Dynamic neutron scattering from conformational dynamics. I. Theory and Markov models

Benjamin Lindner, Zheng Yi, Jan-Hendrik Prinz, Jeremy C. Smith, and Frank Noé

Citation: J. Chem. Phys. **139**, 175101 (2013); doi: 10.1063/1.4824070
View online: http://dx.doi.org/10.1063/1.4824070
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v139/i17
Published by the AIP Publishing LLC.

Dynamic neutron scattering from conformational dynamics. I. Theory and Markov models

Benjamin Lindner,[1] Zheng Yi,[1] Jan-Hendrik Prinz,[2] Jeremy C. Smith,[1] and Frank Noé[2]
[1]*University of Tennessee/Oak Ridge National Laboratory Center for Molecular Biophysics, P.O. Box 2008, 1 Bethel Valley Rd., Oak Ridge, Tennessee 37831, USA and Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, M407 Walters Life Sciences, 1414 Cumberland Avenue, Knoxville, Tennessee 37996, USA*
[2]*Department of Mathematics and Computer Science, FU Berlin, Arnimallee 6, 14159 Berlin, Germany*

(Received 5 June 2013; accepted 10 July 2013; published online 1 November 2013)

The dynamics of complex molecules can be directly probed by inelastic neutron scattering experiments. However, many of the underlying dynamical processes may exist on similar timescales, which makes it difficult to assign processes seen experimentally to specific structural rearrangements. Here, we show how Markov models can be used to connect structural changes observed in molecular dynamics simulation directly to the relaxation processes probed by scattering experiments. For this, a conformational dynamics theory of dynamical neutron and X-ray scattering is developed, following our previous approach for computing *dynamical fingerprints* of time-correlation functions [F. Noé, S. Doose, I. Daidone, M. Löllmann, J. Chodera, M. Sauer, and J. Smith, Proc. Natl. Acad. Sci. U.S.A. **108**, 4822 (2011)]. Markov modeling is used to approximate the relaxation processes and timescales of the molecule via the eigenvectors and eigenvalues of a transition matrix between conformational substates. This procedure allows the establishment of a complete set of exponential decay functions and a full decomposition into the individual contributions, i.e., the contribution of every atom and dynamical process to each experimental relaxation process. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4824070]

I. INTRODUCTION

Neutron spectroscopy (dynamic neutron scattering) probes time correlations on the sub pico- to microsecond timescales (0.1–10^6 ps) and is sensitive to motions on the atomic, molecular, and intermolecular length scales (1–10^3 Å). This kind of spectroscopy has been extensively used to investigate the dynamics of a variety of soft-matter molecular and biomolecular systems, including polymers, proteins, and membranes.[2–8] However, the interpretation of neutron spectroscopic data from complex molecular systems can be difficult owing to the large number of dynamical processes contributing to the underlying correlation functions. In the absence of a theoretical model, experimentalists are limited to fitting the data with an arbitrary set of exponential decay functions or an equivalent representation such as a stretched exponential, leading to a description in terms of a set of poorly characterized relaxation processes with characteristic timescales.

Molecular dynamics (MD) simulation can improve data interpretation, because neutron scattering functions can be directly computed from the MD trajectories.[9–12] This guides the process of functional fitting by providing an *"ab initio"* derived set of conformational changes or structural processes and allowing the decomposition of spectral data into its various components that are distinguishable, such as rotation, diffusion, and intramolecular contributions.[13,14] However, there are two practical limitations. First of all, neutron scattering signals are ensemble averages, and thus their computation is limited by the sampling problem. In order to generate an unbiased ensemble for calculating the time correlation function probed by neutron scattering, a direct MD simulation needs to be a multiple of the slowest detectable relaxation time in the system. Second, even when the experimentally measurable signal can be computed from MD simulation trajectories, the interpretation of the signal is unclear. The time correlation functions probed by neutron scattering often involve a variety of relaxation timescales that are characteristic of the molecule, but which structural processes are these relaxation timescales assigned to?

Both the above difficulties can be overcome by combining Markov models[13,15–18] with dynamical fingerprint analysis.[1,19] In the Markov modeling approach, one discretizes the conformation space of the molecule into small sets (microstates), and uses MD simulation data to compute transition probabilities between microstates. An advantage of this approach is that slow kinetics can be computed with short trajectories and thus even systems with relaxation times longer than the affordable trajectory length are within reach.[1,20–22] Markov models have been applied to peptide dynamics,[18,23–25] protein folding,[1,20,21,26] native-state protein dynamics,[1,20–22,27] and protein-ligand binding.[28,29] See Refs. 30–32 for a review of Markov models,[33] for an overview of the methodology and[34,35] for software to construct Markov models from MD simulation data.

Dynamical fingerprints[1,19] is a theory of computational spectroscopy. The theory provides a way to compute a spectroscopic signal that probes the equilibrium kinetic relaxations of a molecule (such as a time-correlation function) in terms of the eigenvalues and eigenvectors of the Markov

model transition matrix. In contrast to directly computing the time-correlation function from the MD trajectory, this approach yields a direct mapping between the experimentally observable relaxation timescales and amplitudes and an explicit description of the structural transitions involved. Thus, one is able to fully interpret the dynamical features observed in a spectroscopic signal by assigning structural transitions to them. Furthermore, dynamical fingerprint theory also provides a way of designing experiments so as to optimally probe desired individual relaxation processes.[1]

Here we generalize and extend the dynamical fingerprint theory in two ways. First, we provide a formulation of the dynamical fingerprint theory in terms of the exact conformational dynamics of the molecular system studied. This way, most of the equations given here are generally valid, and are also applicable if approaches other than Markov models are used to approximate the molecular kinetics. Markov models are introduced only at the end, in order to approximate the equations derived in the theory part, and the discretization errors introduced by Markov modeling are discussed. Second, we reformulate the quantities of dynamical scattering—the intermediate scattering function, the dynamic structure factor, and the dynamic susceptibility—in terms of conformational dynamics. This results in a one-to-one relationship between a relaxation process of the neutron spectrum and the corresponding relaxation process computable by a Markov model. Besides neutron scattering, which we will focus on, the approach is also applicable to dynamic (inelastic) X-ray scattering.

In Paper II,[36] the theoretical approach developed here is illustrated by reconciling MD simulations of the well-studied alanine dipeptide[37–42] with neutron scattering data via a Markov model. The paper demonstrates how experimentally observable processes can be assigned to individual structural transitions, thus opening the door to studying more complex molecular systems in detail.

II. X-RAY AND NEUTRON SCATTERING

Although dynamic neutron scattering experiments are more commonly performed than dynamic X-ray scattering experiments, we emphasize here that the formalism is valid for both.

A. Intermediate scattering function $F(\mathbf{q}, \omega)$

Dynamic X-ray and neutron scattering experiments usually measure the dynamic structure factor, $S(\mathbf{q}, \omega)$, where $\mathbf{q}$ is the momentum exchange and ω is related to the energy transfer $E = \hbar\omega$. $S(\mathbf{q}, \omega)$ is the time-frequency Fourier transform of the non-normalized intermediate scattering function, $F(\mathbf{q}, \omega)$,[2,43] which can be decomposed into an incoherent and a coherent scattering function:

$$\begin{aligned} F(\mathbf{q}, \tau) &= F_{\text{inc}}(\mathbf{q}, \tau) + F_{\text{coh}}(\mathbf{q}, \tau), \\ F_{\text{inc}}(\mathbf{q}, \tau) &= \sum_{\alpha} b_{\alpha,coh} b^{*}_{\alpha,\text{inc}} \langle \mathrm{e}^{-i\langle \mathbf{q}, \mathbf{r}_{\alpha}(t)\rangle} \mathrm{e}^{i\langle \mathbf{q}, \mathbf{r}_{\alpha}(t+\tau)\rangle} \rangle_{t}, \\ F_{\text{coh}}(\mathbf{q}, \tau) &= \sum_{\alpha,\beta} b_{\alpha,coh} b^{*}_{\beta,\text{coh}} \langle \mathrm{e}^{-i\langle \mathbf{q}, \mathbf{r}_{\alpha}(t)\rangle} \mathrm{e}^{i\langle \mathbf{q}, \mathbf{r}_{\beta}(t+\tau)\rangle} \rangle_{t}, \end{aligned} \quad (1)$$

where $\langle \mathbf{q}, \mathbf{r}_{\alpha}(t)\rangle$ is the scalar product between the scattering vector, $\mathbf{q}$, and the time-dependent Cartesian coordinate, $\mathbf{r}_{\alpha}(t)$, for atom α. $\langle \cdot \rangle_{t}$ denotes an average over time t which, for an ergodic system, is equivalent to the ensemble average. The atomic prefactor, b_{α}, is different for X-ray and neutron scattering and also different for coherent and incoherent scattering.

