Spectral rate theory for two-state kinetics

Jan-Hendrik Prinz, John D. Chodera, and Frank Noé

1 DFG Research Center Matheon, Free University Berlin, Arnimallee 6, 14195 Berlin, Germany
2 Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA
3 Corresponding author

Classical rate theories fail in cases where the observable(s) or order parameter(s) used are poor reaction coordinates or the observed signal is deteriorated by noise, such that no clear separation between reactants and products is possible. Here, we present a general rate theory for ergodic dynamical systems in thermal equilibrium that explicitly takes into account how the system is observed. The theory allows the systematic estimation errors made by standard rate theories to be understood and quantified. We also elucidate the connection of spectral rate theory with the popular Markov state modeling (MSM) approach for molecular simulation studies. An optimal rate estimator is formulated that gives robust and unbiased results even for poor reaction coordinates and can be applied to both computer simulations and single-molecule experiments. No definition of a dividing surface is required, and a measure of the reaction coordinate quality (RCQ) becomes readily available. Additionally, the respective partial probability distributions can be obtained for the reactant and product states along the observed order parameter, even when these strongly overlap. The approach is demonstrated on numerical examples and experimental single-molecule force probe data of the p5ab RNA hairpin and the apo-myoglobin at low pH protein, here focusing on the case of two-state kinetics.

The description of complex molecular motion through simple kinetic rate theories has been a central concern of statistical physics. A common approach, first-order rate theory, treats the relaxation kinetics among distinct regions of configuration space by single-exponential relaxation. There has been a recent interest in estimating such rates from fluctuation trajectories of single molecules, resulting from the recent maturation of measurement techniques that are able to collect extensive traces of single molecule extensions or fluorescence events. When the available observable is a good reaction coordinate in that it allows the slowly-converting states to be clearly separated (see Fig. 2 top left), classical rate theories apply and the robust estimation of transition rates is straightforward using a variety of means. However, in the case in which the slowly-converting states overlap in the observed signal (see Fig. 2 bottom left), either due to the fact that the molecular order parameter used is a poorly separating them, or due to large noise of the measurement (see discussion in [4]), a satisfactory theoretical description is missing and many estimators break down.

Most rate theories and estimators are based on dividing the observed coordinate into a reactant and a product substate and then in some way counting transition events that cross the dividing surface. Transition state theory measures the instantaneous flux across this surface, which is known to overestimate the rate due to the counting of unproductive recrossings over the dividing surface on short timescales. Reactive flux theory has proposed to cope with this by counting a transition event only if it has succeeded to stay on the product side after a sufficiently long lag time \( \tau \). Reactive flux theory involves derivatives of autocorrelation functions that are numerically unreliable to evaluate. In practice, one therefore typically estimates the relaxation rate \( \tau \) via integration or performing an exponential fit to the tail of a suitable correlation function, such as the number correlation function of reactants or the autocorrelation function of the experimentally measured signal. In order to split this relaxation rate into a forward and backward rate constant, a clear definition of the reactant and product substates is needed, which is difficult to achieve when these substates overlap in the observed signal.

Markov models (MSMs) have recently become a popular approach to producing a simplified statistical model of complex molecular dynamics from molecular simulations and can be regarded as an attempt towards a multistate rate theory. MSMs use a transition matrix describing the probability a system initially found in a substate \( i \) is found in substate \( j \) at a lag time \( \tau \) later. When the state division allows the metastable states of the system to be distinguished, the transition matrix with a sufficiently large choice of \( \tau \) can be used to derive a phenomenological transition rate matrix that accurately describes the interstate dynamics. This is explicitly done for the two-state case in [8]. It has been shown in [14, 16] that by increasing the number of substates used to partition state space, and hence using multiple dividing surfaces instead of a single one,
these rate estimates become more precise. In the limit of infinitely many substates, the eigenfunctions of the dynamical propagator in full phase space are exactly recovered, and the rate estimates become exact even for $\tau \to 0^+$ [17]. In practice, however, a finite choice of $\tau$ is necessary in order to have a small systematic estimation error, especially if “uninteresting” degrees of freedom such as momenta or solvent coordinates are discarded. An alternative way of estimating transition rates is by using a state definition that is incomplete and treats the transition region implicitly via committor functions that may better approximate the eigenfunctions of the dynamical propagator in this region [18–20].

The quality of the rate estimates in all of the above approaches relies on the ability to separate the slowly-converting states in terms of some dividing surface or state definition. These approaches break down in the case where the available observables do not permit such a separation, i.e., when kinetically distinct states overlap in the histogram of the observed quantity. However, such a scenario may often arise in single-molecule experiments where the available order parameter depends on what is experimentally observable and may not necessarily be a good indicator of the slow kinetics. Moreover, consequences of the measurement process may increase the overlap between states, for example by bead diffusion in optical tweezer experiments or by shot noise in single-molecule fluorescence measurements. In favorable situations, the signal quality can be improved by binning the data to a coarser timescale, thus reducing the fluctuations from fast processes and shot noise. However, the usefulness of such filtering is limited because the time window used needs to be much shorter than the timescales of interest — otherwise the kinetics will be distorted. In general, one has to deal with a situation where overlap between the slowly-converting states is present, both theoretically and practically.

Hidden Markov Models (HMMs) [21–23] and related likelihood methods [24] are able to estimate transition rates even in such situations, and have been recently successful in distinguishing overlapping states in molecules with complex kinetics [25, 26]. However, HMMs need a probability model of the measurement process to be defined, which can lead to biased estimates when this model is not adequate for the data analyzed. A recent approach, the signal pair-correlation analysis (PCA) [27] provides rate estimates without an explicit probability model, and instead requires to define indicator functions on which the measured signal can uniquely be assigned to one of the kinetically separated states. While this is often easier to achieve than finding an appropriate dividing surface, there is a trade-off between only using only data that is clearly resolved to be in one state or the other (thus minimizing the estimation bias), while avoiding discarding too much data (thus minimizing the statistical error). Despite these slight limitations, both HMM’s and PCA are practically very useful to identify and quantify hidden kinetics in the data. Yet, both are algorithmic approaches rather than a rate theory.

The recent success of single-molecule experiments and the various attempts to estimate accurate transition rates when the relevant states overlap in the observed signal emphasizes the need to have a general rate theory for observed dynamics. Here, we make an attempt towards such a general rate theory for stochastic dynamics that are observed on a possibly poor reaction coordinate - either because the probed molecular order parameter is a poor choice, or because the measurement device creates overlap by noise-broadening the signal. Only very basic properties are assumed to hold for the dynamics underlying the observed signal: It is assumed that the dynamical law governing evolution of the system is a time-stationary Markov process in full-dimensional phase space, obeys microscopic detailed balance, and supports a unique stationary distribution — criteria satisfied for many physical systems of interest. These dynamics are then projected onto some observable in which they are no longer Markovian, and optionally subjected to noise — such that the resulting signal may not easily be separable by a simple dividing surface. This framework allows us to (i) evaluate the quality of existing estimators and propose optimal estimators for the slowest relaxation rate, (ii) provide an absolute measure of reaction coordinate quality (RCQ) of the observed signal, and (iii) derive an optimal estimator for the transition rates between the slowly converting states, as well as their stationary probability densities, even if these strongly overlap in the observation.

The present rate theory is exclusively concerned with the systematic error in estimating rates and proposes “optimal” methods that minimize this systematic rate estimation error. Therefore all statements are strictly valid only in the data-rich regime. Explicit treatment of the statistical error in the data-poor regime is beyond the scope of the present work but discussed at the end of the paper.

Full-space dynamics

We consider a dynamical system that follows a stationary and time-continuous Markov process $x_t$ in a generally large and continuous state space $\Omega$. $x_t$ is assumed to be ergodic with a unique stationary density $\mu(x)$. In order to be independent of specific dynamical models we use the general transition density $p_\tau(x_t, x_{t+\tau})$, i.e., the conditional probability density that given the system is at point $x_t \in \Omega$ at time $t$, it will be found at point $x_{t+\tau} \in \Omega$ a lag time $\tau$ later. We will at this point also assume that the dynamics obey microscopic detailed balance, i.e., $\mu(x_t) p_\tau(x_t, x_{t+\tau}) = \mu(x_{t+\tau}) p_\tau(x_{t+\tau}, x_t)$, which is true for systems that are not driven by external forces. In this case, $\mu(x)$ is a Boltzmann distribution in terms of the system’s Hamiltonian. The following rate theory can also be formulated
for some nonreversible dynamics, but the notation becomes more involved and the quantities are more difficult to interpret; we therefore restrict our discussion to the reversible case here.

Since we are interested in the slow processes we rewrite the transition density as a sum of relaxation processes that are associated with different intrinsic rates by expanding it into terms of the eigenvalues and eigenfunctions $\psi_i$ of the corresponding transfer operator \[14, 16].

\[ p_\tau(x_t, x_{t+\tau}) = \sum_{i=1}^{\infty} e^{-\kappa_i \tau} \psi_i(x_t) \mu(x_{t+\tau}) \psi_i(x_{t+\tau}). \]  

Here,

\[ \lambda_i(\tau) = e^{-\kappa_i \tau} \]  

are eigenvalues of the propagator that decay exponentially with lag time $\tau$. We order relaxation rates according to $\kappa_1 < \kappa_2 \leq \kappa_3 \leq \ldots$ and thus $\lambda_1(\tau) > \lambda_2(\tau) \geq \lambda_3(\tau) \geq \ldots$. The first term is special in that it is the only stationary process: $x_1 = 0$, $\lambda_1(\tau) = 1$, $\psi_1(x) = 1$, thus the first term of the sum is identical to $\mu(x)$. All other terms can be assigned a finite relaxation rate $\kappa_i$, or a corresponding relaxation timescale $t_i = \kappa_i^{-1}$, which are quantities of our interest. The eigenfunctions $\psi_i$ are independent of $\tau$ and determine the structure of the relaxation process occurring with rate $\kappa_i$. The sign structure of $\psi_i(x)$ determines between which substates the corresponding relaxation process is switching and is thus useful for identifying metastable sets, i.e., sets of states that are long-living and interconvert only by rare events \[14, 28]. The eigenfunctions are chosen to obey the normalization conditions

\[ \langle \psi_i, \psi_j \rangle_\mu = \int_\Omega dx \, \psi_i(x) \psi_j(x) \mu(x) = \delta_{ij} \]  

and integration always runs over the full space of the integrated variable if not indicated otherwise. At a given timescale $\tau$ of interest, fast processes with $\kappa \gg \tau^{-1}$ (and correspondingly $t_i \ll \tau$) will have effectively vanished and we are typically left with relatively few slowly-relaxing processes.

