

## 3.1 Continuous Molecular Dynamics

A variety of simulation models that all yield the same stationary properties, but have different dynamical behaviors, are available to study a given molecular model. The choice of the dynamical model must therefore be guided by both a desire to mimic the relevant physics for the system of interest (such as whether the system is allowed to exchange energy with an external heat bath during the course of dynamical evolution), balanced with computational convenience (e.g. the use of a stochastic thermostat in place of explicitly simulating a large external reservoir) [8]. Going into the details of these models is beyond the scope of the present study, and therefore we will simply state the minimal physical properties that we expect the dynamical model to obey. In the following we pursue the theoretical outline from Ref. [31] (Sects. 3.1–3.7) and Ref. [37] (Sects. 3.1–3.8) which should both be used for reference purposes.

Consider a state space  $\Omega$  which contains all dynamical variables needed to describe the instantaneous state of the system.  $\Omega$  may be discrete or continuous, and we treat the more general continuous case here. For molecular systems,  $\Omega$  usually contains both positions and velocities

of the species of interest and surrounding bath particles.  $\mathbf{x}(t) \in \Omega$  will denote the state of the system at time  $t$ . The dynamical process considered is  $(\mathbf{x}(t))_{t \in T}$ ,  $T \subset \mathbb{R}_{0+}$ , which is continuous in space, and may be either time-continuous (for theoretical investigations) or time-discrete (when considering time-stepping schemes for computational purposes). For the rest of the article, the dynamical process will also be denoted by  $\mathbf{x}(t)$  for the sake of simplicity; we assume that  $\mathbf{x}(t)$  has the following properties:

1.  $\mathbf{x}(t)$  is a Markov process in the full state space  $\Omega$ , i.e. the instantaneous change of the system ( $d\mathbf{x}(t)/dt$  in time-continuous dynamics and  $\mathbf{x}(t + \Delta t)$  in time-discrete dynamics with time step  $\Delta t$ ), is calculated based on  $\mathbf{x}(t)$  alone and does not require the previous history. In addition, we assume that the process is time-homogeneous, such that the transition probability density  $p(\mathbf{x}, \mathbf{y}; \tau)$  for  $\mathbf{x}, \mathbf{y} \in \Omega$  and  $\tau \in \mathbb{R}_{0+}$  is well-defined:

$$p(\mathbf{x}, A; \tau) = \mathbb{P}[\mathbf{x}(t + \tau) \in A \mid \mathbf{x}(t) = \mathbf{x}] \quad (3.1)$$

i.e. the probability that a trajectory started at time  $t$  from the point  $\mathbf{x} \in \Omega$  will be in set  $A$  at time  $t + \tau$ . Such a transition probability density for the diffusion process in a one-dimensional potential is depicted in Fig. 3.1b. Whenever  $p(\mathbf{x}, A; \tau)$  has an absolutely continuous probability density  $p(\mathbf{x}, \mathbf{y}; \tau)$  it is given by integrating the transition probability density over region  $A$ :

M. Sarich · J.-H. Prinz · C. Schütte (✉)  
 Freie Universität Berlin, Arnimallee 6, 14195 Berlin,  
 Germany  
 e-mail: [christof.schuette@fu-berlin.de](mailto:christof.schuette@fu-berlin.de)

$$p(\mathbf{x}, A; \tau) = \mathbb{P}[\mathbf{x}(t + \tau) \in A \mid \mathbf{x}(t) = \mathbf{x}] \quad (3.2)$$

$$= \int_A d\mathbf{y} p(\mathbf{x}, \mathbf{y}; \tau). \quad (3.3)$$

2.  $\mathbf{x}(t)$  is ergodic, i.e., the process  $\mathbf{x}(t)$  is aperiodic, the space  $\Omega$  does not have two or more subsets that are dynamically disconnected, and for  $t \rightarrow \infty$  each state  $\mathbf{x}$  will be visited infinitely often. The running average of a function  $f : \Omega \rightarrow \mathbb{R}^d$  then is given by a unique stationary density  $\mu(\mathbf{x})$  in the sense that for almost every initial state  $\mathbf{x}$  we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt f(\mathbf{x}(t)) = \int_{\Omega} d\mathbf{x} f(\mathbf{x}) \mu(\mathbf{x}),$$

that is, the fraction of time that the system spends in any of its states during an infinitely long trajectory is given by the stationary density (invariant measure)  $\mu(\mathbf{x}) : \Omega \rightarrow \mathbb{R}_{0+}$  with  $\int_{\Omega} d\mathbf{x} \mu(\mathbf{x}) = 1$ , where the stationarity of the density means that

$$\int_A d\mathbf{x} \mu(\mathbf{x}) = \int_{\Omega} d\mathbf{x} p(\mathbf{x}, A; \tau) \mu(\mathbf{x}),$$

which takes the simpler form

$$\mu(\mathbf{y}) = \int_{\Omega} d\mathbf{x} p(\mathbf{x}, \mathbf{y}; \tau) \mu(\mathbf{x}),$$

whenever the transition probability has a density. We assume that this stationary density  $\mu$  is unique. In cases relevant for molecular dynamics the stationary density always corresponds to the equilibrium probability density for some associated thermodynamic ensemble (e.g. NVT, NpT). For molecular dynamics at constant temperature  $T$ , the dynamics above yield a stationary density  $\mu(\mathbf{x})$  that is a function of  $T$ , namely the Boltzmann distribution

$$\mu(\mathbf{x}) = Z(\beta)^{-1} \exp(-\beta H(\mathbf{x})) \quad (3.4)$$

with Hamiltonian  $H(\mathbf{x})$  and  $\beta = 1/k_B T$  where  $k_B$  is the Boltzmann constant and  $k_B T$  is the thermal energy.  $Z(\beta) = \int d\mathbf{x} \exp(-\beta H(\mathbf{x}))$  is the partition function. By means of illustration, Fig. 3.1a shows the stationary density  $\mu(\mathbf{x})$  for a diffusion process on a potential with high barriers.

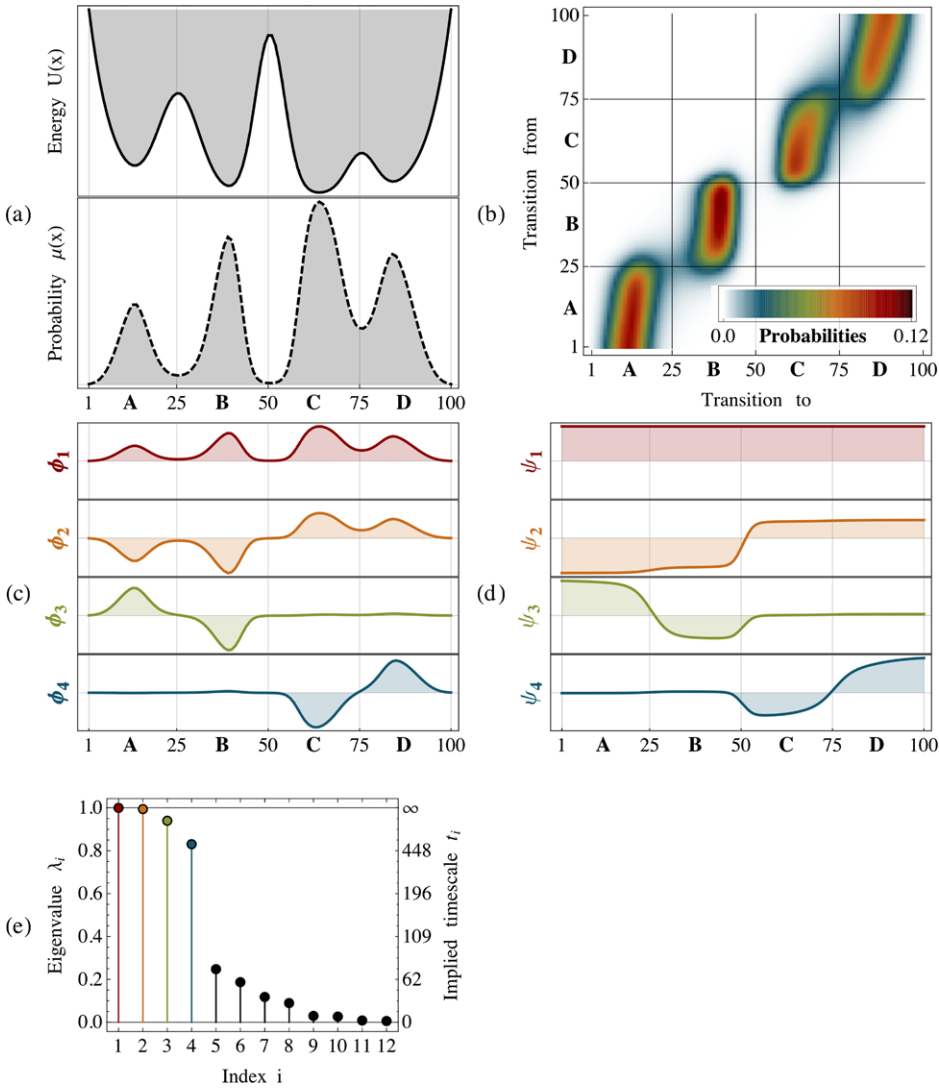
3.  $\mathbf{x}(t)$  is reversible, i.e.,  $p(\mathbf{x}, \mathbf{y}; \tau)$  fulfills the condition of *detailed balance*:

$$\mu(\mathbf{x}) p(\mathbf{x}, \mathbf{y}; \tau) = \mu(\mathbf{y}) p(\mathbf{y}, \mathbf{x}; \tau), \quad (3.5)$$

i.e., in equilibrium, the fraction of systems transitioning from  $\mathbf{x}$  to  $\mathbf{y}$  per time is the same as the fraction of systems transitioning from  $\mathbf{y}$  to  $\mathbf{x}$ . Note that this “reversibility” is a more general concept than the time-reversibility of the dynamical equations e.g. encountered in Hamiltonian dynamics. For example, Brownian dynamics on some potential are reversible as they fulfill Eq. (3.5), but are not time-reversible in the same sense as Hamiltonian dynamics are, due to the stochasticity of individual realizations. Although detailed balance is not essential for the construction of Markov models, we will subsequently assume detailed balance as this allows much more profound analytical statements to be made, and just comment on generalizations here and there. The rationale is that one typically expects detailed balance to be fulfilled in equilibrium molecular dynamics based on a simple physical argument: For dynamics that have no detailed balance, there exists a set of states which form a loop in state space which is traversed in one direction with higher probability than in the reverse direction. This means that one could design a machine which uses this preference of direction in order to produce work. However, a system in equilibrium is driven only by thermal energy, and conversion of pure thermal energy into work contradicts the second law of thermodynamics. Thus, this argument concludes that equilibrium molecular dynamics must be reversible and fulfill detailed balance. Despite the popularity of this argument there are dynamical processes used in molecular dynamics that do *not* satisfy detailed balance in the above sense. Langevin molecular dynamics may be the most prominent example. However, the Langevin process exhibits an *extended detailed balance* [18]

$$\mu(\mathbf{x}) p(\mathbf{x}, \mathbf{y}; \tau) = \mu(A\mathbf{y}) p(A\mathbf{y}, A\mathbf{x}; \tau),$$

where  $A$  is the linear operation that flips the sign of the momenta in the state  $\mathbf{x}$ . This prop-



**Fig. 3.1** (a) Potential energy function with four metastable states and corresponding stationary density  $\mu(x)$ . (b) Density plot of the transfer operator for a simple diffusion-in-potential dynamics defined on the range  $\Omega = [0, 100]$ , *black* and *red* indicates high transition probability, *white* zero transition probability. Of particular interest is the nearly block-diagonal structure, where the transition density is large within blocks allowing rapid transitions within metastable basins, and small or nearly zero for jumps between different metastable basins. (c) The

four dominant eigenfunctions of the transfer operator,  $\psi_1, \dots, \psi_4$ , which indicate the associated dynamical processes. The first eigenfunction is associated to the stationary process, the second to a transition between  $A + B \leftrightarrow C + D$  and the third and fourth eigenfunction to transitions between  $A \leftrightarrow B$  and  $C \leftrightarrow D$ , respectively. (d) The four dominant eigenfunctions of the transfer operator weighted with the stationary density,  $\phi_1, \dots, \phi_4$ . (e) Eigenvalues of the transfer operator, The gap between the four metastable processes ( $\lambda_i \approx 1$ ) and the fast processes is clearly visible

erty allows to reproduce the results for processes with detailed balance to this case, see remarks below.

