THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

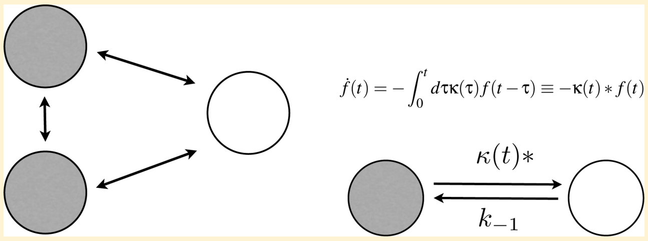
Extracting Conformational Memory from Single-Molecule Kinetic Data

Steve Pressé,*[†] Julian Lee,[‡] and Ken A. Dill[§]

[†]Department of Physics, Indiana University−Purdue University, Indianapolis, Indiana, United States

[‡]Department of Bioinformatics and Life Science, Soongsil University, Seoul, Korea

[§]Laufer Center and Departments of Physics and Chemistry, Stony Brook University, Stony Brook, New York, United States


$\dot{f}(t) = -\int_0^t d\tau\kappa(\tau)f(t-\tau) \equiv -\kappa(t) * f(t)$
$\kappa(t)*$
k_{-1}

ABSTRACT: Single-molecule data often come in the form of stochastic time trajectories. A key question is how to extract an underlying kinetic model from the data. A traditional approach is to assume some discrete state model, that is, a model topology, and to assume that transitions between states are Markovian. The transition rates are then selected according to which ones best fit the data. However, in experiments, each apparent state can be a broad ensemble of states or can be hiding multiple interconverting states. Here, we describe a more general approach called the non-Markov memory kernel (NMMK) method. The idea is to begin with a very broad class of non-Markov models and to let the data directly select for the best possible model. To do so, we adapt an image reconstruction approach that is grounded in maximum entropy. The NMMK method is not limited to discrete state models for the data; it yields a unique model given the data, it gives error bars for the model, and it does not assume Markov dynamics. Furthermore, NMMK is less wasteful of data by letting the entire data set determine the model. When the data warrants, the NMMK gives a memory kernel that is Markovian. We highlight, by numerical example, how conformational memory extracted using this method can be translated into useful mechanistic insight.

1. INTRODUCTION

It is now routine to measure single-molecule (SM) trajectories of ion channel opening/closing events or folding/unfolding transitions of proteins and nucleic acids.[1−5] SM trajectories show transitions of a molecule from one conformational state to another. For instance, in SM force spectroscopy, the trajectory includes transitions between a high force state (folded state) and a low force state (unfolded state). Individual states can exhibit what is called "conformational memory".[1] Broadly speaking, when a state has conformational memory, the dwell time in that state is not distributed as a single exponential. Alternatively, one can say that in bulk, the relaxation is not single exponential. For example, the folding of phosphoglycerate kinase monitored by bulk fluorescence experiments following a temperature jump[6] hints at a kinetically heterogeneous process; SM fluorescence experiments find that adenylate kinase apparently shows no less than six folding intermediates.[8]

Figure 1 is a cartoon illustration of how conformational memory can arise. In this cartoon example, the origin of the conformational memory is simple to understand. The low force state in Figure 1b is made up of two microscopic states; it is called an aggregate. The dwell time distribution in the low force state is therefore not a single exponential because the low force state is composed of two interconverting states indistinguishable in this particular experiment. Such models are called aggregated Markov (AM) models.[9−11] The states within an observable aggregate are assumed to exchange with each other via Markov processes. There are well-known limitations in using AM models in data analysis. First, they require advanced knowledge of the topology of the underlying kinetic states. By topology, we are referring to the kinetic relationships among the states, usually expressed as sets of arrows connecting the underlying states; see Figure 1b. As a consequence, the AM approach does not make full use of the data set; it forces the data onto a model rather than let the data tend toward a model. Second, if there are fewer independent observables than parameters in the model, then AM models are not unique; many different models would fit the data.[12] For example, a three-state system with with two observable aggregates requires six rate coefficients. However, only four independent parameters are determinable from such data. This problem is often solved by using some symmetry relationship, for example, by assuming that some rates are identical[13] or by performing additional experiments that give orthogonal information.[14] Third, AM models require the assumption that transitions are Markovian and that there are a finite number of states. In reality, there could, instead, be a continuous manifold of underlying states,[15] fluctuating rates,[20] or strong memory effects that can give rise to heavy-tail statistics.[21] Points one and three are also limitations of a related strategy tailored for noisy data, called hidden Markov (HM) modeling.[7,25] HM models