Finally, we define the $\mu$-rewighted eigenfunctions.

\[ \phi_i(x) = \mu(x) \psi_i(x) \]  

such that the normalization condition of eigenfunctions can be conveniently written as

\[ \langle \phi_i, \phi_j \rangle = \delta_{ij}. \]  

Finally, the correlation density $c_\tau(x_t, x_{t+\tau})$, i.e., the joint probability density of finding the system at $x_t$ at time $t$ and at $x_{t+\tau}$ at time $t+\tau$ is related to the transition density $p_\tau$ by

\[ c_\tau(x_t, x_{t+\tau}) = \mu(x_t) \, p_\tau(x_t, x_{t+\tau}). \]  

### Observed dynamics and rate theory

Let us consider the case that we are only interested in a single relaxation process – the slowest. Below, we sketch a rate theory for this case. Details of the derivation can be found in the supplementary information (SI). Based on the definitions above, the correlation density can then be written as:

\[ c_\tau(x_t, x_{t+\tau}) = \mu(x_t) \mu(x_{t+\tau}) + e^{-\kappa_2 \tau} \mu(x_t) \psi_2(x_t) \mu(x_{t+\tau}) \psi_2(x_{t+\tau}) + e^{-\kappa_3 \tau} \mu(x_t) p_{\tau,\text{fast}}(x_t, x_{t+\tau}) \]  

with

\[ p_{\tau,\text{fast}}(x_t, x_{t+\tau}) = \sum_{i=3}^{\infty} e^{-\kappa_i \tau} \psi_i(x_t) \mu(x_{t+\tau}) \psi_i(x_{t+\tau}) \]  

subsuming the fast processes.

**Exact rate:** The exact rate of interest $\kappa_2$ can be recovered as follows: If we know the exact corresponding eigenfunction $\psi_2(x)$, it follows from Eq. \[1\] and \[3\] that its autocorrelation function evaluates to:

\[ \lambda_2(\tau) = \langle \psi_2(x_t) \psi_2(x_{t+\tau}) \rangle_t \]

\[ = \int_\Omega dx_t \int_\Omega dx_{t+\tau} c_\tau(x_t, x_{t+\tau}) \psi_2(x_t) \psi_2(x_{t+\tau}) \]

\[ = e^{-\kappa_2 \tau} \]  

where $\langle \cdot \rangle_t$ denotes the time average, that is here identical to the ensemble average due to ergodicity of the dynamics. $\langle \psi_2(0) \psi_2(\tau) \rangle_t$ yields the exact eigenvalue $\lambda_2(\tau)$ and thus also an exact rate estimate of $\kappa_2 = -\tau^{-1} \ln \lambda_2(\tau) = \kappa_2$, independently of the choice of $\tau$.

**Projected dynamics without measurement noise:** Suppose we observe the dynamics of an order parameter $y \in \mathbb{R}$ that is a function of the configuration $x$. Examples are the distance between two groups of the molecule, or a more complex observable such as the Förster resonance transfer efficiency associated to a given configuration. See Fig.\[\square\] for an illustration. We first assume that no additional measurement noise is present. The analysis of a molecular dynamics simulation where a given order parameter is monitored is one example of such a scenario. Here, it is not possible to compute the rate via Eq. \[9\] or some direct approximation of Eq. \[9\], since the full configuration space $\Omega$ in which the eigenfunction $\psi_2$ exists can no longer be recovered once the dynamics has been projected onto an order parameter. Instead, we are forced to work with functions of the observable $y \in \mathbb{R}$.

We have two options for deriving the relevant rate equations for the present scenario: As a first option, we note that a projection that is free of noise can be regarded as a function $y(x) : \Omega \to \mathbb{R}$. Thus, any function $\hat{y}_2(y)$ of elements in observable space $\mathbb{R}$ that aims at approximating the dominant eigenfunction $\psi_2$ can be also regarded as a function in full space $\Omega$ via $\hat{y}_2(y) = \psi_2(y(x))$. When
following this idea, one can use the variational principle of conformation dynamics [29] (see also the discrete-state treatment in [20]), in order to derive the rate equations for the observed space dynamics. See SI for details.

However, since we aim at adding measurement noise in a second step, we derive a more general approach (see SI), which is summarized subsequently. Consider the function $\chi_p(y \mid x)$ that denotes the output probability density with which each configuration of the full state space, $x \in \Omega$, yields a measured value $y \in \mathbb{R}$. In the case of simply projecting $x$-values without noise to specific $y$-values, $\chi$ has the simple form:

$$
\chi_p(y \mid x) = \begin{cases} 1 & y = y(x) \\ 0 & \text{else} \end{cases}
$$

This allows the correlation density in observable space to be written as:

$$
c_\tau(y_0, y_\tau) = \int_\Omega dx_0 \int_\Omega dx_\tau \chi_p(y_\tau \mid x_\tau) \chi_p(y_0 \mid x_0) c_\tau(x_0, x_\tau) = \mu^\psi(y_0) \mu^\psi(y_\tau) + \sum_{i=2}^\infty \lambda_i(\tau) \phi_i^\psi(y_0) \phi_i^\psi(y_\tau)
$$

where we have used superscript $\psi$ to indicate the projection of a full configuration space function onto the order parameter: $\mu^\psi(y)$ is the observed stationary density that can be estimated from a sufficiently long simulation by histogramming the values of $y$. Mathematically, it is given by:

$$
\mu^\psi(y) = \int_\Omega dx \chi_p(y \mid x) \mu(x)
$$

$\phi_i^\psi$ are the projected eigenfunctions:

$$
\phi_i^\psi(y) = \int_\Omega dx \chi_p(y \mid x) \phi_i(x)
$$

In order to arrive at an expression of the rate $\kappa_2$, we propose a trial function in observation space, $\tilde{\psi}_2(y)$, which we require to be normalized by:

$$
\langle \tilde{\psi}_2, 1 \rangle_{\mu^\psi} = 0, \quad \langle \tilde{\psi}_2, \tilde{\psi}_2 \rangle_{\mu^\psi} = 1,
$$

and evaluate its autocorrelation function as:

$$
\langle \tilde{\psi}_2(y_0) \tilde{\psi}_2(y_\tau) \rangle = \int_\mathbb{R} dy_0 \int_\mathbb{R} dy_\tau \tilde{\psi}_2(y_0) c_\tau(y_0, y_\tau) \tilde{\psi}_2(y_\tau) = \alpha_g e^{-\kappa_2 \tau} + \sum_{i>2} \langle \tilde{\psi}_2, \phi_i^\psi \rangle^2 e^{-\kappa_i \tau}
$$

where

$$
\alpha_g = \langle \tilde{\psi}_2, \phi_1^\psi \rangle^2.
$$

In contrast to Eq. [10], both $\tilde{\psi}_2$ and $\phi_i^\psi$ live on the observable space. In the special case that $\tilde{\psi}_2(x)$ is constant in all other variables than $y(x)$, the projection is lossless ($\phi_2^\psi(y(x)) = \phi_2(x)$ and $\psi_2^\psi(y(x)) = \psi_2(x)$ for all $x$), and using the choice $\tilde{\psi}_2 = \psi_2^\psi$, we recover $\langle \tilde{\psi}_2, \psi_2^\psi \rangle = \langle \psi_2, \phi_2 \rangle = 1$, and thus the exact rate estimate via Eq. [2]. In general, however, the eigenfunction $\phi_2^\psi(x)$ does vary in other variables than $y$, and therefore $\tilde{\psi}_2$ can at best approximate the full-space eigenfunction via $\tilde{\psi}_2(y(x)) \approx \psi_2(x)$.

**Observed dynamics with measurement noise:** Suppose that an experiment is conducted in which each actual order parameter value $y(x) \in \mathbb{R}$ is measured with additional noise, yielding the observed value $o \in \mathbb{R}$. In time-binned single-molecule fluorescence experiments such noise may come from photon counting shot noise for a given binning size. In optical tweezer experiments such noise may come from bead diffusion and handle elasticity, assuming that bead and handle dynamics are faster than the kinetics of the molecule of interest. See Fig. 1 for an illustration. Note that we only treat the situation of uncorrelated noise. In situations where the experimental setup changes the kinetics, e.g., when the optical bead diffusion is slow, thus exhibiting different transition rates than the isolated molecule, our analysis always reports the rate of the experimental setup. The task of correcting the measured rates so as to estimate the rates of the pure molecule is beyond the scope of this work and can, for example, be attempted via dynamical deconvolution [30][31].