The above conditions do not place overly burdensome restrictions on the choices of dynamical models used to describe equilibrium dynamics. Many stochastic thermostats are consistent with the above assumptions, e.g. Hybrid Monte Carlo [14, 37], overdamped Langevin (also called Brownian or Smoluchowski) dynamics [15, 16], and stepwise-thermalized Hamiltonian dynamics [42]. When simulating solvated systems, a weak friction or collision rate can be used; this can often be selected in a manner that is physically motivated by the heat conductivity of the material of interest and the system size [1].

We note that the use of finite-timestep integrators for these models of dynamics can sometimes be problematic, as the phase space density sampled can differ from the density desired. Generally, integrators based on symplectic Hamiltonian integrators (such as velocity Verlet [42]) offer greater stability for our purposes.

While technically, a Markov model analysis can be constructed for any choice of dynamical model, it must be noted that several popular dynamical schemes violate the assumptions above, and using them means that one is (currently) doing so without a solid theoretical basis, e.g. regarding the boundedness of the discretization error analyzed in Sect. 3.3 below. For example, Nosé-Hoover and Berendsen are either not ergodic or do not generate the correct stationary distribution for the desired ensemble [10]. Energy-conserving Hamiltonian dynamics on one hand may well be ergodic regarding the projected volume measure on the energy surface but this invariant measure is *not* unique, and on the other hand it is *not* ergodic wrt. the equilibrium probability density for some associated thermodynamic ensemble of interest.

### 3.2 Transfer Operator Approach and the Dominant Spectrum

At this point we shift from focusing on the evolution of individual trajectories to the time evolution of an ensemble density. Consider an en-

semble of molecular systems at a point in time  $t$ , distributed in state space  $\Omega$  according to a probability density  $p_t(\mathbf{x})$  that is different from the stationary density  $\mu(\mathbf{x})$ . If we now wait for some time  $\tau$ , the probability distribution of the ensemble will have changed because each system copy undergoes transitions in state space according to the transition probability density  $p(\mathbf{x}, \mathbf{y}; \tau)$ . The change of the probability density  $p_t(\mathbf{x})$  to  $p_{t+\tau}(\mathbf{x})$  can be described with the action of a continuous operator. From a physical point of view, it seems straightforward to define the *propagator*  $\mathcal{Q}(\tau)$  as follows:

$$p_{t+\tau}(\mathbf{y}) = \mathcal{Q}(\tau) \circ p_t(\mathbf{y}) \quad (3.6)$$

$$= \int_{\Omega} d\mathbf{x} p(\mathbf{x}, \mathbf{y}; \tau) p_t(\mathbf{x}). \quad (3.7)$$

Applying  $\mathcal{Q}(\tau)$  to a probability density  $p_t(\mathbf{x})$  will result in a modified probability density  $p_{t+\tau}(\mathbf{x})$  that is more similar to the stationary density  $\mu(\mathbf{x})$ , to which the ensemble must relax after infinite time. An equivalent description is provided by the *transfer operator*  $\mathcal{T}(\tau)$  [36, 37], which has nicer properties from a mathematical point of view.  $\mathcal{T}(\tau)$  is defined as [35, 36, 39]:

$$u_{t+\tau}(\mathbf{y}) = \mathcal{T}(\tau) \circ u_t(\mathbf{y}) \quad (3.8)$$

$$= \frac{1}{\mu(\mathbf{y})} \int_{\Omega} d\mathbf{x} p(\mathbf{x}, \mathbf{y}; \tau) \mu(\mathbf{x}) u_t(\mathbf{x}). \quad (3.9)$$

$\mathcal{T}(\tau)$  does not propagate probability densities, but instead functions  $u_t(\mathbf{x})$  that differ from probability densities by a factor of the stationary density  $\mu(\mathbf{x})$ , i.e.:

$$p_t(\mathbf{x}) = \mu(\mathbf{x}) u_t(\mathbf{x}). \quad (3.10)$$

The relationship between the two densities and operators is shown in the scheme below:

$$\begin{array}{ccc} p_t & \xrightarrow{\mathcal{Q}(\tau)} & p_{t+\tau} \quad \text{probability densities} \\ \downarrow \cdot \mu^{-1} & & \uparrow \cdot \mu \\ u_t & \xrightarrow{\mathcal{T}(\tau)} & u_{t+\tau} \quad \text{densities in } \mu\text{-weighted space} \end{array}$$

It is important to note that  $\mathcal{Q}$  and  $\mathcal{T}$  in fact do *not* only propagate probability densities but *general* functions  $f : \Omega \rightarrow \mathbb{R}$ . Since both operators have the property to conserve positivity and mass,

a probability density is always transported into a probability density.

Alternatively to  $\mathcal{Q}$  and  $\mathcal{T}$  which describe the transport of densities exactly by a chosen time-discretization  $\tau$ , one could investigate the density transport with a time-continuous operator  $\mathcal{L}$  called *generator* which is the continuous basis of rate matrices that are frequently used in physical chemistry [5, 40, 41] and is related to the Fokker-Planck equation [22, 36]. Here, we do not investigate  $\mathcal{L}$  in detail, but only point out that the existence of a generator implies that we have

$$\mathcal{T}(\tau) = \exp(\tau\mathcal{L}), \quad (3.11)$$

with  $\mathcal{L}$  acting on the same  $\mu$ -weighted space as  $\mathcal{T}$ , while  $\mathcal{Q}(\tau) = \exp(\tau L)$  for  $L$  acting on the unweighted densities/functions. The so-called semigroup-property (3.11) implies that  $\mathcal{T}(\tau)$  and  $\mathcal{L}$  have the same eigenvectors, while the eigenvalues  $\lambda$  of  $\mathcal{T}(\tau)$  and the eigenvalues  $\eta$  of  $\mathcal{L}$  are related via  $\lambda = \exp(\tau\eta)$ . This is of importance since most of the following considerations using  $\mathcal{T}(\tau)$  can be generalized to  $\mathcal{L}$ .

Equation (3.9) is a formal definition. When the particular kind of dynamics is known it can be written in a more specific form [37]. However, the general form (3.9) is sufficient for the present analysis. The continuous operators have the following general properties:

- Both  $\mathcal{Q}(\tau)$  and  $\mathcal{T}(\tau)$  fulfill the Chapman-Kolmogorov Equation

$$p_{t+k\tau}(\mathbf{x}) = [\mathcal{Q}(\tau)]^k \circ p_t(\mathbf{x}), \quad (3.12)$$

$$u_{t+k\tau}(\mathbf{x}) = [\mathcal{T}(\tau)]^k \circ u_t(\mathbf{x}) \quad (3.13)$$

where  $[\mathcal{T}(\tau)]^k$  refers to the  $k$ -fold application of the operator, i.e.  $\mathcal{Q}(\tau)$  and  $\mathcal{T}(\tau)$  can be used to propagate the evolution of the dynamics to arbitrarily long times  $t + k\tau$ .

- We consider the two operators on the Hilbert space of square integrable functions. More specifically, we work with two Hilbert spaces, one with unweighted functions,

$$L^2 = \left\{ u : \Omega \rightarrow \mathbb{C} : \right. \\ \left. \|u\|_2^2 = \int_{\Omega} d\mathbf{x} |u(\mathbf{x})|^2 < \infty \right\},$$

in which we consider  $\mathcal{Q}$ , the other with  $\mu$ -weighted functions

$$L_{\mu}^2 = \left\{ u : \Omega \rightarrow \mathbb{C} : \right. \\ \left. \|u\|_{2,\mu}^2 = \int_{\Omega} d\mathbf{x} |u(\mathbf{x})|^2 \mu(\mathbf{x}) < \infty \right\},$$

where we consider  $\mathcal{T}$ . These spaces come with the following two scalar products

$$\langle u, v \rangle = \int_{\Omega} d\mathbf{x} u(\mathbf{x})^* v(\mathbf{x}),$$

$$\langle u, v \rangle_{\mu} = \int_{\Omega} d\mathbf{x} u(\mathbf{x})^* v(\mathbf{x}) \mu(\mathbf{x}),$$

where the star indicates complex conjugation.

- $\mathcal{Q}(\tau)$  has eigenfunctions  $\phi_i(\mathbf{x})$  and associated eigenvalues  $\lambda_i$  (see Figs. 3.1c and e):

$$\mathcal{Q}(\tau) \circ \phi_i(\mathbf{x}) = \lambda_i \phi_i(\mathbf{x}), \quad (3.14)$$

while  $\mathcal{T}(\tau)$  has eigenfunctions  $\psi_i(\mathbf{x})$  with the same corresponding eigenvalues:

$$\mathcal{T}(\tau) \circ \psi_i(\mathbf{x}) = \lambda_i \psi_i(\mathbf{x}). \quad (3.15)$$

When the dynamics are reversible, all eigenvalues  $\lambda_i$  are real-valued and lie in the interval  $-1 < \lambda_i \leq 1$  [36, 37] (this is only true in  $L_{\mu}^2$  and not in the other function spaces). Moreover, the two types of eigenfunctions are related by a factor of the stationary density  $\mu(\mathbf{x})$ :

$$\phi_i(\mathbf{x}) = \mu(\mathbf{x}) \psi_i(\mathbf{x}), \quad (3.16)$$

and their lengths are defined by the normalization condition that the scalar product is unity for all corresponding eigenfunctions:

$$\langle \phi_i, \psi_i \rangle = \langle \psi_i, \psi_i \rangle_{\mu} = 1$$

for all  $i = 1 \dots m$ . Due to reversibility, non-corresponding eigenfunctions are orthogonal:

$$\langle \phi_i, \psi_j \rangle = 0$$

for all  $i \neq j$ . When  $\mathcal{T}(\tau)$  is approximated by a reversible transition matrix on a discrete state space,  $\phi_i(\mathbf{x})$  and  $\psi_i(\mathbf{x})$  are approximated by

the left and right eigenvectors of that transition matrix, respectively (compare Figs. 3.1c and d).

- In general the spectrum of the two operators contains a continuous part, called the essential spectrum, and a discrete part, called the discrete spectrum that contains only isolated eigenvalues [19]. The essential spectral radius  $0 \leq r \leq 1$  is the minimal value such that for all elements  $\lambda$  of the essential spectrum we have  $|\lambda| \leq r$ . In all of the following we assume that the essential spectral radius is bounded away from 1 in  $L^2_\mu$ , that is,  $0 \leq r < 1$ . Then, every element  $\lambda$  of the spectrum with  $|\lambda| > r$  is in the discrete spectrum, i.e., is an isolated eigenvalue for which an eigenvector exists. Our assumption is not always satisfied but is a condition on the dynamics: For example, for deterministic Hamiltonian systems it is  $r = 1$ , while for Langevin dynamics with periodic boundary conditions or with fast enough growing potential at infinity, we have  $r = 0$ . In the following, we assume  $r < 1$  and ignore the essential spectrum; we only consider a finite number of  $m$  isolated, *dominant* eigenvalue/eigenfunction pairs and sort them by non-ascending eigenvalue, i.e.  $\lambda_1 = 1 > \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_m$ , with  $r < |\lambda_m|$ . In addition we assume that the largest eigenvalue  $\lambda = 1$  is simple so that  $\mu$  is the only invariant measure.
- The eigenfunction associated with the largest eigenvalue  $\lambda = 1$  corresponds to the stationary distribution  $\mu(\mathbf{x})$  (see Fig. 3.1d, top):

$$\mathcal{Q}(\tau) \circ \mu(\mathbf{x}) = \mu(\mathbf{x}) = \phi_1(\mathbf{x}), \quad (3.17)$$

and the corresponding eigenfunction of  $\mathcal{T}(\tau)$  is a constant function on all state space  $\Omega$  (see Fig. 3.1c, top):

$$\mathcal{T}(\tau) \circ \mathbf{1} = \mathbf{1} = \psi_1(\mathbf{x}), \quad (3.18)$$

due to the relationship  $\phi_1(\mathbf{x}) = \mu(\mathbf{x})\psi_1(\mathbf{x}) = \mu(\mathbf{x})$ .