In neutron scattering b_{α} is the atomic scattering length, which is different for each isotope and independent of $|\mathbf{q}|$. The variation of the atomic scattering length due to isotopic distribution and spin orientation gives rise to coherent and incoherent scattering, which is described by two distinct scattering lengths for each isotope type, $b_{\text{neutron, coh}}$ and $b_{\text{neutron, inc}}$.[44] Coherent scattering arises from the constructive interference of the atomic scattering amplitudes from all atoms, and contains information on interatomic distances and motions, while incoherent scattering is due to self-interference of atomic scattering amplitudes, and contains information on the dynamic properties of individual atoms. The intermediate scattering function, $F(\mathbf{q}, \tau)$, can be directly calculated from MD trajectories by using Eq. (1). α and β enumerate individual atoms whose positions are specified by their time-dependent position vectors, $\mathbf{r}_{\alpha}(t)$ and $\mathbf{r}_{\beta}(t)$, respectively.

In X-ray scattering b_{α} represents the form factor of the respective atom and can be approximated by a series of Gaussians:[45]

$$b_{\text{x-ray,coh}}(\mathbf{q}) = \sum_{i} c_i \mathrm{e}^{-d_i |\mathbf{q}|^2},$$

where the set of c_i and d_i are empirically derived constants.

If the scattering particles in the experimental sample have no preferred orientation, then the intermediate scattering function must be orientationally averaged:

$$F(q, \tau) = \langle F(\mathbf{q}, \tau) \rangle_{\mathbf{q}}, \quad (2)$$

where $q = |\mathbf{q}|$ and $\langle \cdot \rangle_{\mathbf{q}}$ denotes the average overall scattering vectors with fixed length q. Note that we do not treat the effect of rigid-body translation or rotation on the dynamical scattering signal here. Thus, we assume that the sample is in a dense medium such as a powder, where diffusion is inhibited, or that rigid-body motions have been subtracted.

B. Dynamic structure factor $S(\mathbf{q}, \omega)$ and dynamic susceptibility $\chi''(\mathbf{q}, \omega)$

Dynamic X-ray and neutron scattering experiments usually measure the dynamic structure factor, i.e., the Fourier transform of $F(\mathbf{q}, \tau)$:

$$S(\mathbf{q}, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\mathbf{q}, \tau) \mathrm{e}^{-i\omega\tau} d\tau. \quad (3)$$

Both X-ray and coherent neutron scattering experiments commonly yield the coherent dynamic structure factor $S_{\text{coh}}(\mathbf{q}, \omega)$, while incoherent neutron scattering provides $S_{\text{inc}}(\mathbf{q}, \omega)$:

$$S_{\text{inc}}(\mathbf{q}, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_{\text{inc}}(\mathbf{q}, \tau) \mathrm{e}^{-i\omega\tau} d\tau, \quad (4)$$

$$S_{\mathrm{coh}}(\mathbf{q},\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F_{\mathrm{coh}}(\mathbf{q},\tau)\mathrm{e}^{-i\omega\tau}d\tau. \quad (5)$$

$S_{\mathrm{inc}}(\mathbf{q},\omega)$ can also be used to calculate the dynamic susceptibility $\chi''_{\mathrm{inc}}(\mathbf{q},\omega)$ for analyzing inelastic neutron scattering spectra as

$$\chi''_{\mathrm{inc}}(\mathbf{q},\omega) = \frac{S_{\mathrm{inc}}(\mathbf{q},\omega)}{n_B(\omega)+1}, \quad (6)$$

where $n_B(\omega) = (\mathrm{e}^{\hbar\omega/kT}-1)^{-1}$ (for neutron energy loss) is the Bose temperature factor. The dynamic susceptibility allows the characteristic relaxation time, t_r, of a molecular structural process to be estimated from a peak, ω_{max}, in the frequency spectrum[46] (see below for details).

C. Instrumental resolution and corrections

At finite instrumental energy resolution, $\Delta\omega$, the experimentally accessible scattering function in ω-space is the convolution of $S(\mathbf{q},\omega)$ with the instrumental resolution function $R(\omega;\Delta\omega)$:

$$\begin{aligned} S(\mathbf{q},\omega;\Delta\omega) &= S(\mathbf{q},\omega)\otimes R(\omega;\Delta\omega) \\ &= \int_{-\infty}^{\infty} S(\mathbf{q},\omega-\omega')\,R(\omega';\Delta\omega)\,d\omega'. \end{aligned}$$

In order to mimic a typical experimental resolution function, here the computed $S(\mathbf{q},\omega)$ is convoluted with a Gaussian function

$$S(\mathbf{q},\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\mathbf{q},\tau)\mathrm{e}^{-i\omega\tau}\mathrm{e}^{-\frac{\tau^2}{2c^2}}d\tau,$$

where c determines the width of the Gaussian function. For a given instrument, only frequencies that are significantly higher than c^{-1} are clearly observable.

III. CONFORMATIONAL DYNAMICS AND DYNAMICAL SCATTERING

A. Conformational dynamics

We briefly sketch the main mathematical ideas underlying the general formalism for describing conformational dynamics in terms of the states. The mathematical groundwork for this was laid in Refs. 47 and 48, and the current conformational dynamics method has been formalized by Schütte and co-workers.[15]

It is assumed that the molecular system studied lives in a continuous state space Ω consisting of positions and momenta. Its time evolution $\mathbf{x}(t)$ obeys the following properties:

1. $\mathbf{x}(t)$ is a Markov process in the full state space Ω, i.e., the instantaneous change in $\mathbf{x}$ only depends on the current value of $\mathbf{x}$ and not its history.
2. $\mathbf{x}(t)$ is ergodic, i.e., all states of Ω can be reached in an infinitely long trajectory and are visited with a probability given by the Boltzmann distribution:

$$\mu(\mathbf{x}) = Z(\beta)^{-1}\exp\left(-\beta H(\mathbf{x})\right). \quad (7)$$

3. $\mathbf{x}(t)$ is reversible, i.e., the conditional probability density of being in state $\mathbf{y}$ in time $t+\tau$ given that the system is in state $\mathbf{x}$ at time t, $p(\mathbf{x},\mathbf{y};\tau)$, fulfills the condition of *detailed balance*:

$$\mu(\mathbf{x})\,p(\mathbf{x},\mathbf{y};\tau) = \mu(\mathbf{y})\,p(\mathbf{y},\mathbf{x};\tau). \quad (8)$$

These conditions are fulfilled by many dynamical models frequently used to simulate molecular dynamics, such as properly thermostatted dynamics or Hybrid Monte Carlo. Langevin dynamics and several other thermostatted dynamics do not obey Eq. (8) but instead obey a generalized detailed balance condition. Generalized detailed balance is also suitable for conformation dynamics, but leads to a complication of the formalism that we will not pursue here.

We can then perform the following, at this stage purely formal trick and describe the evolution of the dynamics in terms of an ensemble distribution $\rho_t(\mathbf{x})$:

$$\rho_{t+\tau}(\mathbf{y}) = \mathcal{P}(\tau)\rho_t(\mathbf{x}) = \int d\mathbf{x}\,\rho_t(\mathbf{x})p(\mathbf{x},\mathbf{y};\tau). \quad (9)$$

This means that when $\rho_t(\mathbf{x})$ is the probability distribution of molecules in the ensemble at time t, then $\rho_{t+\tau}(\mathbf{x})$ is the probability distribution of the ensemble at a later time $t+\tau$. The evolution of the probability density is described by the operator $\mathcal{P}(\tau)$. The important fact of this equation is that the same operator $\mathcal{P}(\tau)$ holds at all times t and that it is a linear operator, i.e., a mathematically simple object, which allows us to propagate to arbitrarily long times by repeated usage:

$$\rho_{t+\tau_1+\tau_2} = \mathcal{P}(\tau_1)\mathcal{P}(\tau_2)\rho_t = \mathcal{P}(\tau_1+\tau_2)\rho_t \quad (10)$$

and so on. The purpose of Markov models is to discretize state space Ω such that $\mathcal{P}(\tau)$ can be approximated by a matrix and $\rho_t(\mathbf{x})$ can be approximated by a vector, such that the equations above are well approximated. However, this discretization involves an approximation error that will be discussed later on. In this section, we will concentrate on the exact propagator and its properties, and then express the dynamical scattering quantities in terms of the Markov propagator.