Like before, the probability of observing a measurement value $o \in \mathbb{R}$ given that the true configuration was $x \in \Omega$ can be given by an output probability:

$$
\chi_{pd}(o \mid x) = \int_\mathbb{R} dy \chi_d(o \mid y) \chi_p(y \mid x),
$$

which concatenates the projection from $x$ to the value of the order parameter, $\chi_p(y \mid x)$, and the subsequent dispersion of the signal by noise, $\chi_d(o \mid y)$. Despite the fact that dispersion operates by a different physical process than projection, the same analysis as above applies. We define the projected and dispersed stationary density and eigenfunctions:

$$
\mu^o(o) = \int_\Omega dx \chi_{pd}(o \mid x) \mu(x)
$$

$$
\phi_i^o(o) = \int_\Omega dx \chi_{pd}(o \mid x) \phi_i(x)
$$

which are “smeared out” by noise compared to the purely projected density and eigenfunctions $\phi_i^\psi$. As above, the autocorrelation function of a probe function $\tilde{\psi}_2(o)$ is given by:

$$
\langle \tilde{\psi}_2(o_0) \tilde{\psi}_2(o_\tau) \rangle = \alpha_g e^{-\kappa_2 \tau} + \sum_{i>2} \langle \tilde{\psi}_2, \phi_i^o \rangle^2 e^{-\kappa_i \tau}
$$
Many rate estimators operate by defining a single dividing surface which splits the state space into reactants A and products B. Calling $h_A(o)$ the indicator function which is 1 for set A and 0 for set B, one may define the normalized fluctuation autocorrelation function of state $A$ \[ \bar{\lambda}_2(t) = \frac{(h_A(0)h_A(t)) - \langle h_A \rangle^2}{\langle h_A^2 \rangle - \langle h_A \rangle^2} = \langle \bar{\psi}_2(0)\bar{\psi}_2(t) \rangle \] (23)

that can also be interpreted as an autocorrelation function $\lambda_2(t)$ with a step function $\bar{\psi}_2(t)$, \[ \bar{\psi}_2(t) = \frac{h_A(t) - \langle h_A \rangle}{\sqrt{\langle h_A^2 \rangle - \langle h_A \rangle^2}}, \]\[ \bar{\psi}_2(0) = 1 \] \[ \bar{\psi}_2(t) = 0 \] \[ \bar{\psi}_2(\omega) = \delta(\omega) \]

Many commonly used rate estimators consist of two steps: (1) they (explicitly or implicitly) calculate an autocorrelation function $\bar{\lambda}_2(\tau)$ of some function $\bar{\psi}_2$, and (2) transform $\bar{\lambda}_2(\tau)$ into a rate estimate $\bar{\kappa}_2$. In order to derive an optimal estimator, it is important to understand how the systematic error of the estimated rate depends on each of the two steps. The SI contains a detailed derivation of the subsequent results.

Existing rate estimators

Many rate estimators operate by defining a single dividing surface which splits the state space into reactants A and products B. Calling $h_A(o)$ the indicator function which is 1 for set A and 0 for set B, one may define the normalized fluctuation autocorrelation function of state $A$ \[ \bar{\lambda}_2(t) = \frac{(h_A(0)h_A(t)) - \langle h_A \rangle^2}{\langle h_A^2 \rangle - \langle h_A \rangle^2} = \langle \bar{\psi}_2(0)\bar{\psi}_2(t) \rangle \] (23)

that can also be interpreted as an autocorrelation function $\lambda_2(t)$ with a step function $\bar{\psi}_2(t)$, \[ \bar{\psi}_2(t) = \frac{h_A(t) - \langle h_A \rangle}{\sqrt{\langle h_A^2 \rangle - \langle h_A \rangle^2}}, \]\[ \bar{\psi}_2(0) = 1 \] \[ \bar{\psi}_2(t) = 0 \] \[ \bar{\psi}_2(\omega) = \delta(\omega) \]

Many rate estimators operate by defining a single dividing surface which splits the state space into reactants A and products B. Calling $h_A(o)$ the indicator function which is 1 for set A and 0 for set B, one may define the normalized fluctuation autocorrelation function of state $A$ \[ \bar{\lambda}_2(t) = \frac{(h_A(0)h_A(t)) - \langle h_A \rangle^2}{\langle h_A^2 \rangle - \langle h_A \rangle^2} = \langle \bar{\psi}_2(0)\bar{\psi}_2(t) \rangle \] (23)

that can also be interpreted as an autocorrelation function $\lambda_2(t)$ with a step function $\bar{\psi}_2(t)$, \[ \bar{\psi}_2(t) = \frac{h_A(t) - \langle h_A \rangle}{\sqrt{\langle h_A^2 \rangle - \langle h_A \rangle^2}}, \]\[ \bar{\psi}_2(0) = 1 \] \[ \bar{\psi}_2(t) = 0 \] \[ \bar{\psi}_2(\omega) = \delta(\omega) \]
in the special case that \( \kappa_3 \gg \kappa_2 \) (time scale separation), the error is approximately given by \( \kappa_2 (1 - \alpha) / \alpha \). Thus, the error of this estimator becomes zero for the ideal case of an ideal reaction coordinate (\( \alpha = 1 \)), but may be very huge if the reaction coordinate is poor (\( \alpha < 1 \)).

**Single-\( \tau \) rate estimators:** A simple rate estimator is to take the value of the autocorrelation function of some function \( \psi_2 \) at a single value of \( \tau \), and transform it into a rate estimate by virtue of Eq. (22). We call these estimators single-\( \tau \) estimators. Ignoring statistical uncertainties, they yield a rate estimate of the form

\[
\tilde{\kappa}_2,\text{single} = -\frac{\ln \tilde{\lambda}_2(\tau)}{\tau} \tag{27}
\]

Quantitatively, the error can be bounded by the expression (see derivation in the SI):

\[
\tilde{\kappa}_2,\text{single} - \kappa_2 \lesssim -\frac{\ln \alpha}{\tau}. \tag{28}
\]

The error becomes identical to this bound for systems with a strong timescale separation, \( \kappa_3 \gg \kappa_2 \). Eq. (28) decays relatively slowly in time (with \( \tau^{-1} \)). See Fig. 2 for a two-state example. It will be shown below that methods that estimate rates from counting the number of transitions across a dividing surface, such as Markov state models, are single-\( \tau \) estimators and are thus subject to the error given by Eq. (28).

The systematic error of single-\( \tau \) estimators results from the fact that Eq. (27) effectively attempts to fit the tail of a multi-exponential decay \( \tilde{\lambda}_2(\tau) \) by a single-exponential with the constraint \( \tilde{\lambda}_2(0) = 1 \). Unfortunately, the ability to improve these estimators by simply increasing \( \tau \) is limited because the statistical uncertainty of estimating Eq. (22) quickly grows in \( \tau \).

**Multi-\( \tau \) rate estimators:** To avoid the error given by Eq. (28), it is advisable to estimate the rate by evaluating the autocorrelation function \( \tilde{\lambda}_2(\tau) \) at multiple values of \( \tau \). This can be done e.g., by performing an exponential fit to the tail of \( \tilde{\lambda}_2(\tau) \), thus avoiding the constraint \( \tilde{\lambda}_2(0) = 1 \). The estimation error \( \tilde{\kappa}_2 - \kappa_2 \) has the
form
\[ \hat{\kappa}_{2,\text{mult}} - \kappa_2 < c \left( 1 - \frac{\tau}{\alpha} \right) e^{-\tau_1(k_3-k_2)} \] (29)

where \( \tau_1 \) is the first lag time from the series \( (\tau_1, ..., \tau_m) \) used for fitting, and the constant \( c \) also depends on the lag times and the fitting algorithm used. The SI shows that for several fitting algorithms, such as a least-squares procedure at the time points \( (\tau, 2\tau, ..., m\tau) \), \( c \) is such that
\[ \hat{\kappa}_{2,\text{mult}} \leq \hat{\kappa}_{2,\text{single}}. \] (30)

Thus, the multi-\( \tau \) estimator is always better than the single-\( \tau \) estimator (see SI). The main advantage of multi-\( \tau \) estimators is that their convergence rate is exponential in \( \tau \) when the time scale separation \( k_3 - k_2 \) is not vanishing (compare to Eq. [28]). Thus, multi-\( \tau \) estimators are better when the timescale separation between the slowest and the other relaxation rates in the system is larger.

Note that the systematic error of a multi-\( \tau \) estimator functions \( \gamma \) when the time scale separation between the slowest and the other relaxation rates in the system is larger.

In the absence of statistical error, all of the above rate estimation methods yield an overestimation of the rate, \( \hat{\kappa}_2 \geq \kappa_2 \).

Optimal choice of \( \psi_2 \)

It was shown above that multi-\( \tau \) estimators are the best choice for converting an autocorrelation function into a rate estimate. However, what is the best possible choice \( \hat{\psi}_2 = \psi_{2,\text{optimal}} \) given a specific observed time series \( o_t \)? In other words, which function should the observed dynamics be projected upon in order to obtain an optimal rate estimator? Following Eq. [28], the optimal choice \( \hat{\psi}_2 \) is the one which maximizes the parameter \( \alpha \), as this will minimize the systematic error from a direct rate estimation by virtue of Eq. [29], and also minimize the systematic error involved in estimating \( \kappa_2 \) from an exponential fit to Eq. [22]. We are thus seeking the solution of:
\[ \hat{\psi}_2 = \arg \max_{\psi_2} \alpha = \arg \max_{\psi_2} \lambda_2(\tau) \] (31)

subject to the normalization Eq. [14]. Here, \( \arg \max_{\psi_2} \alpha \) denotes the function that maximizes \( \alpha \) over the space of functions \( \psi_2(o) \). If the system has two-state kinetics, i.e., only \( \psi_1(x) = 1 \) and \( \psi_2(x) \) are present as dominant eigenfunctions, the problem [31] is solved by the projected eigenfunction:
\[ \hat{\psi}_2 = \psi_2 \] (32)