To see the significance of the other eigenvalue/eigenfunction pairs, we exploit that the dynamics can be decomposed exactly into a superposition of  $m$  individual slow dynamical processes and the remaining fast processes. For

$\mathcal{T}(\tau)$ , this yields:

$$u_{t+k\tau}(\mathbf{x}) = \mathcal{T}_{\text{slow}}(k\tau) \circ u_t(\mathbf{x}) + \mathcal{T}_{\text{fast}}(k\tau) \circ u_t(\mathbf{x}) \quad (3.19)$$

$$= \sum_{i=1}^m \lambda_i^k \langle u_t, \phi_i \rangle \psi_i(\mathbf{x}) + \mathcal{T}_{\text{fast}}(k\tau) \circ u_t(\mathbf{x}) \quad (3.20)$$

$$= \sum_{i=1}^m \lambda_i^k \langle u_t, \psi_i \rangle_\mu \psi_i(\mathbf{x}) + \mathcal{T}_{\text{fast}}(k\tau) \circ u_t(\mathbf{x}). \quad (3.21)$$

Here,  $\mathcal{T}_{\text{slow}}$  is the *dominant*, or slowly-decaying part consisting of the  $m$  slowest processes with  $\lambda_i \geq \lambda_m$ , while  $\mathcal{T}_{\text{fast}}$  contains all (infinitely many) fast processes that are usually not of interest and which decay with geometric rate at least as fast as  $|\lambda_{m+1}|^k$ :

$$\frac{\|\mathcal{T}_{\text{fast}}(k\tau) \circ u_t\|_{2,\mu}^2}{\|u_t\|_{2,\mu}^2} \leq |\lambda_{m+1}|^k.$$

This decomposition requires that subspaces  $\mathcal{T}_{\text{slow}}$  and  $\mathcal{T}_{\text{fast}}$  are orthogonal, which is a consequence of detailed balance. This decomposition permits a compelling physical interpretation: The slow dynamics are a superposition of dynamical processes, each of which can be associated to one eigenfunction  $\psi_i$  (or  $\phi_i$ ) and a corresponding eigenvalue  $\lambda_i$  (see Figs. 3.1c–e). These processes decay with increasing time index  $k$ . In the long-time limit where  $k \rightarrow \infty$ , only the first term with  $\lambda_1 = 1$  remains, recovering to the stationary distribution  $\phi_1(\mathbf{x}) = \mu(\mathbf{x})$ . All other terms correspond to processes with eigenvalues  $\lambda_i < 1$  and decay over time, thus the associated eigenfunctions correspond to processes that decay under the action of the dynamics and represent the dynamical rearrangements taking place while the ensemble relaxes towards the equilibrium distribution. The closer  $\lambda_i$  is to 1, the slower the corresponding process decays; conversely, the closer it is to 0, the faster.

Thus the  $\lambda_i$  for  $i = 2, \dots, m$  each correspond to a physical timescale, indicating how quickly the process decays or transports density toward

equilibrium (see Fig. 3.1e):

$$t_i = -\frac{\tau}{\ln \lambda_i}, \quad (3.22)$$

which is often called the  $i$ th implied timescale [8, 42]. Thus, Eq. (3.19) can be rewritten in terms of implied timescales as:

$$u_{t+k\tau}(\mathbf{x}) = 1 + \sum_{i=2}^m \exp\left(-\frac{k\tau}{t_i}\right) \langle u_t, \psi_i \rangle_{\mu} \psi_i(\mathbf{x}) + \mathcal{F}_{\text{fast}}(k\tau) \circ u_t(\mathbf{x}). \quad (3.23)$$

This implies that when there are gaps amongst the first  $m$  eigenvalues, the system has dynamical processes acting simultaneously on different timescales. For example, a system with two-state kinetics would have  $\lambda_1 = 1$ ,  $\lambda_2 \approx 1$  and  $\lambda_3 \ll \lambda_2$  ( $t_3 \ll t_2$ ), while a system with a clear involvement of an additional kinetic intermediate would have  $\lambda_3 \sim \lambda_2$  ( $t_3 \sim t_2$ ).

In Fig. 3.1, the second process,  $\psi_2$ , corresponds to the slow ( $\lambda_2 = 0.9944$ ) exchange between basins  $A + B$  and basins  $C + D$ , as reflected by the opposite signs of the elements of  $\psi_2$  in these regions (Fig. 3.1c). The next-slowest processes are the  $A \leftrightarrow B$  transition and then the  $C \leftrightarrow D$  transition, while the subsequent eigenvalues are clearly separated from the dominant spectrum and correspond to much faster local diffusion processes. The three slowest processes effectively partition the dynamics into four metastable states corresponding to basins  $A$ ,  $B$ ,  $C$  and  $D$ , which are indicated by the different sign structures of the eigenfunctions (Fig. 3.1c). The metastable states can be calculated from the eigenfunction structure, e.g. using the PCCA method [11, 12, 32].

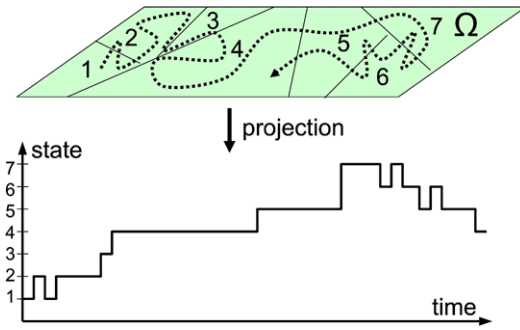
Of special interest is the slowest relaxation time,  $t_2$ . This timescale identifies the *worst case* global equilibration or decorrelation time of the system; no structural observable can relax more slowly than this timescale. Thus, if one desires to calculate an expectation value  $\mathbb{E}(a)$  of an observable  $a(\mathbf{x})$  which has a non-negligible overlap with the second eigenfunction,  $\langle a, \psi_2 \rangle > 0$ , a straightforward single-run MD trajectory would need to be many times  $t_2$  in length in order to compute an unbiased estimate of  $\mathbb{E}(a)$ .

### 3.3 Discretization of State Space

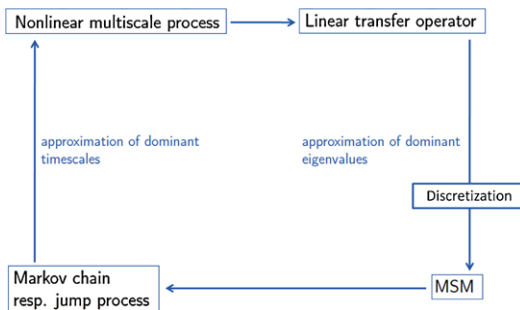
While molecular dynamics in full continuous state space  $\Omega$  is Markovian by construction, the term *Markov State Model* (MSM) or shortly *Markov model* is due to the fact that in practice, state space must be somehow discretized in order to obtain a computationally tractable description of the dynamics as it has first been introduced in [35]. The Markov model then consists of the partitioning of state space used together with the transition matrix modeling the jump process of the observed trajectory projected onto these discrete states. However, this jump process (Fig. 3.2) is no longer Markovian, as the information where the continuous process would be within the local discrete state is lost in the course of discretization. The jump statistics generated by the projection, however, defines a Markov process on the discrete state space associated with the partition. Modeling the long-time statistics of the original process with this discrete state space Markov process is an approximation, i.e., it involves a *discretization error*. In the current section, this discretization error is analyzed and it is shown what needs to be done in order to keep it small.

The discretization error is a *systematic error* of a Markov model since it causes a deterministic deviation of the Markov model dynamics from the true dynamics that persists even when the statistical error is excluded by excessive sampling. In order to focus on this effect alone, it is assumed in this section that the statistical estimation error is zero, i.e., transition probabilities between discrete states can be calculated exactly. The results suggest that the discretization error of a Markov model can be made small enough for the MSM to be useful in accurately describing the relaxation kinetics, even for very large and complex molecular systems. This approach is illustrated in Fig. 3.3.

In practical use, the Markov model is not obtained by actually discretizing the continuous propagator. Rather, one defines a discretization of state space and then estimates the corresponding discretized transfer operator from a finite quantity of simulation data, such as several long or many short MD trajectories that transition between the



**Fig. 3.2** Scheme: The true continuous dynamics (*dashed line*) is projected onto the discrete state space. MSMs approximate the resulting jump process by a Markov jump process



**Fig. 3.3** Illustration of our approach: The continuous dynamics is highly nonlinear and has many scales. It is represented by the linear propagator  $\mathcal{T}$ , whose discretization yields a finite-dimensional transition matrix that represents the Markov State Model (MSM). If the discretization error is small enough, the Markov chain or jump process induced by the MSM is a good approximation of the dominant timescales of the original continuous dynamics

discrete states. The statistical estimation error involved in this estimation will be discussed in the subsequent chapters; the rest of the current chapter focuses only on the approximation error due to discretization of the transfer operator.

Here we consider a discretization of state space  $\Omega$  into  $n$  sets. In practice, this discretization is often a simple partition with sharp boundaries, but in some cases it may be desirable to discretize  $\Omega$  into fuzzy sets [46]. We can describe both cases by defining membership functions  $\chi_i(\mathbf{x})$  that quantify the probability of point  $\mathbf{x}$  to belong to set  $i$  [47] which have the property  $\sum_{i=1}^n \chi_i(\mathbf{x}) = 1$ . We will concentrate on a crisp

partitioning with step functions:

$$\chi_i(\mathbf{x}) = \chi_i^{\text{crisp}}(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in S_i, \\ 0 & \mathbf{x} \notin S_i. \end{cases} \quad (3.24)$$

Here we have used  $n$  sets  $S = \{S_1, \dots, S_n\}$  which entirely partition state space ( $\bigcup_{i=1}^n S_i = \Omega$ ) and have no overlap ( $S_i \cap S_j = \emptyset$  for all  $i \neq j$ ). A typical example of such a crisp partitioning is a Voronoi tessellation [45], where one defines  $n$  centers  $\bar{\mathbf{x}}_i$ ,  $i = 1 \dots n$ , and set  $S_i$  is the union of all points  $\mathbf{x} \in \Omega$  which are closer to  $\bar{\mathbf{x}}_i$  than to any other center using some distance metric (illustrated in Figs. 3.4b and c). Note that such a discretization may be restricted to some subset of the degrees of freedom, e.g. in MD one often ignores velocities and solvent coordinates when discretizing.

The stationary probability  $\pi_i$  to be in set  $i$  is then given by the full stationary density as:

$$\pi_i = \int_{\mathbf{x} \in S_i} d\mathbf{x} \mu(\mathbf{x}),$$

and the local stationary density  $\mu_i(\mathbf{x})$  restricted to set  $i$  (see Fig. 3.4b) is given by

$$\mu_i(\mathbf{x}) = \begin{cases} \frac{\mu(\mathbf{x})}{\pi_i} & \mathbf{x} \in S_i, \\ 0 & \mathbf{x} \notin S_i. \end{cases} \quad (3.25)$$

These properties are local, i.e. they do not require information about the full state space.

