$ at time t_k. That is, $\nu_k(q)$ denotes the probability to be in hidden state $q \in \{1, \ldots, M\}$ at time t_k at each step of the EM iteration based on the previous parameter values Θ_n. In the M-step the next parameter estimates $\Theta_{n+1} = \operatorname{argmax} \mathcal{Q}(\cdot, \Theta_n)$ can be computed based on these occupation probabilities. For HMMGauss and HMMSDE this optimization can be carried out analytically as we shall show next.

M-step in HMMGauss. Denote the ν-weigthed mean and covariance of the time series $\{X_1, \ldots, X_T\}$ in the state q by

$$\begin{aligned}
\bar{x}_N^{(q)} &= \left(\sum_{k=1}^{N-1}\nu_{k+1}(q)\right)^{-1}\sum_{k=1}^{N-1}\nu_{k+1}(q)X_k \\
\mathrm{cov}_N^{(q)}(X) &= \left(\sum_{k=1}^{N-1}\nu_{k+1}(q)\right)^{-1}\sum_{k=1}^{N-1}\nu_{k+1}(q)\left(X_k - \bar{x}_N^{(q)}\right)\left(X_k - \bar{x}_N^{(q)}\right)^T.
\end{aligned}$$

The optimal estimator for Θ_{n+1} hence involves $\theta_q = (\bar{x}_N^{(q)}, \mathrm{cov}_N^{(q)})$ provided that $\mathrm{cov}_N^{(q)}$ is positive definite.

M-step in HMMSDE. Let $\bar{x}_N^{(q)}$, $\mathrm{cov}_N^{(q)}$) be defined as above and introduce in addition the weighted one-step correlation

$$\begin{aligned}
\mathrm{cor}_N^{(q)}(X) = &\left(\mathrm{cov}_N^{(q)}(X)\sum_{k=1}^{N-1}\nu_{k+1}(q)\right)^{-1} \\
&\times \sum_{k=1}^{N-1}\nu_{k+1}(q)\left(X_{k+1} - \bar{x}_T^{(q)}\right)\left(X_k - \bar{x}_N^{(q)}\right)^T.
\end{aligned}$$

The optimal estimators $\hat{F}^{(q)}$, and $\hat{\mu}^{(q)}$ for the parameters $F^{(q)}$, and $\mu^{(q)}$ follow from the next statement that is due to [38]

Theorem 3.1. *Let $\mathrm{cov}_N^{(q)}$ be positive definite at each iteration step in the EM algorithm. Then, at each step, the optimal estimators satisfy*

$$\begin{aligned} \exp(\tau \hat{F}^{(q)}) &= \mathrm{cor}_N^{(q)} && (29) \\ \hat{\mu}^{(q)} &= \bar{x}_N^{(q)} + \left(\mathrm{Id} - \mathrm{cor}_N^{(q)}\right)^{-1} \Delta_N^{(q)}. && (30) \end{aligned}$$

where

$$\Delta_N^{(q)} = \left(\sum_{k=1}^{N-1} \nu_{k+1}(q)\right)^{-1} \sum_{k=1}^{N-1} \nu_{k+1}(q)\,(X_{k+1} - X_k)\,.$$

Equation (30) requires that $\|\mathrm{cor}_N^{(q)}\| < 1$, which will be always the case, if all eigenvalues of the estimated stiffness matrix $\hat{F}^{(q)}$ have strictly negative real part. In addition, we obtain a linear matrix equation for the estimator of the noise covariance $\hat{\Sigma}^{(q)}\hat{\Sigma}^{(q)^T}$,

$$e^{-\tau \hat{F}^{(q)}} W^{(q)} = \hat{\Sigma}^{(q)}\hat{\Sigma}^{(q)^T} e^{\tau \hat{F}^{(q)^T}} - e^{-\tau \hat{F}^{(q)}} \hat{\Sigma}^{(q)}\hat{\Sigma}^{(q)^T}, \qquad (31)$$

where

$$\begin{aligned} W^{(q)} &= \Omega^{(q)} \hat{F}^{(q)^T} + \hat{F}^{(q)} \Omega^{(q)}, \\ \Omega^{(q)} &= \left(\sum_{k=1}^{N-1} \nu_{k+1}(q)\right)^{-1} \sum_{k=1}^{T-1} \nu_{k+1}(q) \hat{d}_k^{(q)} \hat{d}_k^{(q)^T} \\ \hat{d}_k^{(q)} &= \left(X_{k+1} - \hat{\mu}^{(q)} - e^{\tau \hat{F}^{(q)}} \left(X_k - \hat{\mu}^{(q)}\right)\right). \end{aligned}$$

Again, the Lyapunov (31) has a unique and symmetric, positive definite solution, if and only if the eigenvalues of $\hat{F}^{(q)}$ lie in the open left half complex plane.

Theorem 3.1 allows for carrying out the maximization in the EM algorithm by basically computing weighted autocorrelation matrices which is numerically cheap. However, a remark is in order: the computation of $\hat{F}^{(q)}$ from $\exp(\tau \hat{F}^{(q)})$ is not trivial at all, for the matrix logarithm is not unique. We refer to [38] for a detailed discussion of possible difficulties and various algorithmic solutions.