Received: September 21, 2012
Revised: December 18, 2012
Published: December 21, 2012

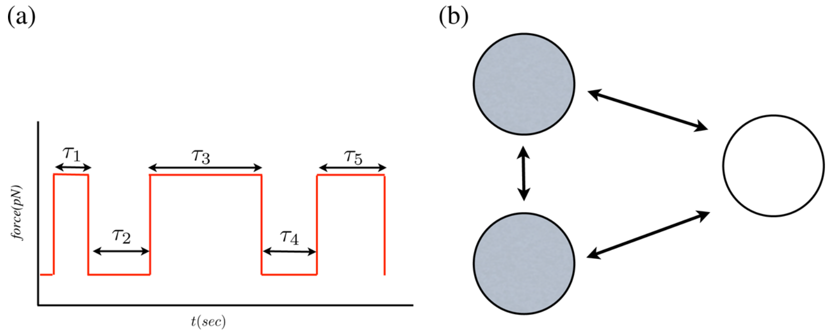


Figure 1. The experiment can only distinguish between two observable states, a high force and a low force state, for this example drawn from SM force spectroscopy. (a) A cartoon of a stochastic time trajectory shows the discrete transitions of SM between two conformational states (one extended and the other not). τ denotes the amount of time that the molecule dwells in each successive state. (b) The AM model. Here, the dark shaded state (the high force state) represents the folded state, while the unshaded state (low force state) represents the unfolded state. For this example, the dark shaded state is composed of two underlying microscopic states that transition between themselves and to all other states via a Markov process. The dwell time distribution in this state is correspondingly biexponential.

are commonly used when the microscopic state in which the molecule is found is obscured by the noise. Thus, HM modeling techniques can be useful in extracting AM models from data. That is, the HM and AM models are not mutually exclusive. The stochastic rate model is another possible description of a system with memory where the escape from a complex state is described using a rate that is a stochastic variable, in general depending on time.[22] Yet, another method is the multiscale state space network, which is useful for mapping out the interconnectivity of conformational state space of a protein from a continuously varying observable.[23] Alternatively, the conformational state space of single proteins has been described using a Langevin equation with a memory kernel fit to the data.[24]

Here, we describe an alternative approach to kinetic modeling. The idea is to start with a very broad class of kinetic models. We subsequently use the data from SM time traces and the tools of maximum entropy (MaxEnt)[15,16,26−28] to select the best model from this class. The purpose of this approach is to avoid biasing our analysis or wasting data by forcing the data ahead of time to fit a predetermined model. Rather, we let the data determine the model.

We call our approach the non-Markov memory kernel (NMMK) method. Here, the underlying system kinetics are modeled using continuous memory kernels rather than using explicit states such as in AM and HM modeling. In the NMMK method, the memory kernel itself is the model. It predicts the full dynamics without the need for any associated assumptions about kinetic rates or topologies (see Figure 2). Thus, a Markov model (a model with no memory) emerges from the NMMK method only if it is warranted by the data. In many ways, this method complements AM and HM modeling methods described above as well as approaches for inferring rate distributions (also called rate spectra) from raw data; these include a hybrid method of MaxEnt and nonlinear least-squares (MemExp)[16] as well as a variety of other methods.[17] Such methods seem less successful when applied to rate distributions that cannot be described by the sum of many exponentials.[17] Other efforts in describing sequential multistep reactions[18] are specific to exponential rates or stochastic rates sampled from a Gaussian distribution.[19]

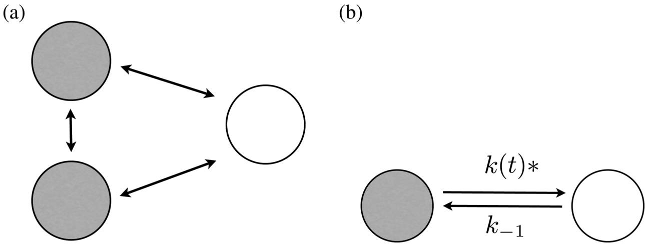


Figure 2. AM models and NMMK models. (a) AM model with an aggregated shaded state. (b) The AM model formulated as a NMMK model. The AM model is a special case of the NMMK model. See the text for details.