Figure 1 (taken from Ref. 33) illustrates how $\mathcal{P}(\tau)$ operates on a one-dimensional example. Figure 1(a) shows a potential energy landscape with associated Boltzmann density $\mu(\mathbf{x})$. Figure 1(b) is an illustration of the transition density $p(\mathbf{x},\mathbf{y};\tau)$ governing the operator $\mathcal{P}(\tau)$: the horizontal and vertical axes correspond to the coordinate $\mathbf{x}$ and the color coding quantifies how much probability density is transported between two points $\mathbf{x}$ in a time τ. The dark color blocks near the diagonal correspond to the fact that there is a high probability of moving around within an energy basin, while the white colors in off-diagonal regions correspond to the fact that there is a small probability of jumping between the basins.

For the present paper, the essential properties of $\mathcal{P}(\tau)$ are its eigenvalues and eigenfunctions. $\mathcal{P}$ fulfills the following eigenvalue equations:

$$\mathcal{P}(\tau)\phi_i = \lambda_i(\tau)\phi_i. \quad (11)$$

We sort eigenvalues by decreasing magnitude, and find that the largest eigenvalue is $\lambda_1(\tau) = 1$, and it is dominant and unique. It is easy to see that its eigenfunction ϕ_1, when

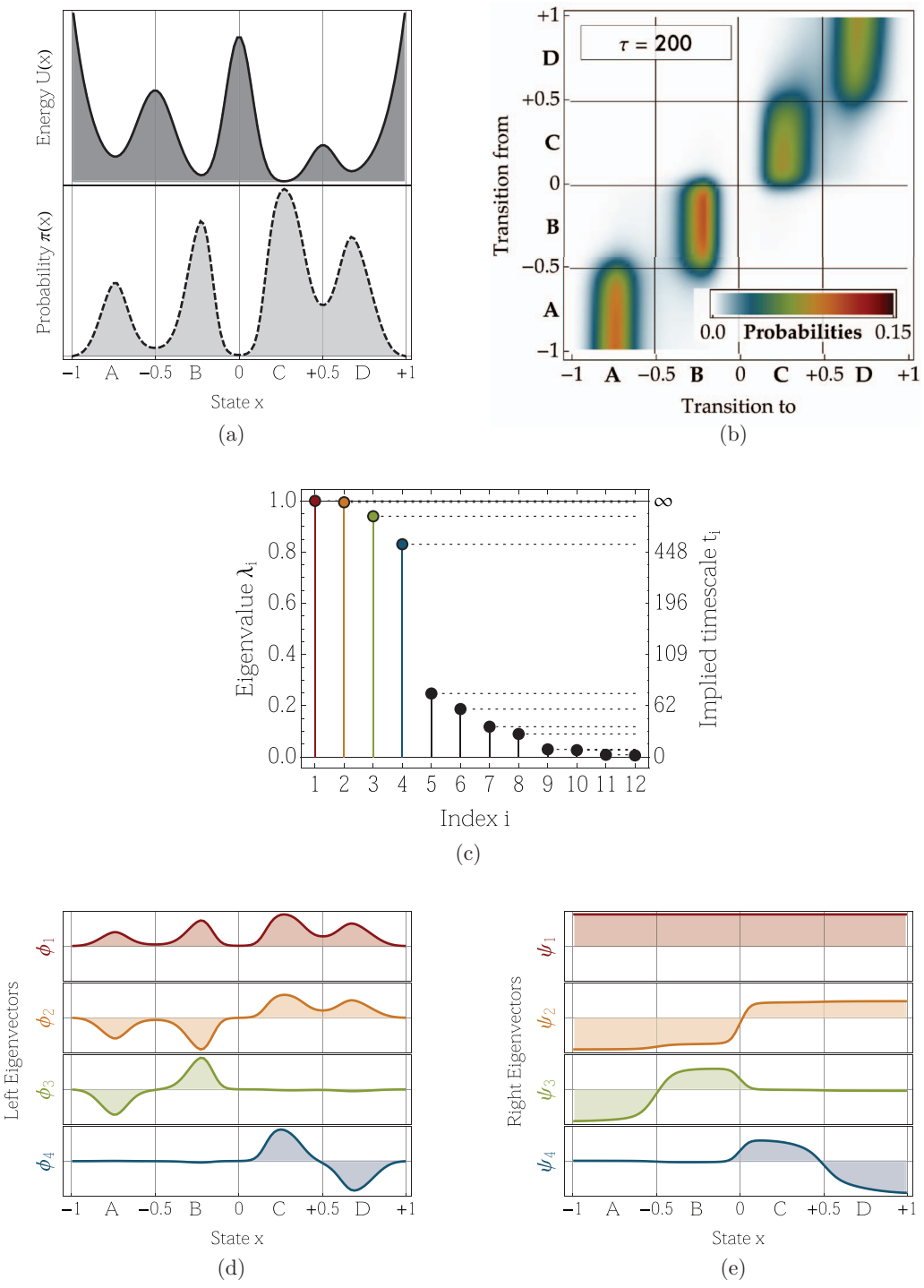

FIG. 1. (a) Potential energy function with four metastable states and corresponding stationary density $\mu(\mathbf{x})$. (b) Density plot of the transfer operator for a simple diffusion-in-potential dynamics defined in the range $\Omega = [-1, 1]$ (see supplementary material of Ref. 33), black and red colors indicate a high transition probability, and white indicates vanishing 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) Eigenvalues of the transfer operator. The gap between the four metastable processes ($\lambda_i \approx 1$) and the fast processes is clearly visible. (d) The four dominant eigenfunctions of the operator $\mathcal{P}(\tau)$, $\phi_1, \ldots, \phi_4$, which indicate the associated dynamical processes. The first eigenfunction is associated with 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. (e) The eigenfunctions weighted with the $\mu(\mathbf{x})^{-1}$. Reprinted with permission from J.-H. Prinz *et al.*, J. Chem. Phys. **134**, 174105 (2011). Copyright 2011 American Institute of Physics.

properly normalized, corresponds to the stationary density $\phi_1 = \mu$

$$\mathcal{P}(\tau)\mu = \mu.$$

All other eigenvalues have norms smaller than 1 and their eigenfunctions correspond to dynamical relaxation processes. Assuming the Markov property (discussed in Eq. (34)) and applying it towards Eq. (11) one can show that the eigenvalues $\lambda_i(\tau)$ decay exponentially in τ, and thus each eigenpair

can be associated with a relaxation timescale t_i, or a relaxation rate κ_i, via

$$\lambda_i(\tau) = \mathrm{e}^{-\tau/t_i} = \mathrm{e}^{-\kappa_i \tau}. \tag{12}$$

The smaller the eigenvalue λ_i, the smaller the relaxation timescale t_i; the larger the relaxation rate κ_i, the faster the corresponding process decays.

To understand the interplay of multiple relevant eigenvalues and eigenvectors let us review again Fig. 1 which shows the diffusion dynamics on an energy landscape with four basins (A, B, C, D) and high intervening energy barriers. Figure 1(c) shows the 12 largest eigenvalues of the transition matrix in Fig. 1(b). There is one eigenvalue, λ_1, which is equal to one, followed by three eigenvalues, λ_2 to λ_4, which are close to one. These four *dominant eigenvalues*are separated by a gap from the remaining eigenvalues. Hence, the transition matrix consists of a stationary process, three slow processes and many processes which decay quickly. After a few time steps, only the four dominant processes contribute to the evolution of the probability vector. How these processes alter this vector is determined by the shape of the corresponding eigenvectors.

In addition to the eigenfunctions ϕ_i, we will sometimes need the weighted eigenvectors ψ_i. ϕ_i and ψ_i are related by

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

ψ_i can also be understood as the eigenfunctions of the transfer operator, which is a weighted version of the propagator.[15]

Figures 1(d) and 1(e) show the four dominant eigenfunctions (both ϕ_i and the weighted version ψ_i). The first eigenfunction corresponds to the stationary process and is, therefore, either the stationary distribution (ϕ_1), or constant (ψ_1). The second eigenvector corresponds to the slowest process and has positive signs in regions A and B and negative signs in regions C and D. This shape effectively moves probability density across the largest barrier in the energy surface. Since the eigenvector is approximately constant within the combined region (A,B) and (C,D) left and right of the barrier, it does not alter the relative probability distribution within these regions. The third eigenvector, analogously, moves density between A and B, the fourth between C and D.