How can the best possible \( \psi_2 \) be determined from the observed time series? For a sufficiently large set of \( n \) basis functions, \( \gamma = \{\gamma_1(o), ..., \gamma_n(o)\} \), the optimal eigenfunction \( \hat{\psi}_2 \) is approximated by a linear combination \( \hat{\psi}_2(o) \approx \sum_{i=1}^n c_i \gamma_i(o) \) with coefficients \( c = \{c_1, ..., c_n\} \). When \( \gamma_i(o) \) is chosen to be an orthogonal basis set, then \( \hat{\psi}_2 = \arg \max_{\psi_2} \alpha \) can be approximated by the Ritz method [29][35]. An easy way to do this approximation in practice is to perform a fine discretization of the observable \( o \) by histogram windows. Using a binning with bin boundaries \( b_1, ..., b_{n+1} \), and the corresponding indicator functions
\[ \gamma_i(o) = \begin{cases} 1 & \text{if } i \in [b_i, b_{i+1}) \\ 0 & \text{else} \end{cases} \] (33)

then the above optimization problem is solved by estimating the transition probability matrix with elements
\[ T_{ij} = P(o(\tau) \in [b_j, b_{j+1}) | o(0) \in [b_i, b_{i+1})] \] (34)

and calculating \( c \) as the second eigenvector:
\[ Tc = \lambda_2 c \] (35)

where \( \lambda_2 < 1 \) is the second-largest eigenvalue of \( T \). If the system has two-state kinetics, i.e., only \( \psi_1(x) = 1 \) and \( \psi_2(x) \) are present as dominant eigenfunctions, the estimate \( \hat{\psi}_2 \) is independent of the choice of \( \tau \) in Eq. [34]. Thus, in real systems, \( \tau \) should be chosen to be at least a multiple of \( k_2^{-1} \) (e.g., \( \tau \geq 3k_2^{-1} \), as indicated by a constant rate \( k_2 \) estimate using a multi-\( \tau \) estimator (Eq. [29]). Note that a given optimal \( \hat{\psi}_2(o) \) can still be used with single-\( \tau \) and multi-\( \tau \) rate estimators that would produce different estimates for \( \kappa_2 \).

Note that \( \hat{\psi}_2 \) according to the procedure described here is only optimal for the case when the observed signal is obtained by projecting the high-dimensional data onto the observable, but is no longer optimal in the presence of noise, and especially large noise. In order to choose \( \hat{\psi}_2 \) optimal when noise is present, a generalized hermitian eigenvalue problem must be solved instead of Eq. [35] which includes a mixing matrix whose elements quantify how much the observable bins are mixed due to measurement noise. Since this approach is not very straightforward and in most practical cases only leads to small improvements, we won’t pursue it here. We rather note that it is often practical to reduce the noise level by carefully filtering the recorded data, provided that the filter length is much shorter than the timescales of interest.

Reaction coordinate quality (RCQ)

Evaluating how suitable a given observable in terms of capturing the slow kinetics is of great general interest. Although there is not a unique way of quantifying this suitability of the observable, the term reaction coordinate quality (RCQ) is often used. Previous studies have proposed ways to measure the RCQ that are based on comparing the observed dynamics to specific dynamical models or testing the ability of the observable to model
the committor or splitting probability between two chosen end-states A and B \cite{11, 15, 38}. These metrics are either only valid for specific models of dynamics or themselves require a sufficiently good separation of A and B by definition, restricting their applicability to observables with rather good RCQs.

The prefactor \( \alpha \) in Eq. (22) is a measure between 0 and 1, quantifying the relative amplitude of the slowest relaxation in the autocorrelation function that is used to estimate the rate from. By virtue of Eqs. (28) and (29), \( \alpha \) quantifies how large the error in our rate estimate can be. Therefore, we propose to that \( \alpha \) may serve as a basis to define the quality of the reaction coordinate. However, \( \alpha \) depends on the way the data is analyzed, namely the functional form \( \bar{\psi}_2(o) \) used to compute the autocorrelation function \( \overline{\lambda}_2(\tau) \). In order to remove this ambiguity we note that for the optimal choice \( \bar{\psi}_2 = \psi_2 \) (Eq. (11)), \( \alpha \) is maximized and in this case we denote this prefactor \( \alpha_0 \), where \( \hat{\alpha}_o = \alpha_0(\psi_2) \geq \alpha_0(\bar{\psi}_2) \). We propose \( \alpha_0 \) to serve as a measure for the quality of the reaction coordinate. This definition of RCQ is very general, as it makes no assumptions about the class of dynamics in the observed coordinate, and does not depend on any subjective choices such as the choice of two reaction end-states A and B in terms of the observable o. Through the derivation above it has also been shown that \( \alpha \) measures the fraction of amplitude by which the slowest process is observable, which is exactly the property one would expect from a measure of the RCQ: \( \hat{\alpha}_0 \) is 1 for a perfect reaction coordinate and 0 if the slowest process is exactly orthogonal to the observable, or has been completely obfuscated by noise.

Using this definition, the reaction coordinate quality is not purely a property of the molecular order parameter used, but rather a property of the measurement signal (including the effects of experimental noise and filtering used to preprocess the data). Yet, an analyst is typically interested in the RCQ \( \hat{\alpha}_y \) that is due to the choice of the molecular order parameter. Unfortunately, when an experiment is conducted, only the effective or apparent RCQ \( \hat{\alpha}_o \) is available. Unless a quantitative model of the dispersion function \( \chi_d(o \mid y) \) is known, the RCQ \( \hat{\alpha}_y \) before adding noise cannot be recovered. (see also Fig. 1 for an illustration). However, we can still quantitatively relate \( \hat{\alpha}_y \) and \( \hat{\alpha}_o \). For this, we derive a theory of observation quality. While the detailed derivation is found in the SI, we just summarize the most important results here:

1. When observing the order parameter y without noise and projecting the observation onto the optimal indicator function \( \bar{\psi}_2 = \psi_2 \), the RCQ can be expressed as the weighted norm of the projected eigenfunction, expressed by the scalar product:

\[
\hat{\alpha}_y = \langle \phi_2^y, \psi_2^y \rangle \tag{36}
\]

2. Unless the projection perfectly preserves the structure of the full-space eigenfunction \( \psi_2 \), we have \( \hat{\alpha}_y < 1 \). Thus, almost every observable reduces the RCQ.

3. When additional noise is present, the RCQ can be expressed as the weighted norm of the projected and noise-distorted eigenfunction:

\[
\hat{\alpha}_o = \langle \phi_2^o, \psi_2^o \rangle \tag{37}
\]

4. The RCQ \( \hat{\alpha}_y \) due to the projection onto the selected molecular order parameter alone, and the apparent RCQ \( \hat{\alpha}_o \) including measurement noise are related by:

\[
\hat{\alpha}_o \leq \hat{\alpha}_y \tag{38}
\]

i.e., adding noise always further reduces the apparent RCQ compared to a situation where the order parameter y can be directly measured.

The inequality [38] implies that we can use the observable RCQ \( \hat{\alpha}_o \) in order to both optimize the experimental setup and the order parameter used. For example in an optical tweezer measurement, we can change laser power and handle length so as to maximize \( \hat{\alpha}_o \), thus making \( \hat{\alpha}_o \) and \( \hat{\alpha}_y \) more similar and reducing the effect of noise on the measurement quality. On the other hand, since \( \hat{\alpha}_o \) is a lower bound for \( \hat{\alpha}_y \), we can also use it to ensure a minimal projection quality: When the measurement setup itself is kept constant we can compare the measurements of different constructs (e.g., different FRET labeling positions or different attachment sites in a tweezer experiment). The best value \( \hat{\alpha}_o \) corresponds to the provably best construct.

Finally, \( \hat{\alpha}_o \) can be determined by fitting the autocorrelation function of \( \bar{\psi}_2 \), as described in the spectral estimation procedure described below. Figs. 2-4 show estimates of the RCQ of different observed dynamics using different rate estimators.

Markov (state) models (MSMs)

MSMs have recently gained popularity in the modeling of stochastic dynamics from molecular simulations \cite{12, 14, 15, 37, 38}. MSMs can be understood as a way of implicitly performing rate estimates via discretizing state space into small substates. Let us consider a MSM obtained by finely discretizing the observed space y into bins and estimating a transition matrix \( T(\tau) \) amongst these bins. We have seen that this procedure approximately solves the optimization problem of Eq. (31), and the leading eigenvector of \( T(\tau) \) approximates the projection of the true second eigenfunction, \( \bar{\psi}_2(o) \), available for the given observable o. Ref [15] has suggested to use the implied timescale \( \tilde{t}_2 = -\tau / \ln(\overline{\lambda}_2(\tau)) \) as an estimate for the system’s slowest relaxation timescale, and at the same time for a test which choice of \( \tau \) leads to a MSM with a small approximation error. These implied
timescales correspond to the inverse relaxation rates and therefore the MSM rate estimate is described by Eq. (27) with the choice $\hat{\psi}_2 = \hat{\psi}_2$. A sufficiently fine MSM thus serves as an optimal single-$\tau$ rate estimator as it uses the maximum RCQ $\hat{\alpha}$ for observed signal that is being discretized. However, when this signal has a poor RCQ $\hat{\alpha}$ since it is poorly separating the slowly-converting states, there is a substantial rate estimator error according to Eq. (28) that decays slowly with $\tau^{-1}$. This explains the slow convergence of implied timescales shown in recent MSM studies [12][15][32].

Estimating state densities and microscopic transition rates

When the rate $k_2$ is exactly known, the microscopic transition rates between the two interchanging states, $k_{AB}$ and $k_{BA}$ could be calculated from the equations:

$$\pi_A k_{AB} = \pi_B k_{BA},$$ (39)

$$k_{AB} + k_{BA} = k_2$$ (40)

where $\pi_A$ and $\pi_B$ are the stationary probabilities of states $A$ and $B$:

$$\pi_A = \int_0^1 d\mu_A(o)$$ (41)

$$\pi_B = \int_0^1 d\mu_B(o) = 1 - \pi_B$$ (42)

with $\mu_A(o)$ and $\mu_B(o)$ being the partial densities of states $A$ and $B$ respectively.