Here, we focus on some clear distinguishing features of the NMMK method; the model extracted is unique, transitions between states are not assumed to be Markovian from the onset, a topology is not assumed a priori, our variational method yields the time it takes for the memory in each state to decay to zero, and the method gives error bars on the model that it predicts. We also discuss what features we should expect in the memory kernel for the special case when the observed state is composed of an aggregate of states.

2. THEORETICAL METHODS

2.1. Non-Markov Memory Kernel Model (NMMK) Approach. Here, we illustrate how the NMMK method works on simulated data. In the analysis of noisy time traces of SM data, investigators first pick out transitions and obtain both marginal dwell time histograms for each state as well as conditional dwell time histograms for each pair of state (processes entirely described by marginal dwell histograms for each state are called renewal processes[30]). Furthermore, for AM models, no additional information is contained in higher-order conditional histograms.[9]

For instance, from Figure 1a, the dwell time distribution in the low force state is obtained by histogramming τ_2, τ_4,... In a follow-up publication where we will tackle real experimental data, we use change-point algorithms[31] for detecting transitions in noisy data in a model-independent way. That is, unlike the HM or AM approach, transitions are picked out from data in a way that does not depend on the topology assumed a priori.

The NMMK model describes the kinetics in terms of a generalized master equation with a memory kernel $\kappa(t)$ as follows

$$\dot{f}(t) = -\int_0^t \mathrm{d}\tau\, \kappa(\tau) f(t-\tau) \equiv -\kappa(t)^* f(t) \tag{1}$$

where $f(t)$ denotes a dwell time distribution in a particular state obtained from data. (For simplicity, we will assume that the distribution from which all dwell times are sampled is stationary. Thus our memory kernel depends on the amount of time spent in a state t, $\kappa(t)$, since first arriving in that state, not on some absolute time. Likewise, if we were discussing conditional dwell distributions, we would only consider the memory kernel for having spent time t in some state A conditioned on having spent time τ in state B, say $\kappa_{AB}(t|\tau)$, and this object would again not depend on some absolute time.) When we talk about a model, we will be referring to the memory kernel. It gives us a dynamical picture of a state by telling us how the memory in a state decays.

Note that if the dwell distribution, $f(t)$, in a particular state is single exponential, then the resulting $\kappa(t)$ is a delta function. This is the signature of a memoryless, that is, Markov, process. Put another way, if the memory kernel for some state is a delta function, we conclude that this state is a single state whose transitions to its neighbors satisfy the Markov property.

2.2. AM Models Show Clear Signatures in the Memory Kernel.

Suppose that some state A is an aggregate of indistinguishable discrete states. We show in the Appendix that AM models always give rise to multiexponential dwell time distributions. What memory kernel should we expect for state A?

In this case, the most general dwell time distribution in this state is

$$f(t) = \sum_i c_i \mathrm{e}^{-\lambda_i t} \tag{2}$$

An exponential decay with N components can arise in cases where the state is an aggregate of N states. In Laplace space, eq 1 is

$$s\hat{f}(s) - f(0) = -\hat{\kappa}(s)\hat{f}(s) \tag{3}$$

Substituting eq 2 into eq 3 yields

$$-\kappa(t) = \mathcal{L}^{-1}\left(s - \frac{\sum_i c_i}{\sum_i \frac{c_i}{\lambda_i + s}}\right) \equiv \mathcal{L}^{-1}O \tag{4}$$

where $\mathcal{L}^{-1}O$ denotes the inverse Laplace transform of O. We will perform the inverse Laplace transformation by expressing O using partial fractions as $\sum_m h_m/(s + \sigma_m) + K$ (for constant K). This gives