The transition density $p(\mathbf{x}, \mathbf{y}; \tau)$ can be represented as an expansion of eigenfunctions, using the eigenvalues as coefficients. Here we directly give the expansion of the correlation density, $c(\mathbf{x}, \mathbf{y}; \tau) = \mu(\mathbf{x})p(\mathbf{x}, \mathbf{y}; \tau)$, which is the joint probability of finding the system in state $\mathbf{x}$ at time t and in state $\mathbf{y}$ at time $t + \tau$:

$$c(\mathbf{x}, \mathbf{y}; \tau) = \sum_{k=1}^{\infty} \lambda_k(\tau)\phi_k(\mathbf{x})\phi_k(\mathbf{y}). \tag{13}$$

The intermediate scattering functions given in Eq. (1) are time correlation functions. In order to rewrite the dynamical scattering quantities such that they are computable with Markov models, it is essential that we will be able to represent time correlation functions in terms of conformational dynamics quantities. Generally, a time-correlation function between observables $f(\mathbf{x})$ and $g(\mathbf{x})$ is given by

$$\langle f_t g_{t+\tau} \rangle_t = \lim_{T\to\infty} \frac{1}{T-\tau} \int_{t=0}^{T-\tau} f(\mathbf{x}(t)) f(\mathbf{x}(t+\tau)),$$

and, using ergodicity, this can be rewritten in terms of a double integral over the state at time t and the state at time $t + \tau$, using the correlation density as a probability weight. Employing the spectral decomposition (13), we arrive at

$$\begin{aligned}\langle f_t g_{t+\tau} \rangle_t &= \int d\mathbf{x} \int d\mathbf{y}\, f(\mathbf{x}) c(\mathbf{x}, \mathbf{y}; \tau) g(\mathbf{y}) \\ &= \sum_{k=1}^{\infty} \lambda_k(\tau) \int d\mathbf{x} \int d\mathbf{y}\, f(\mathbf{x})\phi_k(\mathbf{x})\phi_k(\mathbf{y}) g(\mathbf{y}) \\ &= \sum_{k=1}^{\infty} \lambda_k(\tau) \langle f, \phi_k \rangle \langle g, \phi_k \rangle. \end{aligned} \tag{14}$$

B. Conformational dynamics representation of the intermediate scattering function $F(\mathbf{q}, \tau)$

The fundamental observable for scattering experiments is described by the atomic scattering amplitude, $a(q, x)$:

$$a_\alpha(\mathbf{q}, \mathbf{x}) = b_\alpha \mathrm{e}^{-i\langle \mathbf{q}, \mathbf{r}_\alpha(\mathbf{x}) \rangle}$$

in which the time-dependent Cartesian coordinate, $r_\alpha(t)$, for atom α is expressed as a state-dependent Cartesian coordinate, $r_\alpha(x)$, which requires the description in terms of a state space. Mapping $a_\alpha(\mathbf{q}, \mathbf{x})$ onto the state vector $\phi_k(\mathbf{x})$, leads to quantities describing the scattering amplitude of atom α in relaxation process k:

$$a_{k\alpha}(\mathbf{q}) = \langle a_\alpha(\mathbf{q}, \mathbf{x}), \phi_k(\mathbf{x}) \rangle. \tag{15}$$

The incoherent intermediate scattering function (Row 2 in Eq. (1)) can be rewritten using Eqs. (14) and (15) as

$$\begin{aligned} F_{\mathrm{inc}}(\mathbf{q}, \tau) &= \sum_\alpha b_{\alpha,\mathrm{inc}}^2 \langle \mathrm{e}^{-i\langle \mathbf{q}, \mathbf{r}_\alpha(t) \rangle} \mathrm{e}^{i\langle \mathbf{q}, \mathbf{r}_\alpha(t+\tau) \rangle} \rangle_t \\ &= \sum_{k=1}^{\infty} \mathrm{e}^{-\kappa_k \tau} \sum_\alpha \langle b_{\alpha,\mathrm{inc}} \mathrm{e}^{-i\langle \mathbf{q}, \mathbf{r}_\alpha(\mathbf{x}) \rangle}, \phi_k(\mathbf{x}) \rangle \\ &\quad \times \langle b^*_{\alpha,\mathrm{inc}} \mathrm{e}^{i\langle \mathbf{q}, \mathbf{r}_\alpha(\mathbf{x}) \rangle}, \phi_k(\mathbf{x}) \rangle \\ &= \sum_{k=1}^{\infty} \mathrm{e}^{-\kappa_k \tau} \sum_\alpha a_{k\alpha}^{inc}(\mathbf{q}) a_{k\alpha}^{inc*}(\mathbf{q}) \\ &= \sum_{k=1}^{\infty} \mathrm{e}^{-\kappa_k \tau} A_k^{\mathrm{inc}}(\mathbf{q}) \end{aligned} \tag{16}$$

with

$$A_k^{\mathrm{inc}}(\mathbf{q}) = \sum_\alpha a_{k\alpha}^{inc}(\mathbf{q}) a_{k\alpha}^{inc*}(\mathbf{q}) = \sum_\alpha \left| a_{k\alpha}^{inc}(\mathbf{q}) \right|^2. \tag{17}$$

For the coherent case, we analogously obtain

$$\begin{aligned} F_{\mathrm{coh}}(\mathbf{q},\tau) &= \sum_{k=1}^{\infty} \mathrm{e}^{-\kappa_k \tau} \sum_{\alpha,\beta} \langle b_{\alpha,coh} \mathrm{e}^{-i\langle \mathbf{q},\mathbf{r}_\alpha(\mathbf{x})\rangle}, \phi_k(\mathbf{x})\rangle \\ &\quad \times \langle b^*_{\beta,coh} \mathrm{e}^{i\langle \mathbf{q},\mathbf{r}_\beta(\mathbf{x})\rangle}, \phi_k(\mathbf{x})\rangle \\ &= \sum_{k=1}^{\infty} \mathrm{e}^{-\kappa_k \tau} \left(\sum_{\alpha} \langle b_{\alpha,coh} \mathrm{e}^{-i\langle \mathbf{q},\mathbf{r}_\alpha(\mathbf{x})\rangle}, \phi_k(\mathbf{x})\rangle \right) \\ &\quad \times \left(\sum_{\beta} \langle b^*_{\beta,coh} \mathrm{e}^{i\langle \mathbf{q},\mathbf{r}_\beta(\mathbf{x})\rangle}, \phi_k(\mathbf{x})\rangle \right) \\ &= \sum_{k=1}^{\infty} \mathrm{e}^{-\kappa_k \tau} \sum_{\alpha} a^{coh}_{k\alpha}(\mathbf{q}) \sum_{\beta} a^{coh*}_{k\beta}(\mathbf{q}) \\ &= \sum_{k=1}^{\infty} \mathrm{e}^{-\kappa_k \tau} A^{\mathrm{coh}}_k(\mathbf{q}), \end{aligned} \tag{18}$$

where the amplitude of process k in the intermediate scattering function is given by

$$A^{\mathrm{coh}}_k(\mathbf{q}) = \left(\sum_{\alpha} a^{coh}_{k\alpha}(\mathbf{q}) \right) \left(\sum_{\beta} a^{coh}_{k\beta}(\mathbf{q}) \right)^* \tag{19}$$

$$= \sum_{\alpha,\beta} a^{coh}_{k\alpha}(\mathbf{q}) a^{coh*}_{k\beta}(\mathbf{q}). \tag{20}$$

The total intermediate scattering function is given by the sum of incoherent and coherent scattering, with the total amplitudes $A_k(\mathbf{q})$:

$$\begin{aligned} F(\mathbf{q},\tau) &= \sum_{k=1}^{\infty} \mathrm{e}^{-\kappa_k \tau} \left(A^{\mathrm{inc}}_k(\mathbf{q}) + A^{\mathrm{coh}}_k(\mathbf{q}) \right) \\ &= \sum_{k=1}^{\infty} \mathrm{e}^{-\kappa_k \tau} A_k(\mathbf{q}). \end{aligned} \tag{21}$$