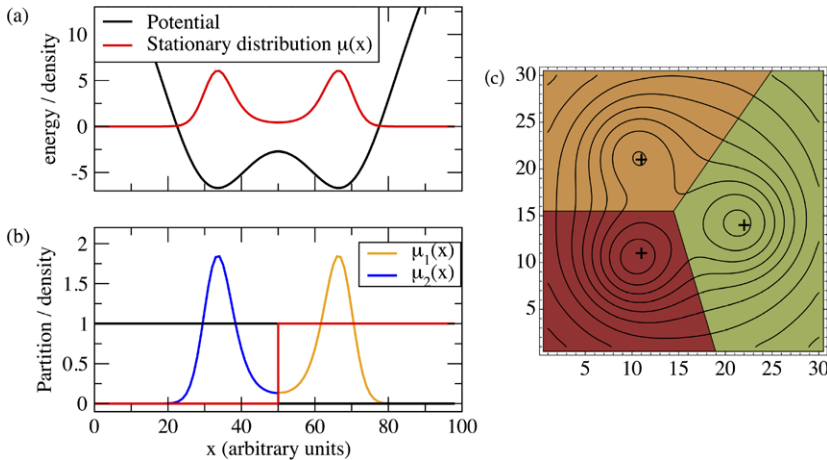
### 3.4 Transition Matrix

Together with the discretization, the Markov model is defined by the row-stochastic transition probability matrix,  $\mathbf{T}(\tau) \in \mathbb{R}^{n \times n}$ , which is the discrete approximation of the transfer operator described in Sect. 3.2 *via*:

$$T_{ij}(\tau) = \frac{\langle \chi_j, (\mathcal{T}(\tau) \circ \chi_i) \rangle_{\mu}}{\langle \chi_i, \chi_i \rangle_{\mu}} \quad (3.26)$$

Physically, each element  $T_{ij}(\tau)$  represents the time-stationary probability to find the system in state  $j$  at time  $t + \tau$  given that it was in state  $i$  at time  $t$ . By definition of the conditional probability, this is equal to:





**Fig. 3.4** Crisp state space discretization illustrated on a one-dimensional two-well and a two-dimensional three-well potential. **(a)** Two-well potential (black) and stationary distribution  $\mu(\mathbf{x})$  (red). **(b)** Characteristic functions  $v_1(\mathbf{x}) = \chi_1(\mathbf{x})$ ,  $v_2(\mathbf{x}) = \chi_2(\mathbf{x})$  (black and red). This dis-

cretization has the corresponding local densities  $\mu_1(\mathbf{x})$ ,  $\mu_2(\mathbf{x})$  (blue and yellow), see Eq. (3.25). **(c)** Three-well potential (black contours indicate the isopotential lines) with a crisp partitioning into three states using a Voronoi partition with the centers denoted (+)

$$T_{ij}(\tau) = \mathbb{P}[\mathbf{x}(t + \tau) \in S_j \mid \mathbf{x}(t) \in S_i] \quad (3.27)$$

$$= \frac{\mathbb{P}[\mathbf{x}(t + \tau) \in S_j \cap \mathbf{x}(t) \in S_i]}{\mathbb{P}[\mathbf{x}(t) \in S_i]} \quad (3.28)$$

$$= \frac{\int_{S_i} d\mathbf{x} \mu_i(\mathbf{x}) p(\mathbf{x}, S_j; \tau)}{\int_{\mathbf{x}} d\mathbf{x} \mu_i(\mathbf{x})}, \quad (3.29)$$

where we have used Eq. (3.2). Note that in this case the integrals run over individual sets and only need the local equilibrium distributions  $\mu_i(\mathbf{x})$  as weights. This is a very powerful feature: In order to estimate transition probabilities, we do not need any information about the global equilibrium distribution of the system, and the dynamical information needed extends only over time  $\tau$ . In principle, the full dynamical information of the discretized system can be obtained by initiating trajectories of length  $\tau$  out of each state  $i$  as long as we draw the starting points of these simulations from a local equilibrium density  $\mu_i(\mathbf{x})$  [24, 37, 47].

The transition matrix can also be written in terms of correlation functions [42]:

$$T_{ij}(\tau) = \frac{\mathbb{E}[\chi_i(\mathbf{x}(t))\chi_j(\mathbf{x}(t + \tau))]}{\mathbb{E}[\chi_i(\mathbf{x}(t))]} = \frac{c_{ij}^{\text{corr}}(\tau)}{\pi_i}, \quad (3.30)$$

where the unconditional transition probability  $c_{ij}^{\text{corr}}(\tau) = \pi_i T_{ij}(\tau)$  is an equilibrium time correlation function which is normalized such that  $\sum_{i,j} c_{ij}^{\text{corr}}(\tau) = 1$ . For dynamics fulfilling detailed balance, the correlation matrix is symmetric ( $c_{ij}^{\text{corr}}(\tau) = c_{ji}^{\text{corr}}(\tau)$ ).

Since the transition matrix  $\mathbf{T}(\tau)$  is a discretization of the transfer operator  $\mathcal{T}$  [35, 36, 36, 37] (Sect. 3.2), we can relate the functions that are transported by  $\mathcal{T}$  (functions  $u_t$  in Eq. (3.8)) to column vectors that are multiplied to the matrix from the right while the probability densities  $p_t$  (Eq. (3.10)) correspond to row vectors that are multiplied to the matrix from the left. Suppose that  $\mathbf{p}(t) \in \mathbb{R}^n$  is a column vector whose elements denote the probability, or population, to be within any set  $j \in \{1, \dots, n\}$  at time  $t$ . After time  $\tau$ , the probabilities will have changed according to:

$$p_j(t + \tau) = \sum_{i=1}^n p_i(t) T_{ij}(\tau), \quad (3.31)$$

or in matrix form:

$$\mathbf{p}^T(t + \tau) = \mathbf{p}^T(t) \mathbf{T}(\tau) \quad (3.32)$$

Note that an alternative convention often used in the literature is to write  $\mathbf{T}(\tau)$  as a column-stochastic matrix, obtained by taking the trans-

pose of the row-stochastic transition matrix defined here.

The stationary probabilities of discrete states,  $\pi_i$ , yield the unique discrete stationary distribution of  $\mathbf{T}$ :

$$\boldsymbol{\pi}^T = \boldsymbol{\pi}^T \mathbf{T}(\tau). \quad (3.33)$$

All equations encountered so far are concerned with the discrete state space given by the partition sets, i.e.,  $\mathbf{p}^T(t)$  and  $\mathbf{p}^T(t + \tau)$  are probability distribution on the discrete state space. The probability distribution on the continuous state space related to  $\mathbf{p}^T(t)$  is

$$u_t(\mathbf{x}) = \sum_i p_i(t) \chi_i(\mathbf{x}).$$

If we propagate  $u_t$  with the true dynamics for time  $\tau$ , we get  $u_{t+\tau} = \mathcal{S}(\tau) \circ u_t$ . However,  $u_{t+\tau}$  and  $\mathbf{p}^T(t + \tau)$  will no longer be perfectly related as above, i.e., we will only have

$$u_{t+\tau}(\mathbf{x}) \approx \sum_i p_i(t + \tau) \chi_i(\mathbf{x}).$$

We wish now to understand the error involved with this approximation. Moreover, we wish to model the system kinetics on long timescales by *approximating* the true dynamics with a Markov chain on the discrete state space of  $n$  states. Using  $\mathbf{T}(\tau)$  as a Markov model *predicts* that for later times,  $t + k\tau$ , the probability distribution will evolve as:

$$\mathbf{p}^T(t + k\tau) = \mathbf{p}^T(t) \mathbf{T}^k(\tau), \quad (3.34)$$

on the discrete state space which can only approximate the true distribution,

$$u_{t+k\tau} = (\mathcal{S}(\tau))^k \circ u_t,$$

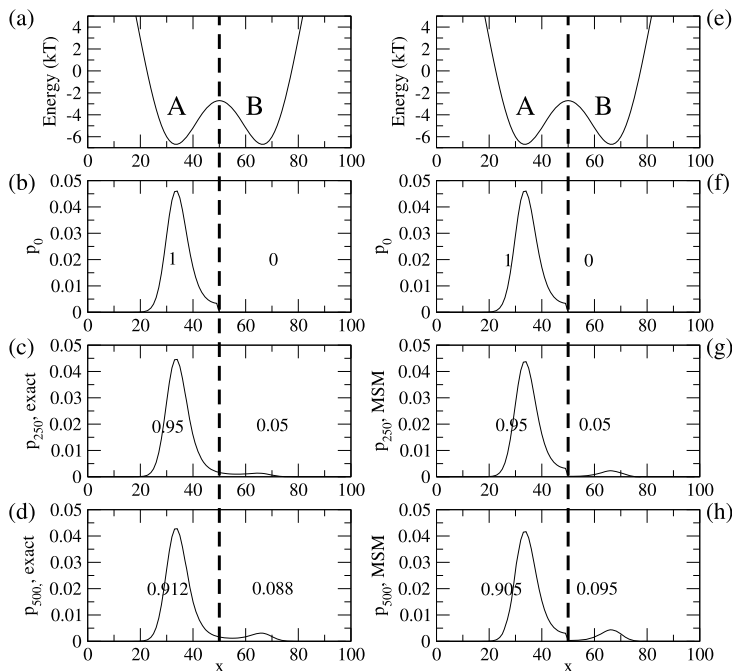
that would have been produced by the continuous transfer operator, as Eq. (3.34) is the result of a state space discretization. The discretization error involved in this approximation thus depends on how this discretization is chosen and is analyzed in detail below. A description alternative to that of transition matrices quite common in chemical physics is using rate matrices and Master equations [5, 24, 26, 40, 41, 48].

### 3.5 Discretization Error and Non-Markovianity

The Markov model  $\mathbf{T}(\tau)$  is indeed a model in the sense that it only approximates the long-time dynamics based on a discretization of state space, making the dynamical equation (3.34) approximate. Here we analyze the approximation quality of Markov models in detail and obtain a description that reveals which properties the state space discretization and the lag time  $\tau$  must fulfill in order to obtain a good model.

Previous works have mainly discussed the quality of a Markov model in terms of its ‘‘Markovianity’’ and introduced tests of Markovianity of the process  $\mathbf{x}(t)$  projected onto the discrete state space. The space-continuous dynamics  $\mathbf{x}(t)$  described in Sect. 3.1 is, by definition, Markovian in full state space  $\Omega$  and it can thus be exactly described by a linear operator, such as the transfer operator  $\mathcal{S}(\tau)$  defined in Eq. (3.8). The problem lies in the fact that by performing a state space discretization, continuous states  $\mathbf{x} \in \Omega$  are grouped into discrete states  $s_i$  (Sect. 3.3), thus ‘‘erasing’’ information of the exact location within these states and ‘‘projecting’’ a continuous trajectory  $\mathbf{x}(t)$  onto a discrete trajectory  $s(t) = s(\mathbf{x}(t))$ . This jump process,  $s(t)$ , is *not Markovian*, but Markov models attempt to approximate  $s(t)$  with a Markov chain.

Thus, non-Markovianity occurs when the full state space resolution is reduced by mapping the continuous dynamics onto a reduced space. In Markov models of molecular dynamics, this reduction consists usually of both neglect of degrees of freedom and discretization of the resolved degrees of freedom. Markov models typically only use atom positions, thus the velocities are projected out [9, 32]. So far, Markov models have also neglected solvent degrees of freedom and have only used the solute coordinates [9, 33], and the effect of this was studied in detail in [23]. Indeed, it may be necessary to incorporate solvent coordinates in situations where the solvent molecules are involved in slow processes that are not easily detected in the solute coordinates [25]. Often, Markov models are also based on distance metrics that only involve a subset of the solute



**Fig. 3.5** Illustration of the discretization error by comparing the dynamics of the diffusion in a double well potential (a, e) (see Supplementary Information for setup) at time steps 0 (b), 250 (c), 500 (d) with the predictions of a Markov model parametrized with a too short lag time  $\tau = 250$  at the corresponding times 0 (f), 250 (g), 500 (h). (b, c, d) show the true density  $p_t(\mathbf{x})$  (solid black line) and the probabilities associated with

the two discrete states left and right of the dashed line. The numbers in (f, g, h) are the discrete state probabilities  $p_i(t + k\tau)$  predicted by the Markov model while the solid black line shows the hypothetical density  $p_i(t + k\tau)\mu_i(\mathbf{x})$  that inherently assumed by the Markov model by using the discrete state probabilities to correspondingly weight the local stationary densities

atoms, such as RMSD between heavy atom or alpha carbon coordinates [4, 9, 33], or backbone dihedral angles [5, 32]. Possibly the strongest approximation is caused by clustering or lumping sets of coordinates in the selected coordinate subspace into discrete states [4, 5, 9, 26, 33]. Formally, all of these operations aggregate sets of points in continuous state space  $\Omega$  into discrete states, and the question to be addressed is what is the magnitude of the discretization error caused by treating the non-Markovian jump process between these sets as a Markov chain.