Number of metastable states. All HMM techniques require to select the unknown number of hidden states in advance. There is no general solution to this problem, and often the best way to handle this problem is a mixture of insight and preliminary analysis. However, we should recall that we can easily cluster hidden (metastable) states following the route taken in the transfer operator approach to metastability. therefore we suggest to start the EM algorithm with any sufficiently large number of hidden states M that should be bigger than the expected number of metastable states. After termination of the EM algorithm, we can take the resulting transition matrix and aggregate the M hidden states into $\hat{M} \leq M$ metastable states using the PCCA method described in Section 2. By construction the thus clustered hidden states are metastable states of the dynamics.

3.1 HMMSDE and VAR Processes

Let us revisit the problem of the estimating optimal parameters for the single n-dimensional OU process (12). Theorem 3.1 makes an assertion about estimating optimal parameters $\theta = (\exp(\tau F), \bar{x}, \Sigma)$ from a given observation $X_1, \ldots, X_N$. The optimal parameters are found by maximizing the likelihood

$$\mathcal{L}(\theta|X) = \prod_{k=1}^{N-1} \rho_\theta(X_{k+1}|X_k),$$

with the one-step transition probability

$$\rho_\theta(X_{k+1}|X_k) \propto \exp\left(-\frac{1}{2}(X_{k+1}-\mu_k)^T R(\tau)^{-1}(X_{k+1}-\mu_k)\right), \tag{32}$$

with μ_k and $R(\tau)$ as given by (28). Unfortunately there is no known analytic solution to the maximization problem of $\mathcal{L}$ with respect to the parameter set $(\bar{x}, F, \Sigma)$; Theorem 3.1 yields the optimal parameters $\theta = (\exp(\tau F), \bar{x}, \Sigma)$, and the matrix logarithm is not surjective. Yet another drawback, from a statistical viewpoint, is that $\mathcal{L}$ is not integrable over the unrestricted parameter space which can be easily seen by setting $F = 0$ and integrating over $\bar{x}$. This imposes certain constraints on the admissible parameters, thereby complicating sampling of the Maximum-Likelihood estimators. A possible loophole consist in rewriting the transition probability (32) according to

$$X_{k+1} \sim \mathcal{N}(\bar{x} + \exp(\tau F)(X_k - \bar{x}), R)$$

which can be equivalently expressed as

$$X_{k+1} \sim (\mathrm{Id} - \exp(\tau F))\bar{x} + \exp(\tau F)X_k + \mathcal{N}(0, R), \tag{33}$$

Equation (33) resembles an autoregressive model of order one, VAR(1). If we define the shorthands

$$\begin{aligned}
\Phi &:= \big((\mathrm{Id} - \exp(\tau F))\bar{x},\, \exp(\tau F)\big) \in \mathbf{R}^{n\times(n+1)} \\
\xi &:= \begin{pmatrix} 1 & \ldots & 1 \\ X_1 & \ldots & X_{N-1} \end{pmatrix} \in \mathbf{R}^{(n+1)\times(N-1)} \\
Y &:= \big(X_2, \ldots, X_N\big) \in \mathbf{R}^{n\times(N-1)} \\
\epsilon &:= \big(\mathcal{N}_1(0,R), \ldots, \mathcal{N}_{N-1}(0,R)\big) \in \mathbf{R}^{n\times(N-1)} \quad \text{(i.i.d.)}\,,
\end{aligned}$$

we can recast (33) in the form

$$Y = \Phi\xi + \epsilon\,.$$

The likelihood of the new parameter set $\tilde{\theta} = (\Phi, R)$ reads

$$\tilde{\mathcal{L}}(\tilde{\theta}|X) = (\det R)^{\frac{N-1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}((Y-\Phi\xi)(Y-\Phi\xi)^T R^{-1})\right). \tag{34}$$

Maximum-likelihood estimators for $\hat{\Phi}$ and $\hat{R}$ can be found in the relevant literature, e.g., [43, 44]. We have

$$\hat{\Phi} = Y\xi^T(\xi\xi^T)^{-1} \text{ and } \hat{R} = \frac{1}{N-1}(Y-\hat{\Phi}\xi)(Y-\hat{\Phi}\xi)^T.$$

Using $\tilde{\theta}$ and $\tilde{\mathcal{L}}$ for the parameter estimation has the advantages that (1) the distribution of the discrete observations is fully characterized by $\tilde{\theta}$, (2) analytical Maximum-Likelihood estimators are available, and (3) the likelihood $\tilde{\mathcal{L}}$ is integrable over the unconstrained parameter space.

Moreover, the VAR(1) model has a straightforward extension to VAR(p) models that allow for adding non-Markovian memory effects to the description. Last but not least, we can do *change-point detection* for VAR processes so as to detect changes in the time series parametrization on-the-fly; such parametrization changes can occur if, for instance, the system makes transition between metastable sets. We refer the reader to [45] for a detailed treatment of this subject in the framework of Bayesian statistics.