The following features of this representation are interesting:

- $F(\mathbf{q},\tau)$ is completely described by a sum of individual exponential decay functions with individual relaxation times t_k or rates κ_k (Eq. (29)).
- The eigenvalues and eigenfunctions in Eq. (21) associated with slow timescales can be approximated with a Markov model analysis of a molecular dynamics simulation (see Sec. IV). The duality of the eigenvalues and eigenfunctions allows us to clearly assign a specific structural conformational transition to each relaxation time, t_k.
- The amplitudes $a_{\alpha k}(\mathbf{q})$ with which each atom α contributes to the visibility in the experiment of process k depends on the overlap of the configuration-dependent scattering intensity $b_\alpha \mathrm{e}^{-i\langle \mathbf{q},\mathbf{r}_\alpha(\mathbf{x})\rangle}$ with the eigenfunction ϕ_k. Thus, if an atom significantly changes its scattering intensity along the conformational transition described by eigenfunction ϕ_k, it will significantly contribute to the amplitude with which this term appears in the multiexponential decay (21).
- The contribution of each atom to any given relaxation process for incoherent scattering can be identified because A^{inc}_k is a simple sum over atoms.
- Since $k = 1$ corresponds to the equilibrium distribution, A^{coh}_1 and A^{inc}_1 instantly provide information about the elastic structure factor (ESF).

If the scattering particles in the experimental sample have no preferred orientation, then the intermediate scattering function must be orientationally averaged according to Eq. (2). This can be carried out at the level of the individual atomistic scattering terms:

$$\bar{A}^{\mathrm{inc}}_k = \sum_{\alpha} \left\langle a^{inc}_{k\alpha}(\mathbf{q}) a^{inc*}_{k\alpha}(\mathbf{q}) \right\rangle_{\mathbf{q}}, \tag{22}$$

$$\bar{A}^{\mathrm{coh}}_k = \sum_{\alpha,\beta} \left\langle a^{coh}_{k\alpha}(\mathbf{q}) a^{coh*}_{k\beta}(\mathbf{q}) \right\rangle_{\mathbf{q}}, \tag{23}$$

$$\bar{A}_k = \bar{A}^{\mathrm{inc}}_k + \bar{A}^{\mathrm{coh}}_k, \tag{24}$$

where $\langle \cdot \rangle_{\mathbf{q}}$ indicates the averaging over orientations. Thus, we obtain the orientationally averaged intermediate scattering functions, expressed in terms of propagator eigenvalues and eigenfunctions:

$$F(\tau) = \sum_{k=1}^{\infty} \mathrm{e}^{-\kappa_k \tau} \bar{A}_k,$$

$$F_{\mathrm{inc}}(\tau) = \sum_{k=1}^{\infty} \mathrm{e}^{-\kappa_k \tau} \bar{A}^{\mathrm{inc}}_k,$$

$$F_{\mathrm{coh}}(\tau) = \sum_{k=1}^{\infty} \mathrm{e}^{-\kappa_k \tau} \bar{A}^{\mathrm{coh}}_k.$$

C. Conformational dynamics representation of the dynamic structure factor $S(\mathbf{q},\omega)$

Since $F(\mathbf{q},\tau)$ can be represented as a sum of exponential decay functions, the dynamic structure factor, $S(\mathbf{q},\omega)$, can also be directly represented by properties of the MSM.

$$\begin{aligned} S(\mathbf{q},\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\tau\, \mathrm{e}^{-i\omega\tau} R(\tau;c) \sum_{k=1}^{\infty} \lambda_k(\tau) A_k(\mathbf{q}) \\ &= \frac{1}{2\pi} A_1(\mathbf{q}) \mathcal{F}\{R(\tau;c)\} \end{aligned} \tag{25}$$

$$+ \frac{1}{2\pi} \sum_{k=2}^{\infty} A_k(\mathbf{q}) \mathcal{F}\{\mathrm{e}^{-\kappa_k \tau} R(\tau;c)\}, \tag{26}$$

where $A_1(\mathbf{q})$ is the elastic component, corresponding to $\lambda_1 = 1$. We use the common Gaussian resolution function $R(\tau,c) = \mathrm{e}^{-\tau^2/c^2}$, obtaining the transformed Gaussian function:

$$\begin{aligned} \mathcal{F}\{R(\tau,c)\} &= \int_{-\infty}^{\infty} R(\tau,c)\, \mathrm{e}^{-i\omega\tau} d\tau \\ &= \sqrt{2\pi}\, c\, \mathrm{e}^{-(c\omega)^2/2}. \end{aligned} \tag{27}$$

Furthermore, we assume that the instrument is capable of well resolving all relaxation processes present. Quantitatively, this means we assume $c \gg t_k$ for all terms $k > 1$, resulting in

$R(\tau = t_k; c) = e^{-t_k^2/c^2} \approx 1$. Therefore, we only need to find the Fourier transform of a single-exponential relaxation for each term $k > 1$, which turns out to be a Lorentzian function:

$$\mathcal{F}\left\{e^{-\tau/t_k}\right\} = \int_{-\infty}^{\infty} e^{-\kappa_k \tau} e^{-i\omega\tau} d\tau = \frac{2\kappa_k}{\omega^2 + \kappa_k^2} \equiv L(\kappa_k, \omega). \quad (28)$$

Using Eqs. (27) and (28) in Eq. (25) results in

$$S(\mathbf{q}, \omega) \approx \sqrt{\frac{c}{2\pi}} A_1(\mathbf{q}) e^{-\frac{c^2\omega^2}{2}} + \frac{1}{2\pi} \sum_{k=2}^{\infty} A_k(\mathbf{q}) L(\kappa_k, \omega). \quad (29)$$

The incoherent and coherent parts of S, and the orientationally averaged versions of S are directly obtained by substituting the corresponding versions of the amplitudes A_k.

D. Conformational dynamics representation of the dynamic susceptibility $\chi''(\mathbf{q}, \omega)$

Finally, we consider the dynamic susceptibility, $\chi''_{\text{inc}}(\mathbf{q}, \omega)$, which can also be directly represented in terms of the propagator eigenvalues and eigenfunctions:

$$\chi''_{\text{inc}}(\mathbf{q}, \omega) = \frac{S_{\text{inc}}(\mathbf{q}, \omega)}{n_B(\omega) + 1} = A_1(\mathbf{q}) \sqrt{\frac{c}{2\pi}} (1 - e^{-\frac{\hbar\omega}{k_B T}}) e^{-\frac{c^2\omega^2}{2}} \quad (30)$$

$$+ \sum_{k=2}^{m} \frac{A_k(\mathbf{q})}{\pi} \frac{L(\kappa_k, \omega)}{n_B(\omega) + 1}. \quad (31)$$

As noted above, the local maxima of χ''_{inc} in ω are useful for determining the relaxation rates of the underlying molecular dynamics. To see this, we consider the individual dynamical terms of χ''_{inc}:

$$\chi''_k(\mathbf{q}, \omega) = \frac{A_k(\mathbf{q})}{\pi} \frac{L(\kappa_k, \omega)}{n_B(\omega) + 1}.$$

Taking the derivative with respect to ω yields

$$\frac{d\chi''_k(\mathbf{q}, \omega)}{d\omega} = 2A_k(\mathbf{q}) \frac{\hbar\left(\kappa_k^2 + w^2\right) - 2\frac{e^{\hbar\omega/kT} - 1}{e^{\hbar\omega/kT}} kT\omega}{\pi\kappa_k kT\left(\kappa_k^2 + w^2\right)^2}$$

and choosing $\omega = \kappa_k$ yields

$$\left.\frac{d\chi''_k(\mathbf{q}, \omega)}{d\omega}\right|_{\omega=\kappa_k} = A_k(\mathbf{q}) \frac{\frac{\hbar\kappa_k}{kT} - \frac{e^{\hbar\kappa_k/kT} - 1}{e^{\hbar\kappa_k/kT}}}{\pi\kappa_k^4}.$$

$\hbar\kappa_k/kT$ is a small number. We can thus make the approximation $e^{\hbar\kappa_k/kT} \approx 1 + \hbar\kappa_k/kT$, resulting in

$$\frac{e^{\hbar\kappa_k/kT} - 1}{e^{\hbar\kappa_k/kT}} \approx \frac{\hbar\kappa_k/kT}{1 + \hbar\kappa_k/kT} \approx \frac{\hbar\kappa_k}{kT}$$

and thus

$$\left.\frac{d\chi''_k(\mathbf{q}, \omega)}{d\omega}\right|_{\omega=\kappa} \approx 0.$$

Therefore, to a very good approximation, the maxima of χ'' indicate the relaxation rates κ_k of the molecule.