Consider the diffusive dynamics model depicted in Fig. 3.5a and let us follow the evolution of the dynamics given that we start from a local equilibrium in basin A (Fig. 3.5b), either with the exact dynamics, or with the Markov model dynamics on the discrete state space A

and B. After a step  $\tau$ , both dynamics have transported a fraction of 0.1 of the ensemble to B. The true dynamics resolves the fact that much of this is still located near the saddle point (Fig. 3.5c). The Markov model cannot resolve local densities within its discrete states, which is equivalent to assuming that for the next step the ensemble has already equilibrated within the discrete state (Fig. 3.5g). This difference affects the discrete state (basin) probabilities at time  $2\tau$ : In the true dynamics, a significant part of the 0.1 fraction can cross back to A as it is still near the saddle point, while this is not the case in the Markov model where the 0.1 fraction is assumed to be relaxed to states mostly around the minimum (Compare Figs. 3.5d and h). As a result, the probability to be in state B is higher in the Markov model than in the true dynamics. The difference between the

Markov model dynamics and the true dynamics is thus a result of discretization, because the discretized model can no longer resolve deviations from local equilibrium density  $\mu_i(\mathbf{x})$  within the discrete state.

This picture suggests the discretization error to have two properties: (a) the finer the discretization, the smaller the discretization error is, and (b) when increasing the coarse-graining time, or time resolution, of our model,  $\tau$ , the errors for any fixed point in time  $t$  should diminish, because we have less often made the approximation of imposing local equilibrium.

### 3.6 Quantifying the Discretization Error

In order to quantify the discretization error, we will exploit the fact that the construction of a Markov State Model can be related to a projection of the transfer operator  $\mathcal{T}(\tau)$ . This projection, denoted  $Q$ , is the orthogonal projection with respect to the scalar product  $\langle \cdot, \cdot \rangle_\mu$  onto a finite dimensional space  $D$  spanned by a given basis  $q_1, \dots, q_n$ , e.g., for  $q_i = \chi_i$  being the characteristic functions from (3.24) that are associated with a crisp partitioning of the state space. For a general function  $u$ ,  $Qu$  is the best possible representation of  $u$  in the space  $D$ . In general, it can be calculated [32, 37, 40] that the *projected propagation operator*, that is, the best representation of the propagator  $\mathcal{T}$  in our space  $D$ , has the form  $Q\mathcal{T}(\tau)Q$ . It can be represented by the matrix  $\mathbf{T}(\tau) = T(\tau)M^{-1}$  with

$$\begin{aligned} T_{ij}(\tau) &= \frac{\langle q_j, (\mathcal{T}(\tau) \circ q_i) \rangle_\mu}{\langle q_i, q_i \rangle_\mu}, \\ M_{ij} &= \frac{\langle q_j, q_i \rangle_\mu}{\langle q_i, q_i \rangle_\mu}. \end{aligned} \quad (3.35)$$

If we choose  $q_i = \chi_i$  being the characteristic functions from (3.24) that are associated with a crisp partitioning of the state space into sets  $S_1, \dots, S_n$ , we find  $M = Id$  because of orthogonality of the characteristic functions. Moreover, in this case, as calculated in (3.29)

$$\mathbf{T}(\tau)_{ij} = T_{ij}(\tau) = \mathbb{P}[\mathbf{x}(t + \tau) \in S_j \mid \mathbf{x}(t) \in S_i].$$

This means that this MSM transition matrix can be interpreted as the projection of the transfer operator with  $Q$  being the projection onto the discretization basis. Together with Fig. 3.5 this suggests a practical measure to quantify the discretization error. Densities, eigenfunctions or any other function  $f(\mathbf{x})$  of the continuous state  $\mathbf{x}$ , are approximated by its best-approximations  $\hat{f}(\mathbf{x}) = Qf(\mathbf{x})$  within the space spanned by the discretization basis  $q_1, \dots, q_n$ . In the case of a crisp partitioning of state space, functions  $f(\mathbf{x})$  are approximated through the discretization  $S_1, \dots, S_n$  by step functions  $\hat{f}(\mathbf{x})$  that are constant within the discrete states:

$$\hat{f}(\mathbf{x}) = Qf(\mathbf{x}) = \sum_{i=1}^n a_i \chi_i(\mathbf{x}) \quad (3.36)$$

where the coefficients are given by the projection weighted by the probability of each state:

$$a_i = \frac{\langle f, \chi_i \rangle_\mu}{\langle \mathbf{1}, \chi_i \rangle_\mu} = \frac{\int_{S_i} d\mathbf{x} \mu(\mathbf{x}) f(\mathbf{x})}{\int_{S_i} d\mathbf{x} \mu(\mathbf{x})}. \quad (3.37)$$

The approximation error that is caused by the discretization is then defined as the  $\mu$ -weighted Euclidean norm  $\|\cdot\|_{\mu,2}$  of the difference between discretized and original function:

$$\delta_f \equiv \|f(\mathbf{x}) - \hat{f}(\mathbf{x})\|_{\mu,2} \quad (3.38)$$

$$= \left( \int_{\Omega} d\mathbf{x} \mu(\mathbf{x}) (f(\mathbf{x}) - \hat{f}(\mathbf{x}))^2 \right)^{1/2}. \quad (3.39)$$

The projection allows the comparison between true and Markov model dynamics to be made exactly as suggested by Fig. 3.5: In both cases we start with an arbitrary initial density projected onto discrete states,  $Qp_0(\mathbf{x})$ , then transport this density either with the true or with the Markov model dynamics for some time  $k\tau$ . Subsequently, the densities is again projected onto discrete states by  $Q$  and then compared:

- The true dynamics transports the initial density  $Qp_0(\mathbf{x})$  to  $[\mathcal{T}(\tau)]^k Qp_0(\mathbf{x})$
- The Markov model dynamics transports the initial density  $Qp_0(\mathbf{x})$  to  $Q\mathcal{T}(\tau)Qp_0(\mathbf{x})$  in one step and to  $Q[\mathcal{T}(\tau)Q]^k p_0(\mathbf{x})$  in  $k$  steps using the identity for projections  $Q \circ Q = Q$ .

- Projecting both densities to local densities and comparing yields the difference

$$\varepsilon(k) = \left\| Q[\mathcal{T}(\tau)]^k Q p_0(\mathbf{x}) - Q[\mathcal{T}(\tau)Q]^k p_0(\mathbf{x}) \right\|_{\mu,2} \quad (3.40)$$

$$= \left\| [Q[\mathcal{T}(\tau)]^k Q - Q[\mathcal{T}(\tau)Q]^k] p_0(\mathbf{x}) \right\|_{\mu,2}. \quad (3.41)$$

In order to remove the dependency on the initial distribution  $p_0(\mathbf{x})$ , we assume the worst case: the maximal possible value of  $\varepsilon(k)$  amongst all possible  $p_0(\mathbf{x})$  is given by the operator-2-norm of the error matrix  $[Q[\mathcal{T}(\tau)]^k Q - Q[\mathcal{T}(\tau)Q]^k]$ , where  $\|A\|_{\mu,2} \equiv \max_{\|x\|=1} \|Ax\|_{\mu,2}$  [17], thus the Markov model error is defined as:

$$E(k) := \left\| Q[\mathcal{T}(\tau)]^k Q - Q[\mathcal{T}(\tau)Q]^k \right\|_{\mu,2}, \quad (3.42)$$

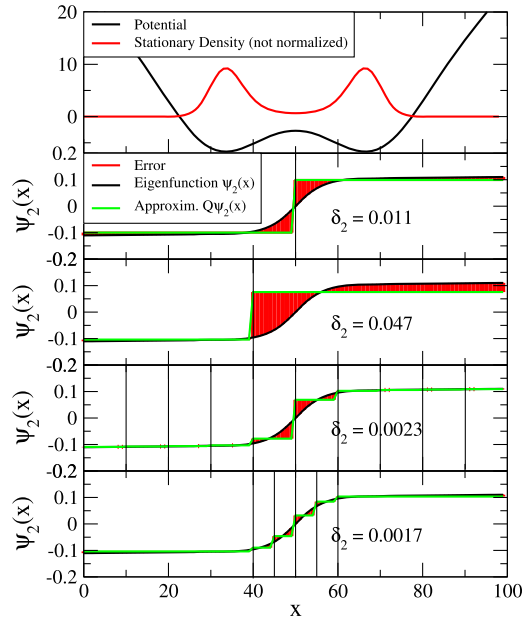
which measures the maximum possible difference between the true probability density at time  $k\tau$  and the probability density predicted by the Markov model at the same time.

In order to quantify this error, we declare our explicit interest in the  $m$  slowest processes with eigenvalues  $1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_m$ . Generally,  $m \leq n$ , i.e. we are interested in less processes than the number of  $n$  eigenvectors that a Markov model with  $n$  states has. We define the following two quantities:

- $\delta_i := \|\psi_i(\mathbf{x}) - Q\psi_i(\mathbf{x})\|_{\mu,2}$  is the *eigenfunction approximation error*, quantifying the error of approximating the true continuous eigenfunctions of the transfer operator,  $\psi_i$  (see Fig. 3.6 for illustration), for all  $i \in \{1, \dots, m\}$ .  $\delta := \max_i \delta_i$  is the largest approximation error amongst these first  $m$  eigenfunctions
- $\eta(\tau) := \frac{\lambda_{m+1}(\tau)}{\lambda_2(\tau)}$  is the *spectral error*, the error due to neglecting the fast subspace of the transfer operator, which decays to 0 with increasing lag time:  $\lim_{\tau \rightarrow \infty} \eta(\tau) = 0$ .

The general statement is that the Markov model error  $E(k)$  can be proven [36] to be bounded from above by the following expression:

$$E(k) \leq \min\{2, [m\delta + \eta(\tau)][a(\delta) + b(\tau)]\} \lambda_2^k \quad (3.43)$$



**Fig. 3.6** Illustration of the eigenfunction approximation error  $\delta_2$  on the slow transition in the diffusion in a double well (top, black line). The slowest eigenfunction is shown in the lower four panels (black), along with the step approximations (green) of the partitions (vertical black lines) at  $x = 50$ ;  $x = 40$ ;  $x = 10, 20, \dots, 80, 90$ ; and  $x = 40, 45, 50, 55, 60$ . The eigenfunction approximation error  $\delta_2$  is shown as red area and its norm is printed

with

$$a(\delta) = \sqrt{m}(k-1)\delta, \quad (3.44)$$

$$b(\tau) = \frac{\eta(\tau)}{1-\eta(\tau)}(1-\eta(\tau)^{k-1}). \quad (3.45)$$

This implies two observations:

1. For long times, the overall error decays to zero with  $\lambda_2^k$ , where  $0 < \lambda_2 < 1$ , thus the stationary distribution (recovered as  $k \rightarrow \infty$ ) is always correctly modeled, even if the kinetics are badly approximated.
2. The error during the kinetically interesting timescales consists of a product whose terms contain separately the discretization error and spectral error. Thus, the overall error can be diminished by choosing a discretization basis  $q_1, \dots, q_n$  that approximates the slow eigenfunctions well, and using a large lag time  $\tau$ . For a crisp partitioning this implies that the

discretization has to be fine enough to trace the slow eigenfunctions well.

Depending on the ratio  $\lambda_{m+1}(\tau)/\lambda_2(\tau)$ , the decay of the spectral error  $\eta(\tau)$  with  $\tau$  might be slow. It is thus interesting to consider a special case of discretization that yields  $n = m$  and  $\delta = 0$ . This would be achieved by a Markov model that uses a fuzzy partition with membership functions  $q_1, \dots, q_n$  derived from the first  $m$  eigenfunctions  $\psi_1, \dots, \psi_m$  [24]. In this case, the space spanned by  $q_1, \dots, q_n$  would equal the dominant eigenspace and hence the projection error would be  $\delta = 0$ . From a more practical point of view, this situation can be approached by using a Markov model with  $n > m$  states located such that they discretize the first  $m$  eigenfunctions with a vanishing discretization error, and then declaring that we are *only* interested in these  $m$  slowest relaxation processes. In this case we do not need to rely on the upper bound of the error from Eq. (3.43), but directly get the important result  $E(k) = 0$ .

In other words, a Markov model can approximate the kinetics of slow processes *arbitrarily well*, provided the discretization can be made sufficiently fine or improved in a way that continues to minimize the eigenfunction approximation error. This observation can be rationalized by Eq. (3.19) which shows that the dynamics of the transfer operator can be exactly decomposed into a superposition of slow and fast processes.