4 Transition Path Theory

Transition Path Theory (TPT) is concerned with transitions in Markov processes. The basic idea is to single out two disjoint subset in the state-space of the chain and ask what is the typical mechanism by which the dynamics transits from one of these states to the other. We may also ask at which rate these transitions occur.

The first object which comes to mind to characterize these transitions is the path of maximum likelihood by which they occur. However, this path can again be not very informative if the two states one has singled out are not metastable states. The main objective herein is to show that we can give a precise meaning to the question of finding typical mechanisms and rate of transition in discrete state spaces for continuous time processes which are neither metastable nor time-reversible. In a nutshell, given two subsets in state-space, TPT analyzes the statistical properties of the associated reactive trajectories, i.e., the trajectories by which transition occur between these sets. TPT provides information such as the probability distribution of these trajectories, their probability current and flux, and their rate of occurrence.

The framework of transition path theory (TPT) has first been developed in [46, 47, 48] in the context of diffusions. However, we will follow [49] and focus on continuous-time Markov chains, but we note that the results to outlined can be straightforwardly extended to the case of discrete-time Markov chains. In the next section we will illustrate TPT with an example from molecular dynamics, but the tools of TPT presented here can be used for data segmentation as well. In this context, TPT provides an alternative to Laplacian eigenmaps [50, 51] and diffusion maps [52, 53] which have become very popular recently in data analysis.

Notation. We consider a Markov jump process on the countable state space S with infinitesimal generator (or rate matrix) $L = (l_{ij})_{i,j\in S}$,

$$\begin{cases} l_{ij} \geq 0 & \text{for all } i, j \in S, i \neq j \\ \sum_{j\in S} l_{ij} = 0 & \text{for all } i \in S. \end{cases} \tag{35}$$

We assume that the thus defined process is irreducible and ergodic with respect to the unique, strictly positive invariant distribution $\pi = (\pi_i)_{i\in S}$ satisfying

$$0 = \pi^T L. \tag{36}$$

We will denote by $\{X_t\}$ a (right-continuous with left limits) trajectory of the Markov jump process. We also denote by $\{\tilde{X}_t\}$ the time-reversed process which has the same invariant distribution and an infinitesimal generator $\tilde{L} = (\tilde{l}_{ij})_{i,j\in S}$ given by

$$\tilde{l}_{ij} = \frac{\pi_j}{\pi_i} l_{ji}. \tag{37}$$

Finally, recall that if the infinitesimal generator satisfies the detailed balance equation $\pi_i l_{ij} = \pi_j l_{ji}$, the process is reversible and the direct and the time-reversed process are statistically undistinguishable. We do not assume reversibility in the subsequent.

Reactive trajectories. Let A and B two nonempty, disjoint subsets of the state space S. By ergodicity, any equilibrium path $\{X_t\}$ oscillates infinitely many times between set A and set B. If we view A as a reactant state and B as a product state, each oscillation from A to B is a reaction event. To properly define and characterize the reaction events, we proceed by cutting a long ergodic trajectory $\{X_t\}$ into pieces that each connect A and B. We shall then try to describe various statistical properties of the statistical ensemble of these pieces. For details on the pruning procedure, see [49].

Committors. The fundamental objects of TPT are the committor functions. The *discrete forward committor* $q^+ = (q_i^+)_{i\in S}$ is defined as the probability that the process starting in $i \in S$ will reach first B rather than A. Analogously, we define the *discrete backward committor* $q^- = (q_i^-)_{i\in S}$ as the probability that the process arriving in state i has been started in A rather than B. It has been proved in [49] that the forward and backward committor satisfy a discrete Dirichlet problem that is the exact finite-dimensional analogue of the respective continuous problem [46], namely,

$$\begin{cases} \sum_{j\in S} l_{ij} q_j^+ = 0, & \forall i \in (A \cup B)^c \\ q_i^+ = 0, & \forall i \in A \\ q_i^+ = 1, & \forall i \in B \end{cases} \tag{38}$$

and

$$\begin{cases} \sum_{j\in S} \tilde{l}_{ij} q_j^- = 0, & \forall i \in (A \cup B)^c \\ q_i^- = 1, & \forall i \in A \\ q_i^- = 0, & \forall i \in B \end{cases} \tag{39}$$

Probability distribution of reactive trajectories. The first relevant object for quantifying the statistical properties of the reactive trajectories is the *distribution of reactive trajectories* $m^R = (m_i^R)_{i\in S}$. The distribution m^R gives the equilibrium probability to observe a reactive trajectory at state i and time t. According to [49] the probability distribution of reactive trajectories is given by

$$m_i^R = \pi_i q_i^+ q_i^-, \quad i \in S. \tag{40}$$

Probability current of reactive trajectories. Next we are interested in the average current of reactive trajectories flowing from state i to state j per unit of time. This *probability current of reactive trajectories* $f^{AB} = (f_{ij}^{AB})_{i,j\in S}$ satisfies $f_{ii}^{AB} = 0$ for all $i \in S$ and is given by [49]

$$f_{ij}^{AB} = \begin{cases} \pi_i q_i^- l_{ij} q_j^+, & \text{if } i \neq j \\ 0, & \text{otherwise} \end{cases} \tag{41}$$