IV. MARKOV STATE MODELS

Markov state models (MSM) describe the kinetic information harvested from Molecular Dynamics (MD) simulations in terms of a transition matrix between conformational substates. We briefly summarize the Markov model approach, following the more extensive theoretical description in Ref. 33. We then transfer the results of Sec. III to MSM quantities and show how the dynamical scattering functions F, S, and χ'' can be practically calculated from an MD simulation, in such a way that detailed insight into the individual relaxation processes is obtained. This approach readily facilitates a decomposition of the neutron or X-ray scattering spectra into different components, each of which is associated with a specific intramolecular transition process.

MD simulations produce trajectories, given by the Cartesian coordinates of particles in time and space. For molecular systems, the motion can be decomposed into translational, rotational, and internal dynamics, where the latter can be conveniently described in the framework of Markov models. The reason for using a Markov model is that it provides a means to approximate the eigenvalues and eigenfunctions of the propagator. It can be shown that Markov models use an ansatz for basis functions which are then used to approximate the eigenfunctions and then finding the optimal linear coefficients for representing them.[49] While the most common basis sets used are step functions that are constant on some clusters of molecular conformations,[15] scaled Gaussian functions[50] and committor functions[25,51] have also been proposed.

A. Discrete-state Markov model

Here we will concentrate on the usual Markov model approach that is based upon a discretization of the internal (molecular) conformational space into a set of substates $S = \{\Omega_1, \ldots, \Omega_n\}$. Such a discretization is typically obtained by applying a data clustering method and subsequently performing a Voronoi partitioning on the cluster centers.[17] The $n \times n$ transition matrix $\mathbf{T}(\tau)$ contains the conditional probabilities T_{ij} of finding the system in state j at time $t + \tau$ given that it was in state i at time t:

$$T_{ij} = \mathbb{P}(\mathbf{x}_{t+\tau} \in \Omega_j \mid \mathbf{x}_t \in \Omega_i) = \frac{1}{\pi_i} \int_{\Omega_i} d\mathbf{x}\, \mu(\mathbf{x}) \int_{\Omega_j} d\mathbf{y}\, p(\mathbf{x}, \mathbf{y}; \tau)$$

with π_i being the stationary probability of state i and is thus a coarse-grained version of the propagator. Note that this version of $\mathbf{T}(\tau)$ is row-dominant, i.e., each row i contains the probabilities of leaving state i for various target states, and thus each row sum is 1. In other texts, especially in physical chemistry, one sometimes finds the column-dominant transition matrix, which is just the transpose of the present $\mathbf{T}(\tau)$. Note also that some other approaches use rate matrices $\mathbf{K}$ which are related to transition matrices via $\mathbf{T}(\tau) = \exp(\tau\mathbf{K})$.[25,51,52] Rate matrices are appealing because they do not use a lag time, τ. However, it is rather difficult to estimate rate matrices,[25,53,54] and they are only meaningful for a coarse, metastable partition of state space for which the

notion of transition rates applies. For our purpose, $\mathbf{T}(\tau)$ is more useful than a rate matrix because it is compatible with a fine partition of state space that is suited to approximate the true propagator eigenfunctions ϕ_k with high precision.

Using the assumption of detailed balance, which holds true for a system under equilibrium conditions, the eigenvalue decomposition (EVD) of the transition matrix $\mathbf{T}(\tau)$ yields

$$\mathbf{T}(\tau)\hat{\psi}_k = \hat{\lambda}_k(\tau)\hat{\psi}_k,$$
$$\hat{\phi}_k\mathbf{T}(\tau) = \hat{\lambda}_k(\tau)\hat{\phi}_k,$$

where

$$\hat{\phi}_k = \mathrm{diag}(\pi_1, ..., \pi_n)\hat{\psi}_k$$

relates the left and right eigenvectors. $\hat{\phi}$ and $\hat{\psi}$ are the coarse-grained versions of the propagator eigenfunctions ϕ and ψ, and therefore provide information about structural changes, as illustrated in Fig. 1, while the corresponding eigenvalues, λ_k, describe the estimated relaxation times:

$$\hat{t}_k = \hat{\kappa}_k^{-1} = -\frac{\tau}{\ln\hat{\lambda}_k(\tau)}. \tag{32}$$

Equation (32) provides an implied timescale, i.e., the timescale that is implied by the estimated eigenvalue $\hat{\lambda}_k(\tau)$ when the exponential decay of eigenvalues of Eq. (12) is directly applied to estimated quantities: $\hat{\lambda}_k = \mathrm{e}^{-\tau/\hat{t}_i}$. However, due to the discretization error of the Markov model, this exponential decay of eigenvalues will only be approximate, and consequently the timescale $\hat{t}_k$ will depend on the lag time, τ. Fortunately, increasing τ reduces the approximation error of the eigenvalues[55] and the implied timescales/rates.[56] Thus, identifying the minimal lag time τ at which the implied timescales $\hat{t}_k$ become approximately constant in τ is a good choice as a Markov model lagtime.[13]

Like the continuous propagator, a Markov model transition matrix allows probability vectors to be propagated in time:

$$\mathbf{p}_{t+\tau}^T = \mathbf{p}_t^T\mathbf{T}(\tau), \tag{33}$$

where $\mathbf{p}_t$ is an n-dimensional row vector containing the probabilities of finding the system in each of its n states at time t. If the time evolution of the probabilities follows the Markov property, then for longer time increments, $n\tau$, $\mathbf{T}(n\tau) = \mathbf{T}^n(\tau)$ holds and Eq. (33) becomes

$$\mathbf{p}_{t+n\tau}^T = \mathbf{p}_t^T\mathbf{T}^n(\tau). \tag{34}$$

Due to the approximation error of the Markov models, Eq. (34) is only approximately correct. The propagation error can be theoretically bounded by expressions given in Ref. 57, but in practice the approximate correctness of Eq. (34) needs to be systematically tested statistically. Approaches to this are described in Refs. 25 and 33.

B. Approximation of eigenvalues and dynamical scattering intensities

In order to approximate the dynamical scattering functions with Markov models, we need to approximate the two quantities relevant for computing the intermediate scattering function via Eqs. (16) and (18): (1) the propagator eigenvalues $\lambda_k(\tau)$ and related timescales t_k or rates κ_k, and (2) the atom-wise scattering intensities $I_{\alpha k}$, given by Eq. (15), which depend on the propagator eigenfunctions ϕ_k. Here we discuss deterministic approximation errors from the state space discretization involved in Markov model construction. Statistical errors will be discussed below.

How well do the Markov model eigenvalues $\hat{\lambda}_k(\tau)$ approximate the true propagator eigenvalues $\lambda_k(\tau)$? Analogously, how well are the true relaxation timescales t_k and rates κ_k approximated by their Markov model counterparts $\hat{t}_k$ and $\hat{\kappa}_k$? As described in Ref. 49, there is a variational principle in conformational dynamics that states that the largest nontrivial Markov model eigenvalue will always be an underestimate of the true propagator eigenvalue:

$$\hat{\lambda}_2(\tau) \leq \lambda_2(\tau). \tag{35}$$

Hence, the relaxation timescale t_2 will also be underestimated while the relaxation rate κ_2 will be overestimated. Reference 55 provides error bounds for the eigenvalue, and Ref. 56 provides a practically computable error estimate for the slowest estimated relaxation rate $\hat{\kappa}_2$. The only situation in which the estimation is exact, $\hat{\lambda}_2 = \lambda_2$, is when the corresponding eigenfunction ϕ_2 is modeled by the Markov model without error. This can only be achieved in the unpractical limit of an infinitely fine state space discretization, but can be achieved to a very good approximation with relatively few states.[57] When the first $k-1$ eigenfunctions are well approximated, the variational principle also applies to the kth eigenvalue

$$\hat{\lambda}_k(\tau) \leq \lambda_k(\tau), \tag{36}$$

and the error expressions for $\hat{\lambda}_2$, $\hat{t}_2$, $\hat{\kappa}_2$ can be translated to $\hat{\lambda}_k$, $\hat{t}_k$, $\hat{\kappa}_k$.