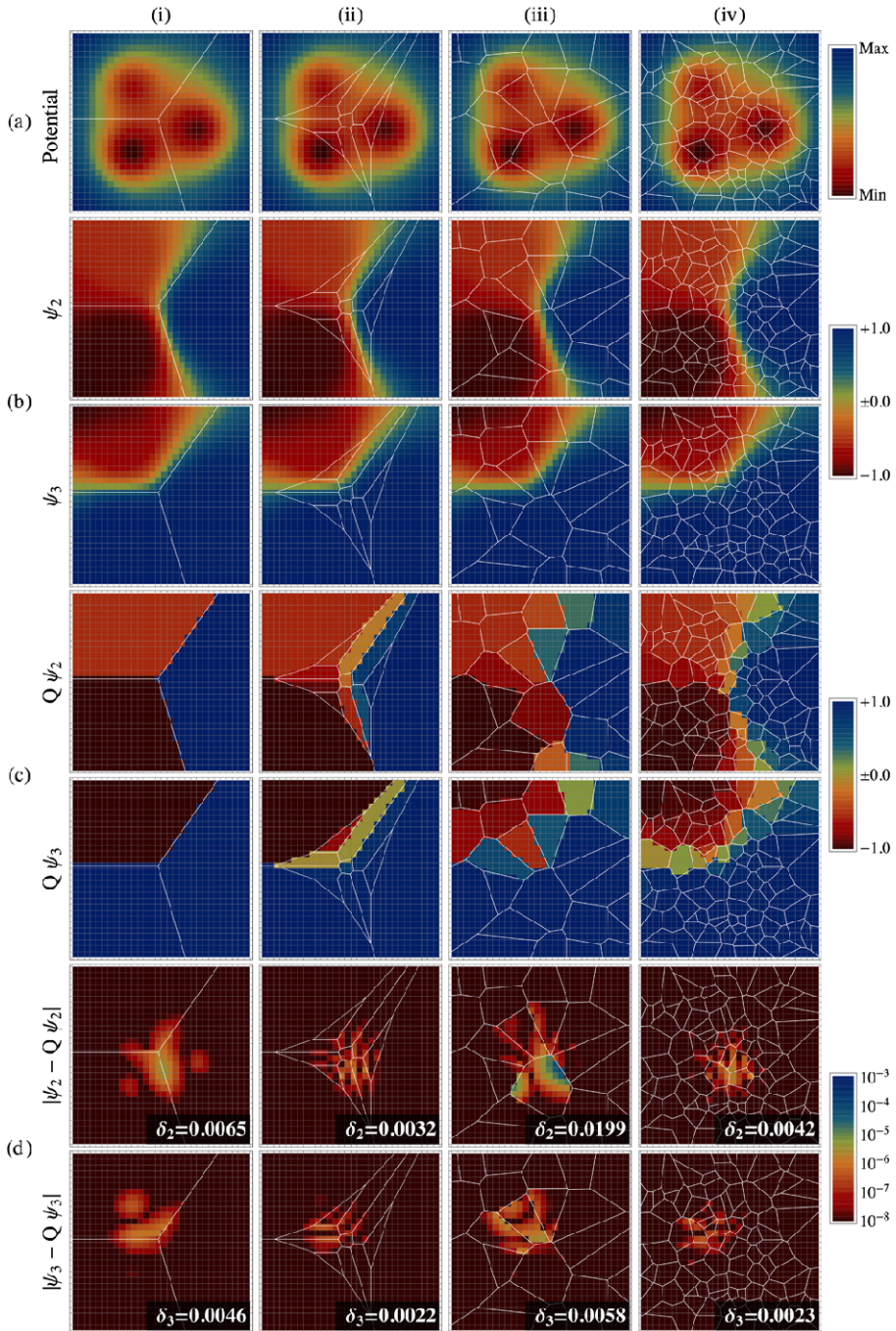
An important consequence of the  $\delta$ -dependence of the error is that the best partition is not necessarily metastable. Previous work [9, 9, 20, 32, 36, 42] has focused on the construction of partitions with high metastability (defined as the trace of the transition matrix  $\mathbf{T}(\tau)$ ), e.g. the two-state partition shown in Fig. 3.6b). This approach was based on the idea that the discretized dynamics must be approximately Markovian if the system remained in each partition sufficiently long to approximately lose memory [9]. While it can be shown that if a system has  $m$  metastable sets with  $\lambda_m \gg \lambda_{m+1}$ , then the most metastable partition into  $n = m$  sets also minimizes the discretization error [36, 36], the expression for the discretization error given here has two further profound ramifications: First, even in the case where there

exists a strong separation of timescales so the system has clearly  $m$  metastable sets, the discretization error can be reduced *even further* by splitting the metastable partition into a total of  $n > m$  sets which are not metastable. And second, even in the *absence* of a strong separation of timescales, the discretization error can be made arbitrarily small by making the partition finer, especially in transition regions, where the eigenfunctions change most rapidly (see Fig. 3.6b).

Figure 3.7 illustrates the Markov model discretization error on a two-dimensional three-well example where two slow processes are of interest. The left panels show a metastable partition with 3 sets. As seen in Fig. 3.7d, the discretization errors  $|\psi_2 - Q\psi_2|(\mathbf{x})$  and  $|\psi_3 - Q\psi_3|(\mathbf{x})$  are large near the transition regions, where the eigenfunctions  $\psi_2(\mathbf{x})$  and  $\psi_3(\mathbf{x})$  change rapidly, leading to a large discretization error. Using a random partition (Fig. 3.7, third column) makes the situation worse, but increasing the number of states reduces the discretization error (Fig. 3.7, fourth column), thereby increasing the quality of the Markov model. When states are chosen such as to well approximate the eigenfunctions, a very small error can be obtained with few sets (Fig. 3.7, second column).

These results suggest that an adaptive discretization algorithm may be constructed which minimizes the  $E(k)$  error. Such an algorithm could iteratively modify the definitions of discretization sets as suggested previously [9], but instead of maximizing metastability it would minimize the  $E(k)$  error which can be evaluated by comparing eigenvector approximations on a coarse discretization compared to a reference evaluated on a finer discretization [36].

One of the most intriguing insights from both Eq. (3.19) and the results of the discretization error is that if, for a given system, only the slowest dynamical processes are of interest, it is sufficient to discretize the state space such that the first few eigenvectors are well represented (in terms of small approximation errors  $\delta_i$ ). For example, if one is interested in processes on timescales  $t^*$  or slower, then the number  $m$  of eigenfunctions that need to be resolved is equal to the number of



**Fig. 3.7** Illustration of the eigenfunction approximation errors  $\delta_2$  and  $\delta_3$  on the two slowest processes in a two-dimensional three-well diffusion model (see Supplementary Information for model details). The *columns* from left to right show different state space discretizations with white lines as state boundaries: (i) 3 states with maximum metastability, (ii) the metastable states were further subdivided manually into 13 states to better resolve the

transition region, resulting in a partition where no individual state is metastable, (iii)/(iv) Voronoi partition using 25/100 randomly chosen centers, respectively. (a) Potential, (b) The exact eigenfunctions of the slow processes,  $\psi_2(\mathbf{x})$  and  $\psi_3(\mathbf{x})$ , (c) The approximation of eigenfunctions with discrete states,  $Q\psi_2(\mathbf{x})$  and  $Q\psi_3(\mathbf{x})$ , (d) Approximation errors  $|\psi_2 - Q\psi_2|(\mathbf{x})$  and  $|\psi_3 - Q\psi_3|(\mathbf{x})$ . The error norms  $\delta_2$  and  $\delta_3$  are given

implied timescales with  $t_i \geq t^*$ . Due to the perfect decoupling of processes for reversible dynamics in the eigenfunctions (see Eqs. (3.20)–(3.21)), no gap after these first  $m$  timescales of interest is needed. Note that the quality of the Markov model does not depend on the dimensionality of the simulated system, i.e. the number of atoms. Thus, if only the slowest process of the system is of interest (such as the folding process in a two-state folder), only a one-dimensional parameter, namely the level of the dominant eigenfunction, needs to be approximated with the clustering, even if the system is huge. This opens a way to discretize state spaces of very large molecular systems.

### 3.7 Approximation of Eigenvalues and Timescales by Markov Models

One of the most interesting kinetic properties of molecular systems are the intrinsic timescales of the system. They can be both experimentally accessed *via* relaxation or correlation functions that are measurable with various spectroscopic techniques [2, 5, 20, 28], but can also be directly calculated from the Markov model eigenvalues as implied timescales, Eq. (3.22). Thus, we investigate the question how well the dominant eigenvalues  $\lambda_i$  are approximated by the Markov model, which immediately results in estimates for how accurately a Markov model may reproduce the implied timescales of the original dynamics. Consider the first  $m$  eigenvalues of  $\mathcal{T}(\tau)$ ,  $1 = \lambda_1(\tau) > \lambda_2(\tau) \geq \dots \geq \lambda_m(\tau)$ , and let  $1 = \hat{\lambda}_1(\tau) > \hat{\lambda}_2(\tau) \geq \dots \geq \hat{\lambda}_m(\tau)$  denote the associated eigenvalues of the Markov model  $\mathbf{T}(\tau)$ . The rigorous mathematical estimate from [13] states that

$$\max_{j=1,\dots,m} |\lambda_j(\tau) - \hat{\lambda}_j(\tau)| \leq (m-1)\lambda_2(\tau)\delta^2, \quad (3.46)$$

where  $\delta$  is again the maximum discretization error of the first  $m$  eigenfunctions, showing that the eigenvalues are well reproduced when the discretization well traces these eigenfunctions. In

particular if we are only interested in the eigenvalue of the slowest process,  $\lambda_2(\tau)$ , which is often experimentally reported *via* the slowest relaxation time of the system,  $t_2$ , the following estimate of the approximation error can be given:

$$\frac{|\lambda_2(\tau) - \hat{\lambda}_2(\tau)|}{|\lambda_2(\tau)|} \leq \delta_2^2. \quad (3.47)$$

As  $\lambda_2(\tau)$  corresponds to a slow process, we can make the restriction  $\lambda_2(\tau) > 0$ . Moreover, the discretization error of Markov models based on full partitions of state space is such that the eigenvalues are always underestimated [13], thus  $\lambda_2(\tau) - \hat{\lambda}_2(\tau) > 0$ . Using Eq. (3.22), we thus obtain the estimate for the discretization error of the largest implied timescale and the corresponding smallest implied rate,  $k_2 = t_2^{-1}$ :

$$\hat{t}_2^{-1} - t_2^{-1} = \hat{k}_2 - k_2 \leq -\tau^{-1} \ln(1 - \delta_2^2), \quad (3.48)$$

which implies that for either  $\delta_2 \rightarrow 0_+$  or  $\tau \rightarrow \infty$ , the error in the largest implied timescale or smallest implied rate tends to zero. Moreover, since  $\lambda_2(\tau) \rightarrow 0$  for  $\tau \rightarrow \infty$ , this is also true for the other processes:

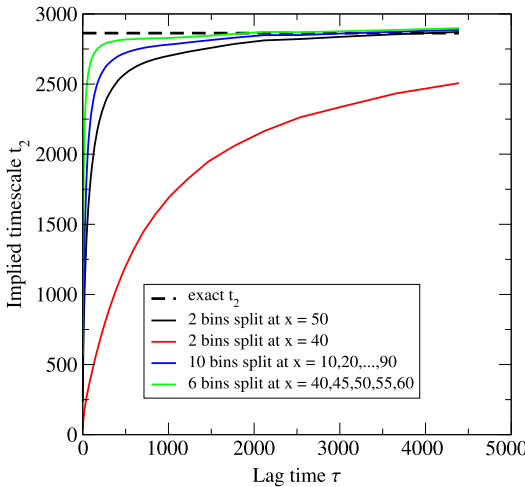
$$\lim_{\tau \rightarrow \infty} \frac{|\lambda_j(\tau) - \hat{\lambda}_j(\tau)|}{|\lambda_j(\tau)|} = 0, \quad (3.49)$$

and also

$$\lim_{\delta \rightarrow 0} \frac{|\lambda_j(\tau) - \hat{\lambda}_j(\tau)|}{|\lambda_j(\tau)|} = 0, \quad (3.50)$$

which means that the error of the implied timescales also vanishes for either sufficiently long lag times  $\tau$  or for sufficiently fine discretization. This fact has been empirically observed in many previous studies [2, 5, 9, 26, 32, 33, 42], but can now be understood in detail in terms of the discretization error. It is worth noting that observing convergence of the slowest implied timescales in  $\tau$  is not a test of Markovianity. While Markovian dynamics implies constancy of implied timescales in  $\tau$  [32, 42], the reverse is not true and would require the eigenvectors to be constant as well. However, observing the lag time-dependence of the implied timescales is a useful approach to choose a lag time  $\tau$  at which  $\mathbf{T}(\tau)$



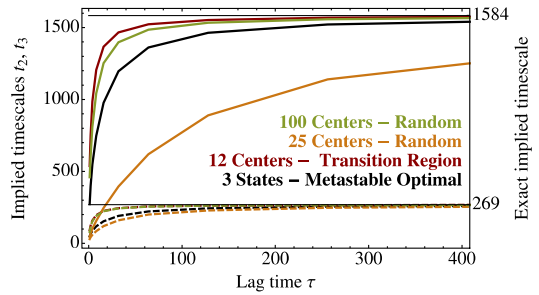


**Fig. 3.8** Convergence of the slowest implied timescale  $t_2 = -\tau/\ln\lambda_2(\tau)$  of the diffusion in a double-well potential depending on the MSM discretization. The metastable partition (*black, solid*) has greater error than non-metastable partitions (*blue, green*) with more states that better trace the change of the slow eigenfunction near the transition state

shall be calculated, but this model needs to be validated subsequently (see Chapter Estimation and Validation).