Transition rate and effective current. Further we may ask for the average number of transitions from A to B per time unit or, equivalently, the average number of reactive trajectories observed per unit of time (transition rate). That is, let N_T be the number of reactive trajectories in the interval $[-T, T]$ in time. The *transition rate* k_{AB} is defined as

$$k_{AB} = \lim_{T\to\infty} \frac{N_T}{2T}. \tag{42}$$

Due to [49] the transition rate is given by

$$k_{AB} = \sum_{i\in A, j\in S} f_{ij}^{AB} = \sum_{j\in S, k\in B} f_{jk}^{AB}. \tag{43}$$

Notice that the rate equals

$$k_{AB} = \sum_{i\in A, j\in S} f_{ij}^{+}, \tag{44}$$

where the *effective current* is defined as

$$f_{ij}^{+} = \max(f_{ij}^{AB} - f_{ji}^{AB}, 0). \tag{45}$$

Reaction Pathways. A *reaction pathway* $w = (i_0, i_1, \ldots, i_n),\ i_j \in S, j = 0, \ldots, n$ from A to B is a simple pathway with the property

$$i_0 \in A, i_n \in B, i_j \in (A \cup B)^c \quad j = 1, \ldots, n-1.$$

The crucial observation which leads to a characterization of bottlenecks of reaction pathways is that the amount of reactive trajectories which can be conducted by a reaction pathway per time unit is confined by the minimal effective current of a transition involved along the reaction pathway: the *min-current* of w is

$$c(w) = \min_{e=(i,j)\in w} \{f_{ij}^{+}\}. \tag{46}$$

Accordingly we shall characterize the "best" reaction pathway as the one with the *maximal min-current*, and, eventually, we can rank all reaction pathways according to the respective weight $c(w)$. Efficient graph algorithms for computing the hierarchy of transition pathways can be found in [49].

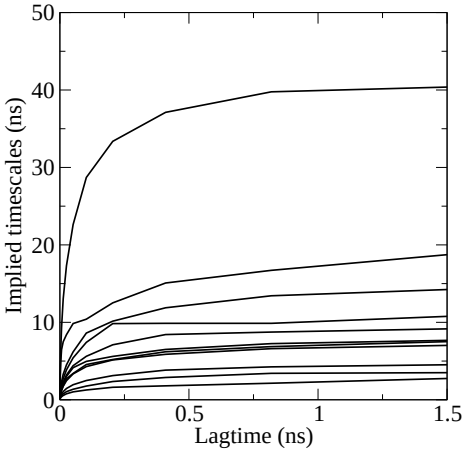

Figure 10: Slowest implied time scales of the MR121-GSGSW peptide dynamics.

5 Application to MD Simulations

The methods introduced in the Sections 2–4 are now illustrated with a biophysically relevant molecular dynamics example, the synthetic hexapeptide MR121 - GSGSW [54]. This is a linear polymer the central part (GSGS) of which contains a repeat of the Glycin and Serine amino acids that are found in the loop regions of many proteins. In order to study the folding of the loop, two additional chemical groups (MR121 and W/Tryptophan) were attached to the peptide in the experiments reported in [54]. These two groups contain ring systems which provide an experimentally detectable signal when forming immediate contact. Here, we study a 1 microsecond molecular dynamics (MD) simulation of this system which was performed in explicit water at the experimental temperature 293 K with the GROMACS software package [31] using the GROMOS96 force field [55]. During this simulation, the peptide frequently folds and unfolds and visits various different conformations. We shall analyse its conformational dynamics in the following.

To distinguish all relevant conformations of the system, the peptide coordinates were fitted to the extended structure, and the state space was partitioned into small regions using a K-means clustering with K=5000. In order to determine the lagtime at which transitions appear Markovian, the microscopic transition matrix $T^{\text{micro}}(\tau) \in \mathbf{R}^{5000\times 5000}$ was computed for different τ and the time scales, t_i^*, implied by the corresponding spectrum $\Lambda(\tau) = (\lambda_1(\tau), ..., \lambda_{5000}(\tau))$ were examined:

$$t_i^* = -\frac{\tau}{\log \lambda_i(\tau)}. \tag{47}$$

At lagtimes large enough for the dynamics to be Markovian, the implied time scales are expected to be constant in τ [56]. As visible from Figure 10, this is the case for about $\tau \geq 1$ ns.