Here we attempt to estimate both the partial densities, $\mu_A(o)$ and $\mu_B(o)$, and from these the microscopic transition rates via Eqs. (39) and (40). The difficulty is that $\mu_A(o)$ and $\mu_B(o)$ can significantly overlap in $o$, both due to the way the order parameter used projects the molecular configurations onto the observable, and due to noise broadening of the measurement device. This reveals a fundamental weakness of dividing-surface approaches. Although a dividing surface estimator can estimate the rate $k_2$ for sufficiently large $\tau$ without bias via Eq. (29), it cannot distinguish between states on one side of the barrier, and thus assumes the partial densities $\mu_A(o)$ and $\mu_B(o)$ to be given by cutting the full density $\mu(o)$ at the dividing surface. When the true partial densities overlap, this estimate can be far off (compare the curves in Fig. 2, second column). Consequently, wrong estimates for the microscopic rates $k_{AB}$ and $k_{BA}$ are obtained when Eqs. (39) and (40) are used with $\pi_A$ and $\pi_B$ that are simply the total densities “left” and “right” of the dividing surface.

Hidden Markov models approach this problem by proposing a specific functional form of $\mu_A(o)$ and $\mu_B(o)$, for example a Gaussian distribution, and then estimating the parameters of this distribution with an optimization algorithm [21][23]. This is approach is very powerful when the true functional form of the partial densities is known, but will give wrong estimates when the wrong functional form is used.

Here, we propose a nonparametric solution that can estimate the form of the partial densities $\mu_A(o)$ and $\mu_B(o)$, and the microscopic transition rates $\hat{k}_{AB}$ and $\hat{k}_{BA}$ in most cases without bias. For this, we employ PCCA+ theory [17][40] which is based on PCCA theory [23][41] which allows for a way of splitting state space into substates and at the same time maintain optimal approximations to the exact eigenfunctions (here $\psi_2$). The state assignment must be fuzzy, i.e., instead of choosing a dividing surface that uniquely assigns points $o$ to either $A$ or $B$, we have fuzzy membership functions $\chi_A(o)$ and $\chi_B(o)$ with the property $\chi_A(o) + \chi_B(o) = 1$. These membership functions can be calculated after $\psi_2$ is known.

In order to compute the membership $\chi_A$ and $\chi_B$, the memberships of two points of the observable $o$ must be fixed. The simplest choice is to propose two points that are pure, i.e., that have a membership of 1 to $A$ and $B$ each. Such an approach is also proposed by the signal-pair correlation analysis approach [27] where the pure points need to be defined by the user. However at this point of our analysis, an optimal choice can be made, because the eigenfunction $\hat{\psi}_2$ has been approximated. Thus we propose to follow the approach of [40] and choose the $o$-values where $\hat{\psi}_2$ achieves a minimum and a maximum, respectively, as purely belonging to $A$ and $B$. Typically, these are the states the are on the left and the right boundary of the histogram in $o$. This approach will only start to give a biased estimate when the overlap of the $A$ and $B$ densities is so large that not even these extreme points are pure (see Fig. 2, last row for such an example).

Let $\hat{\psi}_2$ be the second eigenvector of the Markov model $T(\tau)$ of the finely binned observable (Eq. (35)). $\hat{\psi}_2$ is a discrete approximation to the projected eigenfunction $\hat{\psi}_2$. Following the derivation given in the appendix, the fuzzy membership functions on the discretized observable space are given by

$$\hat{\chi}_{A,i} = \frac{\max \hat{\psi}_2 - \hat{\psi}_{2,i}}{\max \hat{\psi}_2 - \min \hat{\psi}_2}$$ (43)

$$\hat{\chi}_{B,i} = \frac{\hat{\psi}_{2,i} - \min \hat{\psi}_2}{\max \hat{\psi}_2 - \min \hat{\psi}_2}.$$ (44)

Since we are restricted to the projected eigenfunction $\hat{\psi}_2$, we can determine the optimal choice $\hat{\chi}_A(o)$ and $\hat{\chi}_B(o)$ from $\hat{\psi}_2(o)$. Together with the estimated stationary density $\mu(o)$ which can e.g., be obtained by computing a histogram from sufficiently long equilibrium trajectories, the probability density of being in $A$ and $B$ when projected onto the observable $o$ can be calculated:

$$\mu_{A,i} = \hat{\chi}_{A,i} \mu_i$$ (45)

$$\mu_{B,i} = \hat{\chi}_{B,i} \mu_i.$$ (46)
The probability of being in A and B is thus given by:

$$\pi_A = \sum_i \mu_{A,i}$$

(47)

$$\pi_B = \sum_i \mu_{B,i} = 1 - \pi_A$$

(48)

and these probabilities can be used to split $\epsilon_2$ into microscopic transition rates $k_{AB}$ and $k_{BA}$:

$$\hat{k}_{AB} = \pi_B \hat{\kappa}_2$$

(49)

$$\hat{k}_{BA} = \pi_A \hat{\kappa}_2$$

(50)

Note that the assignment of labels “A” and “B” to parts of state space is arbitrary. Eq. (49) is the transition rate from A to B as defined by Eqs. (43)-(44), and Eq. (50) is the corresponding transition rate from B to A.

**Spectral estimation**

The optimal estimator for $\kappa_2$ is thus one that fits the exponential decay of $\lambda_2(\tau)$ while minimizing the fitting error Eq. (29). As analyzed above, the systematic fitting error is minimized by any multi-$\tau$ estimator. In order to obtain a numerically robust fit, especially in the case when statistical noise is present, it is optimal to fit to an exponential decay of $\hat{\kappa}_2$ that have maximum amplitude $\hat{\omega}_0$. This is approximately achieved by constructing a fine discretization MSM on the observed coordinate (see Section “Optimal choice of $\hat{\eta}_2$”). Thus, the optimal estimator of $\kappa_2$ proceeds as outlined in (1-6) below. The full spectral estimation algorithm (1-6) additionally provides estimates for the microscopic rates $k_{AB}, k_{BA}$, and for the partial densities $\mu_A$ and $\mu_B$:

1. Obtain a fine discretization of the observed coordinate $o$ into $n$ bins, say $[o_i, o_{i+1}]$ for $i = 1...n$.

2. Construct a row-stochastic transition matrix $T(\tau)$ for different values of $\tau$. The estimation of transition matrices from data have been described in detail [14]. A simple way of estimating $T(\tau)$ is the following: (i) for all pairs $i, j$ of bins, let $c_{ij}(\tau)$ be the number of times the trajectory has been in bin $i$ at time $t$ and in bin $j$ at time $t + \tau$, summed over all time origins $t$; (ii) estimate the elements of $T(\tau)$ by $T_{ij}(\tau) = c_{ij}(\tau)/\sum_k c_{ik}(\tau)$. A numerically superior approach is to use a reversible transition matrix estimator [14].

3. Calculate the discrete stationary probability $\mu$ and the discrete eigenvector $\hat{\psi}_2$ by solving the eigenvalue equations:

$$T^T(\tau)\mu = \mu$$

(51)

$$T(\tau)\hat{\psi}_2 = \lambda_2 \hat{\psi}_2$$

(52)

with the largest eigenvalues $\lambda_1 = 1 < \lambda_2 \leq \lambda_3$. $T^T$ denotes the transpose of the transition matrix. The $i$-th element of the vectors $\mu$ and $\hat{\psi}_2$ approximate the stationary density $\mu(o)$ and $\hat{\psi}_2$ on the respective point $o = \frac{a_i + a_{i+1}}{2}$. Functions $\hat{\mu}(o)$ and $\hat{\psi}_2(o)$ can be obtained by some interpolation method.

4. Estimate the relaxation rate $\hat{\kappa}_2$ and the RCQ $\hat{\alpha}$ via an exponential fit of $ae^{-\kappa_2 t}$ to the tail of $\lambda_2(\tau) = \langle \hat{\psi}_2(t)\hat{\psi}_2(t + \tau) \rangle_1$.

5. Calculate the partial densities $\mu_A$ and $\mu_B$ from Eqs. (45) and (46) using transition matrix eigenvectors estimated at a lagtime $\tau_{\text{min}}$ at which the rate estimate $\hat{\kappa}_2$ is converged.

6. Calculate the microscopic transition rates $k_{AB}$ and $k_{BA}$ from Eqs. (49) and (50).

Note that this estimator is optimal in terms of minimizing the systematic error. When dealing with real data, the amount of statistics may set restrictions of how fine a discretization is suitable and how large a lag time will yield reasonable signal to noise. For a discussion of this issue refer to e.g., [34].

**Illustrative two-state example**

To illustrate the theory and the concepts of this paper, we compare the behavior of different order parameters, measurement noise and different estimators in Fig. 2. The full-space model here is a two-dimensional model system using overdamped Langevin dynamics in a bistable potential. This choice was made because the exact properties of this system are known and the quality of different estimates can thus be assessed. The potential is chosen such that the eigenfunction associated with the slow process, $\hat{\psi}_2(x)$ varies in $x_1$ and is constant in $x_2$, such that the choice $o = x_1$ represents a perfect projection and the choice $o = x_2$ represents the worst situation in which the slow process is invisible.

Fig. 2I, II and III show three scenarios using $y = x_1$ (I, perfect order parameter - projection angle $0^\circ$), $y = \frac{1}{2}(x_1 + x_2)$ (II, intermediate-quality order parameter - projection angle $45^\circ$), and $y = \frac{1}{2}(x_1 + 3x_2)$ (III, poor order parameter - projection angle $72^\circ$). Additionally, we compare the results when the order parameter $y$ is traced without noise (“a” panels), and when measurement noise is added (“b” panels). Noise here consists of adding a uniformly distributed random number from the interval [-1,1] to the signal, such that the noise amplitude is roughly 25% of the signal amplitude.