Let us turn to the approximation of the dynamical scattering amplitudes $a_{\alpha k}$. We rewrite their definition in terms of state space integrals over the subsets of state space, Ω_s, belonging to each Markov model state, s.

$$\begin{aligned} I_{\alpha k} &= \langle b_\alpha \mathrm{e}^{-i\langle \mathbf{q},\mathbf{r}_\alpha(\mathbf{x})\rangle}, \phi_k(\mathbf{x})\rangle \\ &= b_\alpha \int d\mathbf{x}\mu(\mathbf{x})\psi_k(\mathbf{x})\mathrm{e}^{-i\langle \mathbf{q},\mathbf{r}_\alpha(\mathbf{x})\rangle} \\ &= b_\alpha \sum_{s=1}^{n}\int_{\Omega_s} d\mathbf{x}\mu(\mathbf{x})\psi_k(\mathbf{x})\mathrm{e}^{-i\langle \mathbf{q},\mathbf{r}_\alpha(\mathbf{x})\rangle}. \end{aligned}$$

We approximate ψ_k with its Markov model approximation, which is a step function with values $\hat{\psi}_{k,s}$ that are constant on Markov state s. The approximation thus reads

$$\begin{aligned} a_{\alpha k} &\approx b_\alpha \sum_{s=1}^{n}\hat{\psi}_{k,s}\frac{\pi_s}{\pi_s}\int_{\Omega_s} d\mathbf{x}\,\mu(\mathbf{x})\mathrm{e}^{-i\langle \mathbf{q},\mathbf{r}_\alpha(\mathbf{x})\rangle} \\ &= b_\alpha \sum_{s=1}^{n}\hat{\phi}_{k,s}\int_{\Omega_s} d\mathbf{x}\frac{\mu(\mathbf{x})}{\pi_s}\mathrm{e}^{-i\langle \mathbf{q},\mathbf{r}_\alpha(\mathbf{x})\rangle} \\ &= b_\alpha \sum_{s=1}^{n}\hat{\phi}_{k,s}\langle \mathrm{e}^{-i\langle \mathbf{q},\mathbf{r}_\alpha(\mathbf{x})\rangle}\rangle_{\Omega_s}. \end{aligned} \tag{37}$$

When molecular dynamics simulations are used, and n_s sample configurations are available in state s, the expectation

value is approximated by

$$\langle e^{-i\langle \mathbf{q}, \mathbf{r}_\alpha(\mathbf{x})\rangle}\rangle_{\Omega_s} \approx \frac{1}{n_s}\sum_{j=1}^{n_s} e^{-i\langle \mathbf{q}, \mathbf{r}_\alpha(\mathbf{x}_j)\rangle},$$

which involves statistical error, but no systematic bias provided that the MD implementation at hand itself samples the Boltzmann distribution without bias. The only bias in the approximation of $a_{\alpha k}$ comes from the approximation $\hat{\psi}_k \approx \psi_k$. However, as already indicated by Fig. 1, the eigenfunctions ψ_k are approximately constant within the metastable states, suggesting that a good approximation is indeed feasible.

The variational principle implies that when comparing different state space partitions, and the statistical error is insignificant, the state space partition that gives rise to the maximal eigenvalues $\hat{\lambda}$ (maximal timescales $\hat{t}$, minimal rates $\hat{\kappa}$) is best and should be chosen. That state space partition will also provide a Markov model that best approximates the eigenfunctions ϕ_k with its eigenvectors, and therefore best approximates the atom-wise dynamical scattering amplitudes $I_{\alpha k}$.

C. Dynamical scattering functions from MSM

Given a Markov model transition matrix $\mathbf{T}(\tau)$, the dynamical scattering functions can be computed as follows:

1. Estimate the eigenvalues $\hat{\lambda}_k(\tau)$ and the related relaxation timescales $\hat{t}_k = \tau/\ln\lambda_k(\tau)$ and rates $\hat{\kappa}_k = t_k^{-1}$.
2. Estimate the dynamical scattering intensities for all atoms α via Eq. (37) for coherent and incoherent scattering lengths.
3. Compute the scattering amplitudes via Eqs. (17) and (19).
4. If desired, compute the orientationally averaged scattering amplitudes by approximating Eqs. (22) and (23) via a sum over a large number of randomly generated orientation vectors $\mathbf{q}$.
5. Compute the incoherent, coherent, or total intermediate scattering function $F(\tau)$ using Eqs. (16), (18), or (21), respectively.
6. Compute the corresponding dynamic structure factor $S(\omega)$ using Eq. (25).
7. Compute the dynamic susceptibility $\chi''(\omega)$ using Eq. (6)

Besides the systematic approximation error discussed above, the procedure outlined here will involve statistical error due to the fact that only a limited amount of MD simulation data can be generated in practice. Fortunately, this statistical error can be quantified. The most important source of error in MD simulations typically comes from the fact that rare transitions between metastable states are poorly sampled. This error will significantly affect the estimation of the dominant eigenvalues λ_k, or timescales t_k. We will not give a complete overview of the literature to compute statistical errors in Markov models here, but rather refer to one approach that is most consistent with the framework of this paper: Reference 58 provides a Monte Carlo procedure for estimating the statistical error of any quantity computed from a Markov model transition matrix, including eigenvalues λ_k or the effect of the eigenvector uncertainty on the dynamical scattering amplitudes $a_{\alpha k}$. This error estimation is implemented in the EMMA package[34] (available at simtk.org/home/emma).

In Ref. 59 the error estimation framework was extended to include the estimation error of local state estimates such as involved in estimating the expression $\langle e^{-i\langle \mathbf{q}, \mathbf{r}_\alpha(\mathbf{x})\rangle}\rangle_{\Omega_s}$ in Eq. (37). In many practical cases, however, this latter part of the statistical error is rather small.

V. DISCUSSION AND CONCLUSION

The present study outlines a strategy for rigorously decomposing the quasielastic region of the scattering intensity of neutron and X-ray spectra into individual components, each corresponding to a single structural transition with an associated exponential decay rate. Both the coherent and incoherent parts of neutron scattering are treated. This work builds upon earlier findings that show that MSM can be applied to molecular dynamics simulation for the interpretation of dynamic fluorescence spectroscopy experiments.[1] Here, it is shown that the same approach can also be applied to readily interpret quasielastic neutron scattering experiments with only a few, albeit important, modifications to the underlying mathematical framework.

The use of MD-based Markov model is distinct from other strategies for interpreting quasielastic neutron scattering experiments, most of which depend on an analytical model with fitted parameters.[60–62] The MD Markov approach is relatively assumption free and in its most basic form simply provides a transition matrix with the transition probabilities between the various configurations of the molecule.

The most important advantage of the present analysis of scattering functions via Markov models over the direct computation of the time-correlation functions (1) from molecular dynamics simulations is that the experimentally measurable signal with its observable relaxation rates κ_k and the related amplitudes A_k are separately obtained. Via the duality of eigenvalues (that determine the relaxation rates κ_k) and eigenvectors (that determine the amplitudes A_k and indicate the nature of the transition associated with rate κ_k), one can then explicitly assign structural transitions to experimentally measurable relaxation timescales.

In practice, most of the amplitudes A_k will be near zero. Thus, even complicated kinetics may have the signature of two- or three-state systems in a single given kinetic experiment. Fortunately, dynamic neutron scattering experiments permit changing the amplitudes A_k without changing the system dynamics, by increasing or decreasing the dynamical scattering intensity of selected atoms, $I_{\alpha k}$ using isotope labeling. Therefore, dynamic neutron scattering offers a rather direct approach to the experimental design strategy described in Ref. 1: with MD simulation and a Markov model approximation of the dynamics of the system at hand, a series of m separate isotope labeling states can be proposed that optimally probe the m slowest individual relaxations, each maximizing the relative amplitude of one of the dominant amplitudes $A_1, \ldots, A_m$.

For a practical illustration of the present methodology, please refer to Paper II,[36] which applies the theoretical

approach in order to reconcile MD simulations of a solvated peptide with neutron scattering experiments.

ACKNOWLEDGMENTS

We greatly thank Dr. Yinglong Miao for valuable discussion and help. We would like to acknowledge the financial support from the National Science Foundation Award (No. MCB-0842871). Frank Noé acknowledges support through DFG Grant No. 825/2. Frank Noé and Jan-Hendrik Prinz acknowledge support by the DFG research center MATHEON.