In order to concentrate on the slow conformation dynamics, the Maximum-Likelihood transition matrix at $\tau = 1$ns is used with the PCCA algorithm [19, 24, 28] to cluster the 5000 microstates into 34 metastable states, providing a discrete

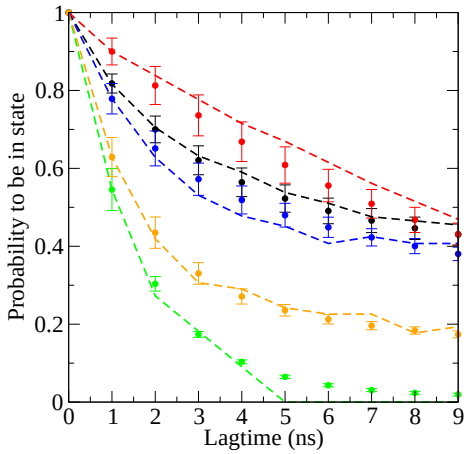

Figure 11: Relaxation of the population out of 5 representative metastable states. The predictions of the transition matrix model (bullets) agree, within the error, well with the actual observations from the simulation trajectory (dashed lines). The confidence intervals correspond to one standard deviation.

trajectory $X_t \in \{1, ..., 34\}$, $t \in [0, 1\mu s]$. The number 34 was arbitrarily chosen, such that the implied time scale of the 34-th eigenvalue is twice the characteristic time scale $\tau = 1$ns of the transition matrix. Since approximately constant time scales are not a sufficient condition for the dynamics to be Markovian, it is checked whether the transition matrix can actually reproduce the observed dynamics on long time scales. For this, the estimated Maximum-Likelihood transition matrix, $T(\tau)$ with $\tau = 1$ns, was used to compute the decrease of population of state i as a function of time, i.e.,

$$p_i(k\tau) = [T^k(\tau)]_{ii}. \tag{48}$$

This is then compared with the corresponding probability that is directly observed in the simulation trajectory,

$$\mathbf{P}[X_{t+k\tau} = i \mid X_t = i]. \tag{49}$$

The result is shown in Figure 11 for 5 representative states. It is seen that the predictions of the transition matrix model are similar to the actual observations from the simulation trajectory.

Since only a limitied number of transitions between metastable states is observed in the MD trajectory, the transition matrix is not uniquely determined but carries some statistical uncertainty. Consequently, the decay curves in Figure 11 are uncertain as well. In order to assess this uncertainty, the distribution of transition matrices (10) induced by the observed transition counts at $\tau = 1$ns, were sampled with a Monte Carlo algorithm [57]. For each matrix of the sample, the relaxation curves were computed using equation (48), and the resulting standard deviations of the distribution give rise to the confidence intervals in Figure 11. The deviations between the predictions from the transition matrix and the observations from the MD trajectory are mostly within 1 standard deviation (except for the green state, for which no long-time observations are

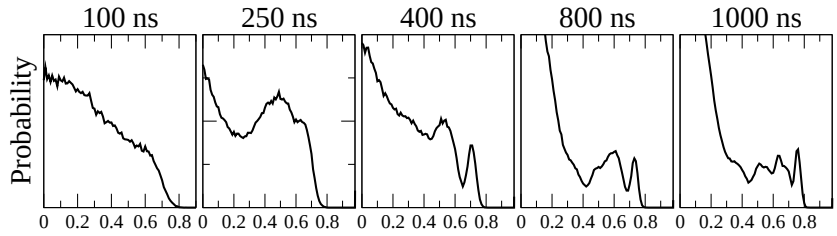

Figure 12: Distributions of the eigenvalue spectrum of T for different simulation lengths.

available in the MD trajectory), thus reassuring the reliability of the transition matrix model.

Next, the transition pathways between the conformational states of the system were studied. For this purpose, the core sets of the 34 conformations were identified as described in Section 2.3.5 and a transition matrix $T^{TP}(\tau)$ with, now, $\tau = 200\,\mathrm{fs}$) was computed employing equation (18) and then further subdividing the transition set in order to optimze metastability; the trajectory was used both forwards and backwards in time, such that the transition matrix is reversible and has a real spectrum. It was then verified that the long-time behaviour ($\Delta t \geq 1\mathrm{ns}$) of this transition matrix also agrees with the observations from the MD trajectory. Since all eigenvalues of T^{TP} are real and positive, its generator L could in principle be computed by taking the matrix logarithm. Since, additionally, the lagtime $\tau = 200\,\mathrm{fs}$ is very short as compared to the typical lifetimes of states, the transition matrix is metastable and all its eigenvalues are close to 1, such that the generator is well approximated by

$$L = \frac{1}{\tau} \log T^{TP}(\tau) \approx \frac{1}{\tau}(T^{TP}(\tau) - I)\,.$$

We study the slowest transition in the system by selecting the two conformations, A and B, with the largest positive and negative element in the second left eigenvector. As is shown in Figure 13, A and B correspond to structures in which the loop is closed and the ring systems of the end-groups are in contact. Hence $A \to B$ corresponds to an exchange of the stacking order of the end-groups, and we can use TPT to study the set of transition pathways for this process. Employing (38) we compute the discrete committor function q for the transition $A \to B$ for all 34 states. Then, from q and L, the TPT effective currents to the transition $A \to B$ are obtained according to (45). The resulting flux network for $A \to B$ is complex, involving significant transition pathways via most of the 32 intermediate states. The network of the 30% most populated transition pathways is shown in Figure 13. It turns out that the most populated pathway is in fact the direct transition, but other pathways are also significant, including pathways via closed-loop intermediates and pathways via unfolded intermediates. Eventually, the total transition rate obtained by TPT is similar to the experimentally-measured slowest rate of the system.