Fig. 2 panels 1a and 1b show the apparent stationary density in the observable $y$, $\mu(y)$, or in the noisy observable $o$, $\mu(o)$, as a black solid line. The partial densities of states $A$ (orange) and $B$ (grey) which comprise the total stationary density are shown as well. The
lower part of the figure shows the observed eigenfunction associated with the two-state transition process ($\psi^o_2$ or $\psi^o_3$) as black solid line with grey background. It is apparent that when the quality of the observation is reduced, either by choosing a poor order parameter, or by adding experimental noise, the overlap of the partial densities increases and the continuous projected eigenfunction $\psi^o_3$ becomes smoother and thus increasingly deviates from the dividing surface model which is a step function switching at the dividing surface (dashed line).

Panels 2a and 2b show the reaction coordinate qualities (RCQ) in these different scenarios. The fact that the blue and red lines are approximately constant after $\tau = 5$ (when the fast processes have relaxed) shows that the RCQ can be reliably estimated at these lag time ranges using either the dividing surface or the spectral estimation approach. The red line (spectral estimation) gives the optimal, i.e., maximal, RCQ, which varies between 1 (perfect order parameter I) and 0.15 (poor order parameter III with additional measurement noise). It is seen that the optimal RCQ given by the spectral estimator can be much larger than the suboptimal RCQ estimated by the dividing surface estimator that uses a fit to the number correlation function Eq. (23) (blue line). This is especially apparent in the case of an intermediate-quality order parameter (Fig. 2-I-2a and II-2b).

Panels 3a and 3b show the estimate of the relaxation rate $k_2$ obtained for different scenarios. Each panel compares the Markov model (MSM) estimate (solid black line), the dividing surface estimate (solid blue line) and the spectral estimate (solid red line) to the exact value (dashed black line). For the multi-$\tau$ estimators (dividing surface and spectral estimation), the lag time $\tau$ specifies the start of a time range [$\tau$, $\tau + 10$] used for an exponential fit. In the case of a perfect order parameter (I), all estimators yield the correct rate at lagtimes $\tau > 5$ time steps (where the fast processes with rates $k_3$ or greater have disappeared). In this case, the noise does not deteriorate the estimate, because the partial densities of states $A$ and $B$ are still well separated. For intermediate-quality and poor order parameters, the MSM estimate breaks down dramatically, providing a strongly overestimated rate for $0 < \tau < 100$ time steps. Panels II-3a,b, and III-3a,b show the typical behavior $\tau^{-1}$-convergence of the MSM estimate predicted by the theory (Eq. (28)). Clearly, the MSM estimate will converge to the true value for very large values of $\tau$, but especially for the situation of a poor order parameter, the minimal $\tau$ required to obtain a small estimation error if typically larger than the timescale $k_3^{-1}$ of the slowest process, thus rendering a reliable estimation impossible. It is seen that the magnitude of the error for a given value of $\tau$ increases when going from panel II-3a to II-3b, from II-3a to III-3a, and from III-3a to III-3b. This is because in this sequence the RCQ deteriorates, as predicted by the theory of reaction coordinate qualities (see above), and hence the prefactor of the MSM error increases (see Eq. (28)). As predicted by Eq. (30), the multi-$\tau$ estimators (dividing surface and spectral estimate) are always better than the single-$\tau$ (MSM) estimate. As predicted by Eq. (29), both the dividing surface and spectral estimate of $k_2$ converge when the fast processes have died out (here approximately at $\tau > 5$ time steps). As seen in panels II-3b, III-3a and III-3b, the spectral estimate is more stable than the dividing surface estimate, i.e., it exhibits less strong fluctuations around the true value $k_2$. This is because the spectral estimate uses a higher RCQ $\hat{\kappa}_{o\tau}$ and thus the exponential tail of the autocorrelation function can be fitted using a larger amplitude of the process relaxation with rate $k_2$, achieving a better signal-to-noise ratio.

Panels 1a and 1b show the partial densities estimated with a dividing surface and PCCA+ (see above) and panels 4a and 4b show the microscopic transition rates estimated from Eqs. (19) and (50), either using the partial densities from the dividing surface, or PCCA+ (spectral estimate). As expected, the partial densities from the dividing surface estimate are wrong as soon as the states overlap in the observable, either due to choosing a poor order parameter, or to experimental noise. As a result, the dividing surface estimates for the microscopic rates $k_{AB}$ and $k_{BA}$ are biased for all these cases (panels II-4a,b, panels III-4a,b). The spectral estimate gives an unbiased estimate for intermediate overlap (panels II-4a,b). For strong overlap, even the spectral estimator starts to have a bias because no pair of observable states can be found that are uniquely assignable to states $A$ and $B$. Nonetheless, the spectral estimator still yields a smaller bias than the dividing surface estimator. Like for the relaxation rate ($k_2$) estimate, the spectral estimator exhibits less fluctuations due to a better signal-to-noise ratio.

Applications to optical tweezer data

In order to illustrate the performance of spectral estimation on real data, it is applied to optical tweezer measurements of the extension fluctuations of two biomolecules examined in a recent optical force spectroscopy study: the p5ab RNA hairpin [42] and the H36Q mutant of sperm whale apo-myoglobin at low pH [43]. The p5ab hairpin forms stem-loop structure with a bulge under native conditions (Fig. 3.5 and 3.6a) and zips/unzips repeatedly under the conditions used to collect data (Fig. 3.6a), while apo-myoglobin (crystal structure shown in Fig. 4.5) hops between unfolded and molten globule states at the experimental pH of 5 (Fig. 4.6a) [43]. Experimental force trajectory data were generously provided by the authors. Experimental details are given in Refs. [42,43], but we briefly summarize aspects of the apparatus and experimental data collection procedure relevant to our analysis.

The instrument used to collect both datasets was a dual-beam counter-propagating optical trap [44]. The
Figure 2: Estimation results using overdamped Langevin dynamics in a two-dimensional two-well potential that is projected onto different observables: (I) perfect projection, (II) intermediate-quality projection, (III) poor projection. Results are compared without noise (a panels) and with additional measurement noise (b panels). (1): Stationary density $\rho(y)$ in the observable (black) and the two partial densities of states $A$ and $B$. (2): Estimated RCQ factor $\alpha$ from spectral estimation (red), and from exponential fitting to the number correlation function using a dividing surface at $y = 0$ (blue). (3): Estimated relaxation rate $\kappa_2$ using the second eigenvalue of a transition matrix at $\tau$ (solid black), using an exponential fit to correlation function at the dividing surface at $\varphi = 0$ (blue), and spectral estimation (red). The black dotted line is the reference solution, obtained from a direct MSM estimate for $\tau = 50$ in row 1. (4): The transition rates $k_{AB}$ (dashed) and $k_{BA}$ (solid) calculated using the dividing surface estimate (blue) and from the partial densities of the spectral estimator (red).
molecule of interest was tethered to polystyrene beads by means of dsDNA handles, with one bead suctioned onto a pipette and the other held in the optical trap. A piezoactuator controlled the position of the trap and allowed position resolution to within 0.5 nm, with the instrument operated in passive (equilibrium) mode such that the trap was stationary relative to the pipette during data collection. The force on the bead held in the optical trap was recorded at 50 kHz, with each recorded force trajectory 60 s in duration.

Fig. 3 compares the results of different rate estimators for optical tweezer measurement of the p5ab RNA hairpin extension fluctuations. Fig. 3-5 shows a sketch of the RNA molecule under analysis and Fig. 3-6 shows the trajectory that was analyzed. The trajectory exhibits a two-state like behavior with state lifetimes on the order of tens of milliseconds. Fig. 3-1a shows the stationary probability density of measured pulling forces, exhibiting two nearly separated peaks. Fig. 3-2a shows the estimated reaction coordinate quality (RCQ) \( a_o \), which is approximately constant at lag times \( \tau > 5 \text{ ms} \), indicating a reliable estimate for this quantity at lag times greater than 5 ms. An optimum value of \( \hat{a}_o \approx 0.96 \) (spectral estimator) is found while the best possible dividing surface results in \( a_o \approx 0.94 \). These values indicate that the present reaction coordinate is well suited to separate the slowly interconverting states, and that different approaches, including a Markov model, a dividing surface estimate and a spectral estimate should yield good results.

Fig. 3-3a compares the estimates of the relaxation rate \( \kappa_2 \) using the direct MSM estimate (black), a fit to the fluctuation autocorrelation function using a dividing surface at the histogram minimum \( o = 12.75 \text{ pN} \) (blue), and spectral estimation (red). For the multi-\( \tau \) estimators (dividing surface and spectral estimation), the lag time \( \tau \) specifies the start of a time range \( [\tau, \tau+2.5 \text{ ms}] \) that was used for an exponential fit. All estimators agree on a relaxation rate of about \( \kappa_2 \approx 58 \text{ s}^{-1} \), corresponding to a timescale of about 17 ms. The MSM estimate is strongly biased for short lagtimes, exhibiting the slow \( \tau^{-1} \) convergence predicted by the theory for single-\( \tau \) estimators (Eq. 28). It converges to an estimate within 10% of the value from multi-\( \tau \) estimates after a lagtime of about 10 ms. The dividing surface and spectral estimators behave almost identical, and converge after about \( \tau = 5 \text{ ms} \). According to the error theory of multi-\( \tau \) estimators (Eq. 29), this indicates that there are additional, faster kinetics in the data, the slowest of which have timescales of 2-3 ms. In agreement with the theory (Eq. 30), the multi-\( \tau \) estimators (dividing surface, spectral estimate) converge faster than the single-\( \tau \) estimate (MSM).

As indicated in Fig. 3-1a, the substates estimated from PCCA+ are almost perfectly separated and can be well distinguished by a dividing surface at the histogram minimum \( o = 12.75 \text{ pN} \). Consequently, both the dividing surface estimate and the spectral estimate yield almost identical estimates of the microscopic transition rates - the folding rate being \( k_{AB} \approx 45 \text{ s}^{-1} \) and the unfolding rate being \( k_{BA} \approx 15 \text{ s}^{-1} \) (Fig. 3-4). In summary, the two-state kinetics of p5ab can be well estimated by various different rate estimators because the slowly-converting states are well separated in the experimental observable.