Figure 3.8 shows the slowest implied timescale  $t_2$  for the diffusion in a two-well potential (see Fig. 3.6) with discretizations shown in Fig. 3.6. The two-state partition at  $x = 50$  requires a lag time of  $\approx 2000$  steps in order to reach an error of  $<3\%$  with respect to the true implied timescale, which is somewhat slower than  $t_2$  itself. When the two-state partition is distorted by shifting the discretization border to  $x = 40$ , this quality is not reached before the process itself has relaxed. Thus, in this system two states are not sufficient to build a Markov model that is at the same time precise and has a time resolution good enough to trace the decay of the slowest process. By using more states and particularly a finer discretization of the transition region, the same approximation quality is obtained with only  $\tau \approx 1500$  (blue) and  $\tau \approx 500$  steps (green).

Figure 3.9 shows the two slowest implied timescales  $t_2, t_3$  for the diffusion in a two-dimensional three-well potential with discretizations shown in Fig. 3.7a. The metastable 3-state partition requires a lag time of  $\approx 1000$  steps in



**Fig. 3.9** Implied timescales for the two slowest processes in the two-dimensional three-well diffusion model (see Fig. 3.7a for potential and Supplementary Information for details). The colors *black, red, yellow, green* correspond to the four choices of discrete states shown in columns 1 to 4 of Fig. 3.7. A fine discretization of the transition region clearly gives the best approximation to the timescales at small lag times

order to reach an error of  $<3\%$  with respect to the true implied timescale, which is somewhat shorter than the slow but longer than the fast timescale, while refining the discretization near the transition states achieves the same precision with  $\tau \approx 200$  using only 12 states. A  $k$ -means clustering with  $k = 25$  is worse than the metastable partition, as some clusters cross over the transition region and fail to resolve the slow eigenfunctions. Increasing the number of clusters to  $k = 100$  improves the result significantly, but is still worse than the 12 states that have been manually chosen so as to well resolve the transition region. This suggests that excellent MSMs could be built with rather few states when an adaptive algorithm that more finely partitions the transition region is employed.

Note that the estimate (3.46) bounds the maximal eigenvalue error for the dominant eigenvalues by the maximal projection error of the dominant eigenfunction. In [34], it is also shown that if  $u$  is a selected, maybe *non*-dominant eigenfunctions for an eigenvalue  $\lambda(\tau)$  and  $\delta = \|u - Qu\|$  is its discretization error, the associated MSM will inherit an eigenvalue  $\hat{\lambda}(\tau)$  with

$$|\lambda(\tau) - \hat{\lambda}(\tau)| \leq 2\lambda_2(\tau)\delta \quad (3.51)$$

if  $\delta = \|u - Qu\| \leq 3/4$ . That is, it is also possible to approximate selected timescales well by choosing the discretization such that it traces the

associated eigenfunctions well without having to take all slower eigenfunctions into account [34].

### 3.8 An Alternative Set-Oriented Projection

In the last sections, we have derived an interpretation of Markov models as projections of the transfer operator  $\mathcal{T}(\tau)$  and connected their discretization error in terms of density propagation and in terms of eigenvalue and timescale approximation to projection errors. Estimates for these quality measures show that the discretization basis should be chosen such that it traces the slow eigenfunctions well. For a crisp partitioning this means that the eigenfunctions should be well approximated by step-functions induced by the sets  $S_1, \dots, S_n$ . On the other hand, the results are not restricted to projections onto step-functions. Theoretically, one could choose an arbitrary discretization basis  $q_1, \dots, q_n$  for constructing the Markov model, see [40], and the corresponding MSM matrix would formally be given by  $\mathbf{T}(\tau) = T(\tau)M^{-1}$  (3.35). In praxis, the basis  $q_1, \dots, q_n$  has been chosen such that it leads to interpretable matrices  $T(\tau)$  and  $M$  in terms of transition probabilities between sets. Otherwise, one will not be able to compute estimates for these matrices and thus for the resulting Markov model. For a crisp partitioning  $S_1, \dots, S_n$  and the associated characteristic functions  $\chi_1, \dots, \chi_n$  we have this property

$$T_{ij}(\tau) = \mathbb{P}[\mathbf{x}(t+\tau) \in S_j \mid \mathbf{x}(t) \in S_i], \quad M = Id.$$

The drawback of this method is that coarse partitionings always lead to coarse step-functions that might not approximate the eigenfunctions well. Therefore, a refinement might be necessary in regions where the slow eigenfunctions are varying strongly.

In this section, we will show how to derive another set-oriented discretization basis where a rather coarse partitioning does not lead to a coarse discretization basis. The main idea goes back to [32, 37, 40]: simply decrease the size of the sets  $S_1, \dots, S_n$  on which the constancy of the discretization basis is enforced. Of course, the resulting sets do not cover the whole state space

and hence do not form a crisp partitioning anymore. We will call such sets *core sets* in the following and denote them by  $C_1, \dots, C_n$  in order to distinguish from a crisp partitioning  $S_1, \dots, S_n$ .

The two questions that we have to answer are (a) how these core sets induce a discretization basis that approximates the slow eigenfunctions well, and (b) how to interpret the transition probabilities between the core sets to calculate the transition matrices  $T(\tau)$  and  $M$  with respect to this basis. The idea is to attach a fuzzy partitioning to the core sets that is connected to the dynamics of the process itself. For every core set  $C_i$  we define the so called *committor function*

$$q_i(x) = \mathbb{P}[x(\sigma) \in C_i \mid x(0) = x], \quad (3.52)$$

where  $\sigma$  is the first time one of the core sets is entered by the process.

That is,  $q_i(x)$  is the probability that the process will hit the set  $C_i$  next rather than the other core sets when being started in point  $x$ . From the definition it follows that [27, 40]

$$\begin{aligned} q_i(x) &= 1 \quad \text{for all } x \in C_i, \\ q_i(x) &= 0 \quad \text{for all } x \in C_j, j \neq i, \\ \sum_i q_i(x) &= 1 \quad \text{for all } x. \end{aligned}$$

The advantage of taking committor functions as discretization basis is that the core sets, on which the committor functions equal to characteristic functions, do not have to cover the whole state space. It is allowed to have a region  $C = \Omega \setminus \bigcup_j C_j$  that is not partitioned and where the values of the committor functions can continuously vary between 0 and 1. This means that the part of state space can be shrunk, where the slow eigenfunctions need to be similar to step-functions. Moreover, it has been shown in [34] that the approximation of the slow eigenfunctions by the committors inside of the fuzzy region  $C$  is accurate if the region  $C$  is left by the process quickly enough. Being more precise, this approximation error is dominated by the ratio of the expected time the process needs to leave the

region  $C$  and the implied timescale that is associated with the target eigenfunction. We will see that it is computationally very suitable that this is the main constraint on  $C$ .

Beside the good approximation of the slow eigenfunctions, the committor discretization has another advantage. We can calculate the Markov model from transition probabilities between the core sets without having to compute the committor functions explicitly. It is shown in [32, 37, 40] that for the committor basis we can interpret the matrices  $T(\tau)$  and  $M$  as follows:

Define  $x^+(t)$  as the index of the core set that is hit next after time  $t$ , and  $x^-(t)$  as the index of the core set that the process has visited last before time  $t$ , then

$$M_{ij} = \mathbb{P}[x^+(t) = j \mid x^-(t) = i],$$

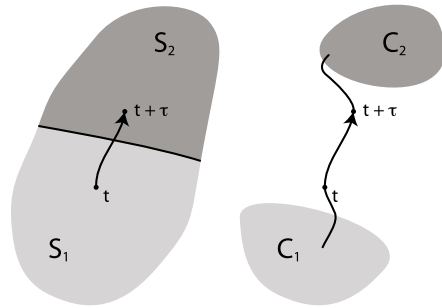
and

$$T_{ij}(\tau) = \mathbb{P}[x^+(t + \tau) = j \mid x^-(t) = i].$$

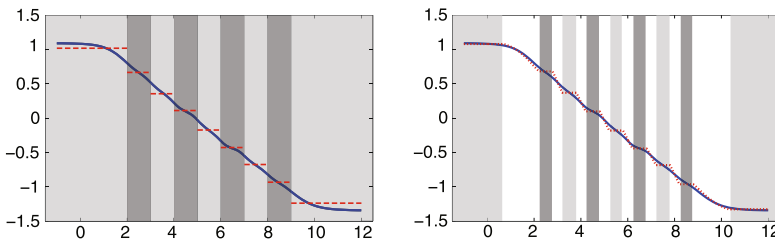
Note that this can be interpreted in terms of a transition behavior. If we interpret  $T_{ij}(\tau)$  as the probability that a transition occurs from index  $i$  at time  $t$  to index  $j$  at time  $t + \tau$ , then we say a transition has occurred if the process was in core set  $i$  at time  $t$  or at least came last from this set and after a time-step  $\tau$  it was in core set  $j$  or at least went next to this set afterwards. Figure 3.10 illustrates this interpretation. As mentioned above, these transition probabilities can be directly estimated from realizations without having to compute the committor functions.

The effect on the approximation of the slowest eigenfunction can be seen in Fig. 3.11.

Computationally, there is an important insight. Assume a crisp partitioning of state space into  $n$  sets  $S_1, \dots, S_n$  is given. Now, a committor discretization would allow to avoid a part of state space from being discretized, as long as the process leaves this part typically much faster than the interesting timescales. On the one hand, these parts of state space exactly correspond to regions where the slow eigenfunctions are varying strongly. So starting from the crisp partitioning we can benefit most by simply ignoring this part of state space and treating the remaining sets as core sets. On the other hand, removing such part of state space, where the process does not spend a lot of time in, does not effect the computational effort in order to generate transitions with respect to the crisp partitioning or the resulting core sets as in Fig. 3.10. Summarizing, one can always enhance a model based on a crisp partitioning by simply declaring a part of state space as not be-



**Fig. 3.10** Counting a transition in the sense of  $T_{12}(\tau)$  for a crisp partitioning (left hand side) and for core sets (right hand side) that do not cover the whole state space



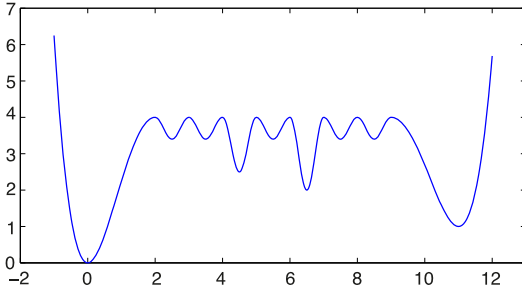
**Fig. 3.11** The benefit of removing a part of state space that is typically left quickly by the process. On the left hand side: step-function approximation of the first non-

trivial eigenfunction. Right hand side: committor approximation of the same eigenfunction

longing to the discretization anymore, as long as this part of state space is usually left quickly by the process, and computationally one can get this enhancement for free.