Finally, we emphasize that statistical errors due to lack of convergence is an important issue regarding all MD simulations: By sampling the sampling error of the estimated transition matrix, we have examined how the length of the

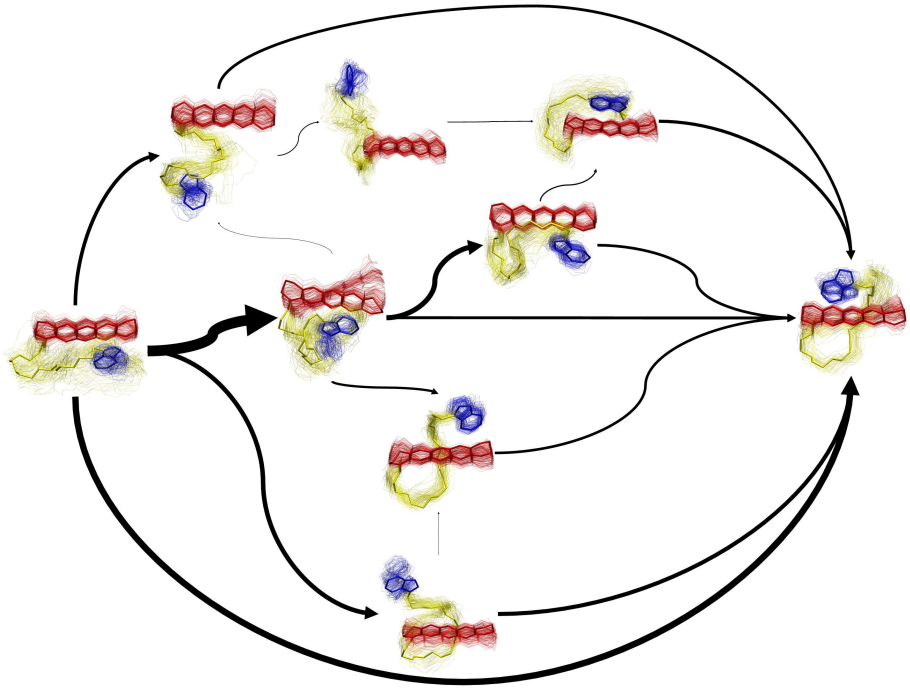

Figure 13: The network of the 30% most populated transition pathways for the slowest transition process in the MR121-GSGSW peptide. The thickness of arrows is proportional to the net flux along each edge. The loop segment is shown in yellow, the MR121 and W end-groups in red and blue, respectively. The transition end-states A and B are shown on the left and right side.

MD trajectory affects the uncertainties of any quantity computed from T. In particular, we were interested in the spectrum of T for fixed lagtime of $\tau = 1$ns, i.e., its eigenvalues $\Lambda = (\lambda_1 ... \lambda_{34})$ and the implied time scales of the transition processes, t_i^*, that were computed according to equation (47). The time scales $t_2^*, t_3^*, ...$ thereby correspond to the time scales of the slowest and next-slowest transition processes. However, since there is a whole distribution of transition matrices T, the spectrum of eigenvalues for a given observed transition count C is not unique, but rather a distribution of spectra. Certainly, the distribution of the λ_i will get sharper as the number of observed transitions increases, thus explaining that some eigenvalues λ_i are sharper than others. Figure 12 shows the spectral distribution for several simulation lengths. For simulation times up to 100ns, the spectral distribution has no distinctive features. But as the simulation length is increased, some of the larger eigenvalues become distinguishable. From 400ns on, the slowest transition process at $\lambda_2 \approx 0.75$ can be clearly distinguished and continues to narrow as the simulation gets longer. At 1000ns, eventually, the spectrum exhibits a lot of structure in the range $\lambda \geq 0.5$. However, apart from λ_2 no peaks are clearly separated which indicates that even for our small peptide, 1μs simulation time is rather short if one wants to obtain good convergence of the kinetics (e.g., rates).

References

[1] A. Ostermann, R. Waschipky, F. G. Parak, and G. U. Nienhaus. Ligand binding and conformational motions in myoglobin. *Nature* **404**, pp. 205–208, 2000.

[2] S. Fischer, B. Windshügel, D. Horak, K. C. Holmes, and J. C. Smith. Structural mechanism of the recovery stroke in the myosin molecular motor. *Proc. Natl. Acad. Sci. USA* **102**, pp. 6873–6878, 2005.

[3] F. Noé and D. Krachtus and J. C. Smith and S. Fischer. Transition networks for the comprehensive characterization of complex conformational change in proteins. *J. Chem. Theory and Comput.* **2**, pp. 840–857, 2006.