Fig. 3 panels b show estimation results for data that has been filtered by averaging over 50 frames (1 ms). This averaging further reduces the already small overlap between substates A and B while the filter length is much below the timescale of \( A - B \) interconversion. Therefore, filtering has a positive result on the analysis: The effective RCQ \( \hat{a}_o \) increases and is now approximately equal to 1 according to spectral estimation. The estimation results are largely identical to the case with noise. In Fig. 3-3b, the error made by the Markov model estimate has become smaller because the error prefactor reported in Eq. 28, in \( \hat{a}_o \), has become smaller. Note that in contrast to the unfiltered data analysis, some of the rate estimates (MSM and spectral estimate) underestimate the rate for small lag times \( \tau \). This is not in contradiction with our theory which predicts an overestimation of the rate for Markovian processes. By using the filter, one has effectively introduced memory into the signal, and the present theory will only apply at a lagtime \( \tau \) that is a sufficiently large multiple of the filter length, such that the introduced memory effects have vanished.

Fig. 4 shows estimation results for an optical tweezer experiment that probes the extension fluctuations of apo-myoglobin [45]. Fig. 4-5 shows a sketch of the experimental pulling coordinate (green arrows) depicted at the crystal structure of apo-myoglobin. Fig. 4-6 shows the trajectory that was analyzed. Out of the trajectories reported in [45] we have here chosen one where the two slowest-converting states have the largest overlap. While the trajectory indicates that there are at least two kinetically separated states, the stationary probability density of measured pulling forces (Fig. 4-6) does not exhibit a clear separation between these states in the measured pulling force. Thus, the order parameter used here, in combination with signal broadening by the experiment, does not clearly separate the kinetically distinct states. This is also indicated by Fig. 4-2a, which shows that the optimal reaction coordinate quality (RCQ) has a value of \( a_o \approx 0.5 \) (spectral estimator) at \( \tau = 15 \text{ ms} \) while the best possible dividing surface results in only \( a_o \approx 0.4 \) at \( \tau = 15 \text{ ms} \). Thus, the quality of the apo-myoglobin data is similar to that of the two-state model with intermediate-quality order parameter and noise (Fig. 2I-b). This data thus represents a much harder test for rate estimators than the p5ab hairpin and should show clear differences between different rate estimators.

Fig. 4 panels 3 and 4 compare the estimates of \( k_2 \) from the direct MSM estimate (black), a fit to the fluctuation autocorrelation function using a dividing surface at the
local histogram minimum of the binned data at $o = 4.6$ pN (blue), and spectral estimation (red). For the multi-$\tau$ estimators (dividing surface and spectral estimation), the lag time $\tau$ specifies the start of a time range $[\tau, \tau + 2.5$ ms] that was used for an exponential fit.

Fig. 4-3a shows again, that the MSM estimate of $k_2$ exhibits the slow $\tau^{-1}$ convergence predicted by the theory (Eq. (29)) and does not yield a converged estimate using lagtimes of up to 20 ms. Since the MSM estimate still significantly overestimates the rate at $\tau = 50$ ms when the relaxation process itself has almost entirely decayed, this estimator is not useful to analyze the apomyoglobin data. In contrast, both, the dividing surface multi-$\tau$ approach and the spectral estimator do yield a converged estimate of $k_2 \approx 26$ s$^{-1}$, corresponding to a timescale of about 38.5 ms (Fig. 4-3a). In Ref. 45 a Hidden Markov model with Gaussian output functions was used and the rate was estimated to be $k_2 \approx 46$ s$^{-1}$ corresponding to a timescale of $k_2 \approx 21$ ms. These results are consistent with our theory which shows that rate estimation errors lead to a systematic overestimation of the rate (and underestimation of the timescale). Fig. 4-4a clearly shows that the partial probabilities are not Gaussians, and enforcing Gaussian output probabilities will lead to a systematic error in the rate estimate. The smallest rate estimates are therefore the best estimates, which are here provided by the multi-$\tau$ estimators using either dividing surface or spectral estimation approaches.

In agreement with the theory (Eq. (30)), the multi-$\tau$ estimators (dividing surface, spectral estimate) converge faster than the single-$\tau$ estimate (MSM). The multi-$\tau$ estimators can be considered converged after a lagtime of about $\tau = 15$ ms. According to the Eq. (29), this indicates that there are additional, faster kinetics in the data, the slowest of which have timescales in the order of 5-10 ms. As the resulting timescale separation to the 25 ms process is rather small, it may be worthwhile to investigate further substates of the $A$ and $B$ states with multistate
approaches such as Hidden Markov Models [23] or pair correlation analysis [27]. Such an analysis is beyond the scope of the present paper on two-state rate theory.

As indicated in Fig. 4-a, the substates A and B estimated from PCCA+ do strongly overlap. Thus, even though the dividing surface estimator can recover the true relaxation rate \( \kappa_2 \), the estimated microscopic rates \( k_{AB} \) and \( k_{BA} \) strongly depend on the choice of the position of the dividing surface. Fig. 4-a shows the estimates the dividing surface multi-\( \tau \) estimator, evaluating to \( k_{AB} \approx 25 \text{s}^{-1} \) and \( k_{BA} \approx 5 \text{s}^{-1} \). In contrast, the spectral estimator yields estimates of \( k_{AB} \approx k_{BA} \approx 15 \text{s}^{-1} \). Like for the two-state model results shown in Fig. 3, the spectral estimate is numerically more stable in \( \tau \) compared to the dividing surface estimate as a result of achieving a better signal-to-noise ratio. Clearly, in the dividing surface approach it is possible to pick a dividing surface position which yields the same estimates for \( k_{AB} \) and \( k_{BA} \) like the spectral estimator. However, the dividing surface estimator itself does not provide any information which is the correct choice, and therefore this theoretical possibility is of no practical use. Supplementary Figure 2 compares the estimation results of \( k_2 \) and \( k_{AB}, k_{BA} \) for different choices of the dividing surface. In contrast to the dividing surface approach, the spectral estimator only assumes that the extreme values of \( o \) are pure, which is a much weaker requirement than assuming that an appropriate dividing surface exists (see theory), and hence provides more reliable rate estimates.

The “b”-panels in Fig. 4 show the effect of filtering the data on the estimation results. Here, the data was averaged over a window length of 1 \( \text{ms} \) corresponding to an averaging of 50 data points of the original 50 kHz data. Fig. 4-b shows that this filtering enhances the separation of states, and the apparent RCQ’s increase to about \( \hat{\alpha}_o \approx 0.7 \) (spectral estimate) and \( \hat{\alpha}_o \approx 0.6 \) (dividing surface). The relaxation rate \( \kappa_2 \) is still estimated to have \( \kappa_2 \approx 40 \text{s}^{-1} \) and the estimate becomes more robust for both the dividing surface and spectral estimates (Fig. 4-b). The MSM estimate slightly improves but is still significantly too high. Fig. 4-b shows that the dividing surface derived rate estimates \( k_{AB} \) and \( k_{BA} \) have slightly, while the spectral estimate remains at \( k_{AB} \approx k_{BA} \approx 15 \text{s}^{-1} \) independent of the filtering, which is in support of the reliability of the spectral estimate.

Conclusions

We have described a rate theory for observed two-state dynamical systems. The underlying system is assumed to be ergodic, reversible, and Markovian in full phase space, as fulfilled by most physical systems in thermal equilibrium. The observation process takes into account that the system is not fully observed, but rather by tracing one order parameter (the extension to multiple order multidimensional order parameters is straightforward). During the observation process, the observed order parameter may be additionally distorted or dispersed, for example by experimental noise. Such observed dynamical systems occur frequently in the molecular sciences, and appear both in the analysis of molecular simulations as well as of single-molecule experiments.

The presented rate theory for observed two-state dynamics is a generalization to classical two-state rate theories in two ways: First, most available rate theories assume that the system of interest is either fully observable, or the relevant indicators of the slowest kinetic process can be observed without projection error or noise. Secondly, most classical rate theories are built on specific dynamical models such as Langevin or Smoluchowski dynamics. The present theory explicitly allows the two kinetic states to overlap in the observed signal (either due to using a poor order parameter or to noise broadening), and does not require a specific dynamical model, but rather works purely based on the spectral properties of a reversible ergodic Markov propagator - hence the name spectral rate theory.

Given the spectral rate theory, the systematic errors of available rate estimators can be quantified and compared. For example, the relatively large systematic estimation error in the implied timescales / implied rates of Markov state models is explained. Additionally, the theory provides a measure for the reaction coordinate quality (RCQ) \( \hat{\alpha}_o \) of the observed signal, which is independent of any specific dynamical model and also does not need the definition of an “A” or “B” state and bounds the error in rates estimated from the observed signal. \( \hat{\alpha}_o \) includes effects of the order parameter measured as well as the effect of the experimental construct on the signal quality, such as experimental noise. It is shown that \( \hat{\alpha}_o \) is a lower bound to the RCQ due to choosing the order parameter, and can thus be used as an indicator to both improve the quality of the experimental setup and the choice of the order parameter.

The theory suggests steps to be taken to construct an optimal rate estimator, that minimizes the systematic error in the estimation of rates from an observed dynamical system. We propose such an estimator and refer to it as spectral estimator. It provides rather direct and optimal estimates for the following three types of quantities:

1. The reaction coordinate quality (RCQ) \( \hat{\alpha}_o \) of the observed signal.
2. The dominant relaxation rate \( \kappa_2 \), as well as the microscopic transition rates \( k_{AB} \) and \( k_{BA} \), even if A and B strongly overlap in the observable.
3. The partial probability densities, and hence projections of the states A and B in the observable, \( \mu^o_{i\mu}(o) \) and \( \mu^o_{i\mu}(o) \), as well as their total probabilities, \( \pi_A \) and \( \pi_B \). This information is also obtained if A and B strongly overlap in the observable.