Let us illustrate this feature by an example. We consider again a diffusion in the potential that is illustrated in Fig. 3.12. Now, Fig. 3.13 shows an optimal crisp partitioning into 9 sets, where every well of the potential falls into another partitioning set. Moreover, it shows the core set discretization where only a part of the transition region between the wells was excluded from the crisp discretization.

The approximation by a step-function is too coarse while removing the transition region and shrinking the size of the sets leads to a smoother



**Fig. 3.12** Example: diffusion in this multi-well potential

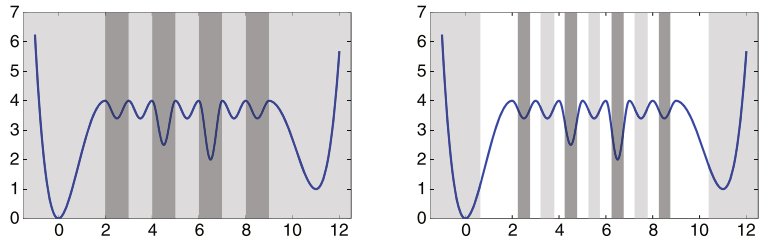
and better interpolation of the eigenfunction. As we discussed in the previous sections, this has a direct impact on the approximation quality of the associated Markov model. For example, the following table shows the implied timescales  $t_i$  of the original Markov process and the approximations by the crisp partitioning and the enhanced core set MSM.

	$t_2$	$t_3$	$t_4$	$t_5$
original	17.5267	3.1701	0.9804	0.4524
core sets	17.3298	3.1332	0.9690	0.4430
crisp partition	16.5478	2.9073	0.8941	0.4006

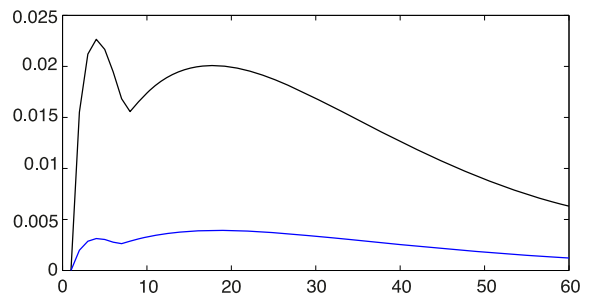
As expected, the approximation quality in terms of timescale approximation could be increased by simply removing a small part of state space from the discretization. The same enhancement is achieved with respect to the density propagation error  $E(k)$  (3.42). Figure 3.14 compares the resulting error for the crisp partitioning (black) and the core set discretization (blue) and increasing  $k$ .

$$E(k) := \|Q[\mathcal{F}(\tau)]^k Q - Q[\mathcal{F}(\tau)Q]^k\|_{\mu,2}.$$

**Fig. 3.13** A metastable crisp partitioning into 9 sets (left), and the derived core set discretization by removing the transition region between the wells (right)



**Fig. 3.14** Density propagation error  $E(k)$  over  $k$ . Black: crisp partitioning. Blue: core sets



## References

1. Andersen HC (1980) Molecular dynamics simulations at constant pressure and/or temperature. *J Chem Phys* 72(4):2384–2393. doi:[10.1063/1.439486](https://doi.org/10.1063/1.439486)
2. Bieri O, Wirz J, Hellrung B, Schutkowski M, Drewello M, Kiefhaber T (1999) The speed limit for protein folding measured by triplet-triplet energy transfer. *Proc Natl Acad Sci USA* 96(17):9597–9601. <http://www.pnas.org/content/96/17/9597.abstract>
3. Bowman GR, Beauchamp KA, Boxer G, Pande VS (2009) Progress and challenges in the automated construction of Markov state models for full protein systems. *J Chem Phys* 131(12):124,101+. doi:[10.1063/1.3216567](https://doi.org/10.1063/1.3216567)
4. Buchete NV, Hummer G (2008) Coarse Master Equations for Peptide Folding Dynamics. *J Phys Chem B* 112:6057–6069
5. Chan CK, Hu Y, Takahashi S, Rousseau DL, Eaton WA, Hofrichter J (1997) Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proc Natl Acad Sci USA* 94(5):1779–1784. <http://www.pnas.org/content/94/5/1779.abstract>
6. Chodera JD, Dill KA, Singhal N, Pande VS, Swope WC, Pitera JW (2007) Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J Chem Phys* 126:155,101
7. Chodera JD, Noé F (2010) Probability distributions of molecular observables computed from Markov models, II: uncertainties in observables and their time-evolution. *J Chem Phys* 133:105,102
8. Chodera JD, Swope WC, Noé F, Prinz JH, Pande VS (2010) Dynamical reweighting: improved estimates of dynamical properties from simulations at multiple temperatures. *J Phys Chem*. doi:[10.1063/1.3592152](https://doi.org/10.1063/1.3592152)
9. Chodera JD, Swope WC, Pitera JW, Dill KA (2006) Long-time protein folding dynamics from short-time molecular dynamics simulations. *Multiscale Model Simul* 5:1214–1226
10. Cooke B, Schmidler SC (2008) Preserving the Boltzmann ensemble in replica-exchange molecular dynamics. *J Chem Phys* 129:164,112
11. Deuffhard P, Huisinga W, Fischer A, Schütte C (2000) Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Linear Algebra Appl* 315:39–59
12. Deuffhard P, Weber M (2005) Robust Perron cluster analysis in conformation dynamics. *Linear Algebra Appl* 398:161–184
13. Djurdjevic N, Sarich M, Schütte C (2012) Estimating the eigenvalue error of Markov state models. *Multiscale Model Simul* 10(1):61–81
14. Duane S (1987) Hybrid Monte Carlo. *Phys Lett B* 195(2):216–222. doi:[10.1016/0370-2693\(87\)91197-X](https://doi.org/10.1016/0370-2693(87)91197-X)
15. Ermak DL (1975) A computer simulation of charged particles in solution, I: technique and equilibrium properties. *J Chem Phys* 62(10):4189–4196
16. Ermak DL, Yeh Y (1974) Equilibrium electrostatic effects on the behavior of polyions in solution: polyion-mobile ion interaction. *Chem Phys Lett* 24(2):243–248
17. Golub GH, van Loan CF (1996) *Matrix computations*, 3rd edn. John Hopkins University Press, Baltimore
18. Herau R, Hitrik M, Sjostrand J (2010) Tunnel effect and symmetries for Kramers Fokker-Planck type operators. [arXiv:1007.0838v1](https://arxiv.org/abs/1007.0838v1)
19. Huisinga W (2001) *Metastability of Markovian systems: a transfer operator based approach in application to molecular dynamics*. PhD thesis, Fachbereich Mathematik und Informatik, FU Berlin
20. Huisinga W, Meyn S, Schütte C (2004) Phase transitions and metastability for Markovian and molecular systems. *Ann Appl Probab* 14:419–458
21. Jäger M, Nguyen H, Crane JC, Kelly JW, Gruebele M (2001) The folding mechanism of a beta-sheet: the WW domain. *J Mol Biol* 311(2):373–393. doi:[10.1006/jmbi.2001.4873](https://doi.org/10.1006/jmbi.2001.4873)
22. van Kampen NG (2006) *Stochastic processes in physics and chemistry*, 4th edn. Elsevier, Amsterdam
23. Keller B, Hünenberger P, van Gunsteren W (2010) An analysis of the validity of Markov state models for emulating the dynamics of classical molecular systems and ensembles. *J Chem Theor Comput*. doi:[10.1021/ct200069c](https://doi.org/10.1021/ct200069c)
24. Kube S, Weber M (2007) A coarse graining method for the identification of transition rates between molecular conformations. *J Chem Phys* 126:024,103–024,113
25. Meerbach E, Schütte C, Horenko I, Schmidt B (2007) Metastable conformational structure and dynamics: peptides between gas phase and aqueous solution. *Series in chemical physics*, vol 87. Springer, Berlin, pp 796–806
26. Metzner P, Horenko I, Schütte C (2007) Generator estimation of Markov jump processes based on incomplete observations nonequidistant in time. *Phys Rev E* 76(6):066,702+. doi:[10.1103/PhysRevE.76.066702](https://doi.org/10.1103/PhysRevE.76.066702)
27. Metzner P, Schütte C, Vanden-Eijnden E (2009) Transition path theory for Markov jump processes. *Multiscale Model Simul* 7(3):1192–1219
28. Neuweiler H, Doose S, Sauer M (2005) A microscopic view of miniprotein folding: enhanced folding efficiency through formation of an intermediate. *Proc Natl Acad Sci USA* 102(46):16,650–16,655. doi:[10.1073/pnas.0507351102](https://doi.org/10.1073/pnas.0507351102)
29. Noé F, Horenko I, Schütte C, Smith JC (2007) Hierarchical analysis of conformational dynamics in biomolecules: transition networks of metastable states. *J Chem Phys* 126:155,102
30. Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR (2009) Constructing the full ensemble of folding pathways from short off-equilibrium simulations. *Proc Natl Acad Sci USA* 106:19,011–19,016
31. Prinz JH et al (2011) Markov models of molecular kinetics: generation and validation. *J Chem Phys* 134:174,105

32. Sarich M (2011) Projected transfer operators. PhD thesis, Freie Universität Berlin
33. Sarich M, Noé F, Schütte C (2010) On the approximation error of Markov state models. *Multiscale Model Simul* 8:1154–1177
34. Sarich M, Schütte C (2012) Approximating selected non-dominant timescales by Markov state models. *Comm Math Sci* (in press)
35. Schütte C (1998) Conformational dynamics: modelling, theory, algorithm, and applications to biomolecules. Habilitation thesis, Fachbereich Mathematik und Informatik, FU Berlin
36. Schütte C, Huisinga W (2003) Biomolecular conformations can be identified as metastable sets of molecular dynamics. In: *Handbook of numerical analysis*. Elsevier, Amsterdam, pp 699–744
37. Schütte C, Sarich M (2013) Metastability and Markov state models in molecular dynamics: modeling, analysis, algorithmic approaches. *Courant lecture notes*, vol 24. American Mathematical Society, Providence
38. Schütte C, Fischer A, Huisinga W, Deuffhard P (1999) A direct approach to conformational dynamics based on hybrid Monte Carlo. *J Comput Phys* 151:146–168
39. Schütte C, Noé F, Meerbach E, Metzner P, Hartmann C (2009) Conformation dynamics. In: Jeltsch RGW (ed) *Proceedings of the international congress on industrial and applied mathematics (ICIAM)*. EMS Publishing House, New York, pp 297–336
40. Schütte C, Noé F, Lu J, Sarich M, Vanden-Eijnden E (2011) Markov state models based on milestoning. *J Chem Phys* 134(19)
41. Sriraman S, Kevrekidis IG, Hummer G (2005) Coarse master equation from Bayesian analysis of replica molecular dynamics simulations. *J Phys Chem B* 109:6479–6484
42. Swope WC, Andersen HC, Berens PH, Wilson KR (1982) A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: application to small water clusters. *J Chem Phys* 76(1):637–649
43. Swope WC, Pitera JW, Suits F (2004) Describing protein folding kinetics by molecular dynamics simulations, 1: theory. *J Phys Chem B* 108:6571–6581
44. Swope WC, Pitera JW, Suits F, Pitman M, Eleftheriou M (2004) Describing protein folding kinetics by molecular dynamics simulations, 2: example applications to alanine dipeptide and beta-hairpin peptide. *J Phys Chem B* 108:6582–6594
45. Voronoi MG (1908) Nouvelles applications des parametres continus a la theorie des formes quadratiques. *J Reine Angew Math* 134:198–287
46. Weber M (2006) Meshless methods in conformation dynamics. PhD thesis, Freie Universität Berlin
47. Weber M (2010) A subspace approach to molecular Markov state models via an infinitesimal generator. ZIB report 09-27-rev
48. Weber M (2011) A subspace approach to molecular Markov state models via a new infinitesimal generator. Habilitation thesis, Fachbereich Mathematik und Informatik, Freie Universität Berlin