[1]F. Noé, S. Doose, I. Daidone, M. Löllmann, J. Chodera, M. Sauer, and J. Smith, Proc. Natl. Acad. Sci. U.S.A. **108**, 4822 (2011).
[2]S. W. Lovesey, *Theory of Neutron Scattering from Condensed Matter* (Oxford University Press, New York, 1987).
[3]M. Bee, *Quasielastic Neutron Scattering: Principles and Applications in Solid State Chemistry, Biology and Materials Science* (Adam Hilger, Bristol and Philadelphia, PA, 1988).
[4]W. Doster, S. Cusack, and W. Petry, Nature (London) **337**, 754 (1989).
[5]J. Smith, Q. Rev. Biophys. **24**, 227 (1991).
[6]S. Chen, Proc. Natl. Acad. Sci. U.S.A. **103**, 9012 (2006).
[7]M. Jasnin, M. Moulin, M. Haertlein, G. Zaccai, and M. Tehei, Biophys J. **95**, 857 (2008).
[8]M. Jasnin, M. Moulin, M. Haertlein, G. Zaccai, and M. Tehei, EMBO J. **9**, 543 (2008).
[9]B. Lindner and J. Smith, Comput. Phys. Commun. **183**, 1491 (2012).
[10]G. Kneller, V. Keiner, M. Kneller, and M. Schiller, Comput. Phys. Commun. **91**, 191 (1995).
[11]T. Róg, K. Murzyn, K. Hinsen, and G. Kneller, J. Comput. Chem. **24**, 657 (2003).
[12]K. Hinsen, E. Pellegrini, S. Stachura, and G. Kneller, J. Comput. Chem. **33**, 2043 (2012).
[13]W. C. Swope, J. W. Pitera, and F. Suits, J. Phys. Chem. B **108**, 6571 (2004).
[14]S. P. Elmer, S. Park, and V. S. Pande, J. Chem. Phys. **123**, 114902 (2005).
[15]C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard, J. Comput. Phys. **151**, 146 (1999).
[16]N. Singhal and V. S. Pande, J. Chem. Phys. **123**, 204909 (2005).
[17]J. D. Chodera, K. A. Dill, N. Singhal, V. S. Pande, W. C. Swope, and J. W. Pitera, J. Chem. Phys. **126**, 155101 (2007).
[18]F. Noé, I. Horenko, C. Schütte, and J. C. Smith, J. Chem. Phys. **126**, 155102 (2007).
[19]B. Keller, J.-H. Prinz, and F. Noé, Chem. Phys. **396**, 92 (2012).
[20]V. Voelz, M. Jäger, L. Zhu, S. Yao, O. Bakajin, S. Weiss, L. Lapidus, and V. Pande, Biophys. J. **100**, 515a (2011).
[21]G. Bowman, V. Voelz, and V. Pande, J. Am. Chem. Soc. **133**, 664 (2011).
[22]S. Sadiq, F. Noé, and G. de Fabritiis, Proc. Natl. Acad. Sci. U.S.A. **109**, 20449 (2012).
[23]N. Singhal, C. Snow, and V. S. Pande, J. Chem. Phys. **121**, 415 (2004).
[24]J. D. Chodera, W. C. Swope, J. W. Pitera, and K. A. Dill, Multiscale Model. Simul. **5**, 1214 (2006).
[25]N. Buchete and G. Hummer, J. Phys. Chem. B **112**, 6057 (2008).
[26]K. A. Beauchamp, R. McGibbon, Y.-S. Lin, and V. S. Pande, Proc. Natl. Acad. Sci. U.S.A. **109**, 17807 (2012).
[27]G. R. Bowman and P. L. Geissler, Proc. Natl. Acad. Sci. U.S.A. **109**, 11681 (2012).
[28]M. Held, P. Metzner, J.-H. Prinz, and F. Noé, Biophys. J. **100**, 701 (2011).
[29]I. Buch, T. Giorgino, and G. de Fabritiis, Proc. Natl. Acad. Sci. U.S.A. **108**, 10184 (2011).
[30]J.-H. Prinz, B. Keller, and F. Noé, Phys. Chem. Chem. Phys. **13**, 16912 (2011).
[31]F. Noé and S. Fischer, Curr. Opin. Struct. Biol. **18**, 154 (2008).
[32]V. Pande, K. Beauchamp, and G. Bowman, Methods **52**, 99 (2010).
[33]J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. Chodera, C. Schütte, and F. Noé, J. Chem. Phys. **134**, 174105 (2011).
[34]M. Senne, B. Trendelkamp-Schroer, A. Mey, C. Schütte, and F. Noé, J. Chem. Theory Comput. **8**, 2223 (2012).
[35]K. Beauchamp, G. Bowman, T. Lane, L. Maibaum, I. Haque, and V. Pande, J. Chem. Theory Comput. **7**, 3412 (2011).
[36]Y. Zheng, B. Lindner, J.-H. Prinz, F. Noé, and J. Smith, J. Chem. Phys. **139**, 175102 (2013).
[37]G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, J. Mol. Biol. **7**, 95 (1963).
[38]H. Hu, M. Elstner, and J. Hermans, Proteins **50**, 451 (2003).
[39]J. Hermans, Proc. Natl. Acad. Sci. U.S.A. **108**, 3095 (2011).
[40]R. J. Lavrich, D. F. Plusquellic, R. D. Suenram, G. T. Fraser, A. R. H. Walker, and M. J. Tubergen, J. Chem. Phys. **118**, 1253 (2003).
[41]M.-P. Gaigeot, J. Phys. Chem. B **113**, 10059 (2009).
[42]P. Smith, J. Chem. Phys. **111**, 5568 (1999).
[43]L. Van Hove, Phys. Rev. **95**, 249 (1954).
[44]V. Sears, Neutron News **3**, 26 (1992).
[45]J. M. Cowley, B. L. M. Peng, I. G. Ren, J. S. L. Dudarevc, and M. J. Whelanc, Int. Tables Crystallogr. **C**, 262 (2006).
[46]J. H. Roh, J. E. Curtis, S. Azzam, V. N. Novikov, I. Peral, Z. Chowdhuri, R. B. Gregory, and A. P. Sokolov, Biophys. J. **91**, 2573 (2006).
[47]E. B. Davies, Proc. London Math. Soc. **s3-45**, 133 (1982).
[48]E. B. Davies, J. Lond. Math. Soc. **s2-26**, 541 (1982).
[49]F. Noé and F. Nüske, Multiscale Model. Simul. **11**, 635 (2013).
[50]M. Weber, "Meshless methods in conformation dynamics," Ph.D. thesis, Freie Universität Berlin, Berlin, Germany, 2006.
[51]C. Schütte, F. Noé, J. Lu, M. Sarich, and E. Vanden-Eijnden, J. Chem. Phys. **134**, 204105 (2011).
[52]S. Sriraman, I. G. Kevrekidis, and G. Hummer, J. Phys. Chem. B **109**, 6479 (2005).
[53]P. Metzner, I. Horenko, and C. Schütte, Phys. Rev. E **76**, 066702 (2007).
[54]D. Crommelin and E. Vanden-Eijnden, Multiscale Model. Simul. **7**, 1751 (2009).
[55]N. Djurdjevac, M. Sarich, and C. Schütte, Multiscale Model. Simul. **10**, 61 (2012).
[56]J.-H. Prinz, J. Chodera, and F. Noé, "Spectral rate theory for two-state kinetics," Phys. Rev. X (to be published).
[57]M. Sarich, F. Noé, and C. Schütte, SIAM Multiscale Model. Simul. **8**, 1154 (2010).
[58]F. Noé, J. Chem. Phys. **128**, 244103 (2008).
[59]J. Chodera and F. Noé, J. Chem. Phys. **133**, 105102 (2010).
[60]M. Tehei, B. Franzetti, K. Wood, F. Gabel, E. Fabiani, M. Jasnin, M. Zamponi, D. Oesterhelt, G. Zaccai, M. Ginzburg, and B.-Z. Ginzburg, Proc. Natl. Acad. Sci. U.S.A. **104**, 766 (2007).
[61]X.-Q. Chu, E. Mamontov, H. O'Neill, and Q. Zhang, J. Phys. Chem. Lett. **3**, 380 (2012).
[62]H. Jansson, F. Kargl, F. Fernandez-Alonso, and J. Swenson, J. Chem. Phys. **130**, 205101 (2009).