Other rate estimators that rely on fitting the exponential tail of a time-correlation function calculated from
be made but goes beyond the present study.

The presented idea of building an optimal estimator for a single relaxation rate upon the transition matrix estimate of the projected lowest eigenfunction, $\psi_2$, is readily extensible to multiple relaxation rates. This lays the basis for a multi-state rate theory for low-dimensional observations of dynamical systems that will be pursued in future studies. Such a theory will help us to understand and enhance the behavior of existing data analysis methods such as Hidden Markov models when applied to observed physical systems.

Acknowledgements

The authors thank Susan Marqusee and Phillip J. Elms (UC Berkeley) for sharing the single-molecule force probe data. FN and JHP acknowledge funding from the DFG center MATHEON. JDC acknowledges funding from a QB3-Berkeley Distinguished Postdoctoral Fel-
We are grateful to Christof Schütte, Attila Szabo, Sergio Bacallado, Vijay Pande and Heidrun Prantel for enlightening discussions and support.

[1] William J. Greenleaf, Michael T. Woodside, and Steven M. Block. High-Resolution, Single-Molecule Measurements of Biomolecular Motion. *Annual Review of Biophysics and Biomolecular Structure*, 36(1):171–190, 2007.

[2] Lisa J. Lapidus, William A. Eaton, and James Hofrichter. Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl. Acad. Sci. USA*, 97(13):7220–7225, June 2000.

[3] Huan-Xiang Zhou. Rate theories for biologists. *Quart. Rev. Biophys.*, 43, 2010.

[4] Olga K. Dudko, Thomas G. W. Graham, and Robert B. Best. Locating the barrier for folding of single molecules under an external force. *Phys. Rev. Lett.*, 107:208301, 2011.

[5] Greg Morrison, Changbong Hyeon, Michael Hinczewski, and D. Thirumalai. Compaction and tensile forces determine the accuracy of folding landscape parameters from single molecule pulling experiments. *Phys. Rev. Lett.*, 106:138102, 2011.

[6] H. Eyring. The activated complex in chemical reactions. *J. Chem. Phys.*, 3:107–115, 1935.

[7] David Chandler. Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *The Journal of Chemical Physics*, 68(6):2959–2970, 1978.

[8] John D. Chodera, Phillip J. Elms, William C. Swope, Jan-Hendrik Prinz, Susan Marquese, Carlos Bustamante, Frank Noé, and Vijay S. Pande. A robust approach to estimating rates from time-correlation functions. [http://arxiv.org/abs/1108.2304](http://arxiv.org/abs/1108.2304), 2011.

[9] Elliot L. Elson and Douglas Magde. Fluorescence Correlation Spectroscopy. I. Conceptual Basis and Theory. *Biopolymers*, 13:1–27, 1974.

[10] James L. Skinner and Peter G. Wolynes. Relaxation processes and chemical kinetics. *J. Chem. Phys.*, 69:2143+, 1978.

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

[12] J. D. Chodera, K. A. Dill, N. Singhal, V. S. Pande, W. C. Swope, and J. W. Pitera. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.*, 126:155101, 2007.

[13] Frank Noé, Ilia Horenko, Christof Schütte, and Jeremy C. Smith. Hierarchical Analysis of Conformational Dyanmics in Biomolecules: Transition Networks of Metastable States. *J. Chem. Phys.*, 126:155102, 2007.

[14] Jan-Hendrik Prinz, Hao Wu, Marco Sarich, Bettina Keller, Martin Fischbach, Martin Held, John D. Chodera, Christof Schütte, and Frank Noé. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.*, 134:174105, 2011.

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

[16] Marco Sarich, Frank Noé, and Christof Schütte. On the approximation error of markov state models. *SIAM Multiscale Model. Simul.*, 8:1154–1177, 2010.

[17] Susanna Kube and Marcus Weber. A coarse graining method for the identification of transition rates between molecular conformations. *J. Chem. Phys.*, 126(2):024103+, 2007.

[18] Nicoalae V. Buchete and Gerhard Hummer. Coarse Master Equations for Peptide Folding Dynamics. *J. Phys. Chem. B*, 112:6057–6069, 2008.

[19] C. Schütte, F. Noé, J. Lu, M. Sarich, and E. Vanden-Eijnden. Markov state models based on milestoneing. *J. Chem. Phys.*, 134:204105, 2011.

[20] Ginka S. Buchner, Ronan D. Murphy, Nicolae-Viorel Buchete, and Jan Kubelka. Dynamics of protein folding: Probing the kinetic network of folding–unfolding transitions with experiment and theory. *Biochimica et Biophysica Acta*, 1814:1001–1020, 2011.

[21] Lawrence R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition, pages 257–286. 1989.

[22] Sean A. McKinney, Chirilmin Joo, and Taekjip Ha. Analysis of Single-Molecule FRET Trajectories Using Hidden Markov Modeling. *Biophysical Journal*, 91(5):1941–1951, September 2006.

[23] John D. Chodera, Phillip Elms, Frank Noé, Bettina Keller, Christian M. Kaiser, Aaron Ewall-Wice, Susan Marquese, Carlos Bustamante, and Nina Singhal Hinrichs. Bayesian hidden markov model analysis of single-molecule force spectroscopy: Characterizing kinetics under measurement uncertainty. [http://arxiv.org/abs/1108.1430](http://arxiv.org/abs/1108.1430), 2011.

[24] Irina V. Gopich and Attila Szabo. Decoding the pattern of photon colors in single-molecule FRET. *The journal of physical chemistry B*, 113(31):10965–10973, August 2009.

[25] Johannes Stigler, Fabian Ziegler, Anja Gieseke, J. Christof M. Gebhardt, and Matthias Rief. The complex folding network of single calmodulin molecules. *Science*, 334:512–516, 2011.

[26] Menahem Pirchi, Guy Ziv, Inbal Riven, Sharona Sedghani Cohen, Nir Zohar, Yoav Barak, and Gilad Haran. Single-molecule fluorescence spectroscopy maps the folding landscape of a large protein. *Nature Comms.*, 2:493, 2011.

[27] Armin Hoffmann and Michael T. Woodside. Signal-pair correlation analysis of single-molecule trajectories. *Angew. Chem. Int. Ed.*, 50:12643–12646, 2011.

[28] C. Schütte, A. Fischer, W. Huisingsa, and P. Deuflhard. A Direct Approach to Conformational Dynamics based on Hybrid Monte Carlo. *J. Comput. Phys.*, 151:146–168, 1999.

[29] Frank Noé and Feliks Nüsse. A variational approach to modeling slow processes in stochastic dynamical systems. *ArXiv e-prints*, November 2012.

[30] M. Hinczewski, Y. von Hansen, and R.R. Netz. Deconvolution of dynamic mechanical networks. *Proc. Natl. Acad. Sci. USA*, 107:21493–21498, 2010.

[31] Yann von Hansen, Alexander Mehlich, Benjamin Pelz, Matthias Rief, and Roland R. Netz. Auto- and cross-power spectral analysis of dual trap optical tweezer experiments using bayesian inference. *Rev. Sci. Instrum.*, 83:095116, 2012.
[32] L. Onsager. Reciprocal relations in irreversible processes II. *Phys. Rev.*, 38:2265–2279, 1931.

[33] John A. Montgomery, David Chandler, and Bruce J. Berne. Trajectory analysis of a kinetic theory for isomerization dynamics in condensed phases. *The Journal of Chemical Physics*, 70(9):4056–4066, 1979.

[34] Robert Zwanzig and Narinder K. Ailawadi. Statistical Error Due to Finite Time Averaging in Computer Experiments. *Physical Review Online Archive (Prola)*, 182:280–283, June 1969.

[35] W. Ritz. Über eine neue methode zur lösung gewisser variationsprobleme der mathematischen physik. *J. Reine Angew. Math.*, 135:1–61, 1909.

[36] Baron Peters. Using the histogram test to quantify reaction coordinate error. *The Journal of Chemical Physics*, 125(24):241101+, 2006.

[37] Frank Noé, Christof Schütte, Eric Vanden-Eijnden, Lothar Reich, and Thomas R. Weikl. Constructing the full ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. USA*, 106:19011–19016, 2009.

[38] Vincent A. Voelz, Gregory R. Bowman, Kyle Beauchamp, and Vijay S. Pande. Molecular Simulation of ab Initio Protein Folding for a Millisecond Folder NTL9. *J. Am. Chem. Soc.*, 132(5):1526–1528, February 2010.

[39] Gregory R. Bowman, Kyle A. Beauchamp, George Boxer, and Vijay S. Pande. Progress and challenges in the automated construction of Markov state models for full protein systems. *J. Chem. Phys.*, 131(12):124101+, September 2009.

[40] P. Deuflhard and M. Weber. Robust Perron cluster analysis in conformation dynamics. *ZIB Report*, 03-09, 2003.

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

[42] Phillip J. Elms, John D. Chodera, Carlos Bustamante, and Susan Marqusee. The limitations of constant-force feedback experiments. *Biophys. J.*, 103:1490, 2012.

[43] Phillip J. Elms, John D. Chodera, Carlos Bustamante, and Susan Marqusee. The molten globule state is unusually deformable under mechanical force. *Proc. Natl. Acad. Sci. USA*, 109:3796, 2012.

[44] Carlos J. Bustamante and Steven B. Smith. Light-force sensor and method for measuring axial optical-trap forces from changes in light momentum along an optic axis, 2006.

[45] Phillip J. Elms, John D. Chodera, Carlos Bustamante, and Susan Marqusee. The molten globule state is unusually deformable under mechanical force. *Proc. Natl. Acad. Sci. USA*, 109:3796–3801, 2012.