[4] M. Jäger, Y. Zhang, J. Bieschke, H.Nguyen, M. Dendle, M. E. Bowman, J. Noel, M. Gruebele, and J. Kelly. Structure-function-folding relationship in a ww domain. *Proc. Natl. Acad. Sci. USA* **103**, pp. 10648–10653, 2006.

[5] A. Y. Kobitski, A. Nierth, M. Helm, A. Jäschke, and G. U. Nienhaus. Mg^2+-dependent folding of a Diels-Alderase ribozyme probed by single-molecule FRET analysis. *Nucleic Acids Res.* **35**, pp. 2047–2059, 2007.

[6] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes. The energy landscapes and motions of proteins. *Science* **254**, pp. 1598–1603, 1991.

[7] D. Wales. *Energy Landscapes*. Cambridge University Press, Cambridge, 2003.

[8] N. Singhal and V. S. Pande. Error analysis and efficient sampling in Markovian state models for molecular dynamics. *J. Chem. Phys.* **123**, p. 204909, 2005.

[9] N. S. Hinrichs and V. S. Pande. Calculation of the distribution of eigenvalues and eigenvectors in Markovian state models for molecular dynamics. *J. Chem. Phys.* **126**, p. 244101, 2007.

[10] F. Noé, M. Oswald, G. Reinelt, S. Fischer, and J. C. Smith. Computing best transition pathways in high-dimensional dynamical systems: application to the $\alpha_L \leftrightharpoons \beta \leftrightharpoons \alpha_R$ transitions in octaalanine. *Mult. Mod. Sim.* **5**, pp. 393–419, 2006.

[11] F. Noé, M. Oswald, and G. Reinelt Optimizing in graphs with expensive computation of edge weights. *to appear in Operations Research*, 2007.

[12] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability.* Springer, 1993.

[13] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in stochastic dynamics of disordered mean-field models. *Probab. Theor. Rel. Fields* **119**, pp. 99–161, 2001.

[14] E.B. Davies. Metastable states of symmetric Markov semigroups I. *Proc. London Math. Soc.* **45**(3), pp. 133–150, 1982.

[15] Ch. Schütte, W. Huisinga, and P. Deuflhard. Transfer operator approach to conformational dynamics in biomolecular systems. In: B. Fielder (ed.) *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems*, pp. 191–223, Springer, 2001.

[16] G. Singleton. Asymptotically exact estimates for metastable Markov semigroups. *Quart. J. Math. Oxford* **35**(2), pp. 321–329, 1984.

[17] Ch. Schütte and W. Huisinga. Biomolecular conformations can be identified as metastable sets of molecular dynamics. In: P. G. Ciarlet and C. Le Bris (eds.) *Handbook of Numerical Analysis*, vol. X, pp. 699-744, Elsevier, 2003.

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

[19] Ch. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard. A direct approach to conformational dynamics based on hybrid Monte Carlo. *J. Comput. Phys., Special Issue on Computational Biophysics* **151**, pp. 146–168, 1999.

[20] Ch. Schütte and W. Huisinga. On conformational dynamics induced by Langevin processes. In: B. Fiedler, K. Gröger, and J. Sprekels (eds.) *EQUADIFF 99 - International Conference on Differential Equations*, vol. 2, pp. 1247–1262, World Scientific, 2000.

[21] W. Huisinga and B. Schmidt. Metastability and dominant eigenvalues of transfer operators. In: C. Chipot, R. Elber, A. Laaksonen, B. Leimkuhler, A. Mark, T. Schlick, C. Schütte, and R. Skeel (eds.) *Advances in Algorithms for Macromolecular Simulation*, LNCSE, vol. 49, pp. 167–182, Springer, 2005.

[22] W. Huisinga, S. Meyn, and Ch. Schütte. Phase transitions & metastability in Markovian and molecular systems. *Ann. Appl. Probab.* **14**, pp. 419-458, 2004.

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

[24] P. Deuflhard and M. Weber. Robust Perron cluster analysis in conformation dynamics. *Lin. Alg. App.* **398**(c), pp. 161–184, 2005.

[25] P. Lezaud. Chernoff and Berry-Esséen inequalities for Markov processes. *ESIAM: P & S* **5**, pp. 183–201, 2001.

[26] B.J. Berne and J.E. Straub. Novel methods of sampling phase space in the simulation of biological systems. *Curr. Opinion in Struct. Biol.*. **7**, pp. 181–189, 1997.

[27] D.M. Ferguson, J.I. Siepmann, and D.G. Truhlar (eds.). *Monte Carlo Methods in Chemical Physics*, Advances in Chemical Physics, vol. 105. Wiley, 1999.

[28] F. Noé, I. Horenko, C. Schütte, and J. C. Smith. Hierarchical analysis of conformational dynamics in biomolecules: transition networks of metastable states. *J. Chem. Phys.*. **126**, 155102, 2007.

[29] J. Chodera, N. Singhal, V. Pande, K. Dill, and W. Swope. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Comp. Chem.*, 126:155101, 2007.

[30] I. Horenko, C. Hartmann, Ch. Schütte, and F. Noé. Data-based parameter estimation of generalized multidimensional Langevin processes. *Phys. Rev. E* **76**(1), 016706, 2007.

[31] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: Fast, flexible and free. *J. Comp. Chem.*, 26:1701–1718, 2005.

[32] J.S. Liu. *Monte Carlo Strategies in Scientific Computing.* Springer, New York, 2001.

[33] I. Maravic and M. Vetterli. Exact sampling results for some classes of parametric nonbandlimited 2-D signals. *IEEE Trans. Signal Proc.* **52** (1), pp. 175–189, 2004.

[34] I. Maravic and M. Vetterli. Sampling and reconstruction of signals with finite rate of innovation in the presence of noise. *IEEE Trans. Signal Proc.* **53**(8), pp. 2788-2805, 2005.

[35] D. Achlioptas, F. McSherry, B. Schölkopf. Sampling techniques for kernel methods. In: T.G. Diettrich, S. Becker, and Z. Ghahramani (eds.) *Advances in Neural Information Processing Systems*, vol. 14, pp. 335–342, MIT Press, 2002.

[36] N.-V. Buchete and G. Hummer. Coarse master equations for peptide folding dynamics. *to appear in J. Phys. Chem. B*, 2008.

[37] A. Fischer, S. Waldhausen, I. Horenko, E. Meerbach, and Ch. Schütte. Identification of biomolecular conformations from incomplete torsion angle observations by Hidden Markov Models. *J. Comp. Chem.* **28**(8), pp. 1384–1399, 2007.

[38] I. Horenko and Ch. Schütte. Likelihood-based estimation of multidimensional Langevin models and its application to biomolecular dynamics. *to appear in Mult. Mod. Sim.*, 2008.

[39] I. Horenko, E. Dittmer, F. Lankas, J. Maddocks, P. Metzner, and Ch. Schütte. Macroscopic dynamics of complex metastable systems: theory, algorithms, and application to b-DNA. *to appear in SIADS*, 2008.

[40] J. Bilmes. *A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models.* ICSI-TR-97-021, 1997.

[41] I. Horenko, E. Dittmer, A. Fischer, and Ch. Schütte. Automated model reduction for complex systems exhibiting metastability. *Mult. Mod. Sim.* **5**(3), pp. 802–827, 2006.

[42] A.J. Viterbi. Error bounds for convolutional codes and an asymptotically optimal decoding algorithm. *IEEE Trans. Inf. Process.* **13**, pp. 260–269, 1967.

[43] H. Lütkepohl. *Introduction to Multiple Time Series Analysis.* Springer, 1991.

[44] S. Ni and D. Sun. Bayesian estimates for vector autoregressive models. *JBES* **23**, pp. 105–117, 2005.

[45] E. Meerbach and Ch. Schütte. Sequential change-point detection in molecular dynamics trajectories. *submitted*, 2008.

[46] W. E and E. Vanden-Eijnden. Towards a theory of transition paths. *J. Stat. Phys.* **123**, pp. 503–523, 2006.

[47] E. Vanden-Eijnden. Transition path theory. In: M. Ferrario, G. Ciccotti, and K. Binder (eds.) *Computer Simulations in Condensed Matter: From Materials to Chemical Biology*, vol. 2, LNP **703**, pp. 439–478, Springer, 2006.

[48] 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.

[49] Ph. Metzner, Ch. Schütte, and E. Vanden-Eijnden. Transition Path Theory for Markov Jump Processes. *submitted to Mult. Mod. Sim.*, 2007.

[50] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, **15**(6), pp. 1373–1396, 2003.

[51] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In: T.G. Diettrich, S. Becker, and Z. Ghahramani (eds.) *Advances in Neural Information Processing Systems*, vol. 14, pp. 585–591, MIT Press, 2002.

[52] B. Nadler, S. Lafon, R.R. Coifman, I.G. Kevrekidis. Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. Appl. Comput. Harmon. Anal. **21**, pp.113–127, 2006.

[53] S. Lafon and A.B. Lee. Diffusion maps and coarse-graining: a unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(9), pp. 1393–1403, 2006.

[54] H. Neuweiler, M. Löllmann, S. Doose, and M. Sauer. Dynamics of unfolded polypeptide chains in crowded environment studied by fluorescence correlation spectroscopy. *J. Mol. Biol.* **365**, pp. 856–869, 2007.

[55] W. F. van Gunsteren and H. J. C. Berendsen. Computer simulation of molecular dynamics: Methodology, applications and perspectives in chemistry. *Angew. Chem. Int. Ed. Engl.* **29**, pp. 992–1023, 1990.

[56] W. C. Swope, J. W. Pitera, and F. Suits. Describing protein folding kinetics by molecular dynamics simulations: 1. theory. *J. Phys. Chem. B* **108**, pp. 6571–6581, 2004.

[57] F. Noé. Probability Distributions of Molecular Observables computed from Markov Models. *to appear in J. Chem. Phys.*, 2008.

[58] M. Weber. Meshless Methods in Confirmation Dynamics. Dissertation thesis, FU Berlin, 2006