$$-\kappa(t) = \sum_m \mathrm{e}^{-\sigma_m t} h_m + K\tilde{\delta}(t) \tag{5}$$

where $\tilde{\delta}(t)$ is defined by the property $\int_0^\infty y(t)\tilde{\delta}(t)\,\mathrm{d}t = y(0)$. The explicit form of eq 5 is

$$-\kappa(t) = \sum_{m=1}^{n-1} \mathrm{e}^{-\sigma_m t}\left(\frac{-\prod_{i=1}^{n}(\lambda_i - \sigma_m)}{\prod_{j=1,j\neq m}^{n-1}(\sigma_j - \sigma_m)}\right) - \frac{\sum_i c_i \lambda_i}{\sum_i c_i}\tilde{\delta}(t) \tag{6}$$

where the $\{\sigma_i\}$ are the zeroes of $\sum_i c_i 1/(\lambda_i + s)$.

For the special case that all $\{c_i\}$'s in eq 2 are positive, all parameters $\{\sigma_m, h_m\}$ are also positive. To see this explicitly, consider a plot of $\hat{f}(s)$ (see Figure 3), where it is clear that $\lambda_1 < \sigma_1 < \lambda_2 < \sigma_2 \ldots < \sigma_{n-1} < \lambda_n$. The signs of the $\{h_m\}$ are then determined from its explicit form in eq 6. From this ordering and eq 6, it follows that all $\{\sigma_m, h_m\}$'s are positive.

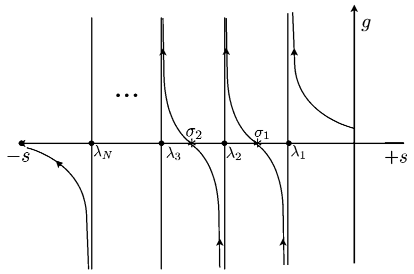


Figure 3. Pole and zero structure of $g \equiv \hat{f}(s)$. Dots denote poles, and asterisks denote zeroes.

The structure of the zeroes and poles is more complicated when some of the $\{c_i\}$'s are negative, which can apply in instances where detailed balance is violated in the underlying AM model. Nonetheless, we can still rank order the zeroes and poles to determine the sign of $\{\sigma_m, h_m\}$.

The type of structure that we expect in the memory kernel from an AM model is given by eq 6. Equation 6 captures a sharp spike in the memory kernel at $t = 0$ (the hallmark of a Markov process) in addition to components present at later times that describe the relaxation of the memory kernel.

While in its full generality eq 6 looks daunting, consider a simple biexponential dwell time distribution with positive $\{c_1, c_2\}$. This dwell describes the decay from an aggregated state with two different time scales, a faster and a slower time scale. In this case, the memory kernel is a sharp positive spike at the origin (the delta function) followed by a dip below zero and then followed by an exponential rise back to zero. See the green curve in Figure 4b for an example of this behavior. At the origin, the memory kernel's sharp positive spike says that the decay is memoryless. At the next time step, the slower time scale becomes relevant and reduces the effective rate of escape from this state. This coincides with the negative component of the memory kernel. In other words, the slower time scale introduces memory in the memory kernel; it introduces mathematical structure beyond the single spike at the origin.

Memory kernels arising from multiexponential dwell time distributions are helpful in building intuition because the forms for the coinciding memory kernels have analytic mathematical forms. We have just shown how additional states yield effective memory in the system. Given an AM model, the structure of the memory kernel can then be interpreted according to the presence of aggregates of discrete states. However, AM models are idealizations of real data, and the structure in the memory kernel can be interpreted differently. For example, the structure in the memory kernel can be interpreted in the context of a diffusion model in a rough energy landscape, a model recently used to describe the dwell in the unfolded state of phosphoglycerate kinase that we discussed earlier.[6] Alternatively, we are free to interpret the memory kernel as a model where the rates themselves are stochastic variables. We argue here that the memory kernel in itself provides a model; it dictates the way in which the memory decays to zero within error as well as the amount of time that this takes.

We now turn our attention to developing a robust algorithm for extracting memory kernels from noisy histograms. Answering this question will help us answer questions like, "is a Markov model (the most basic of all models normally taken for granted) warranted by the data?".

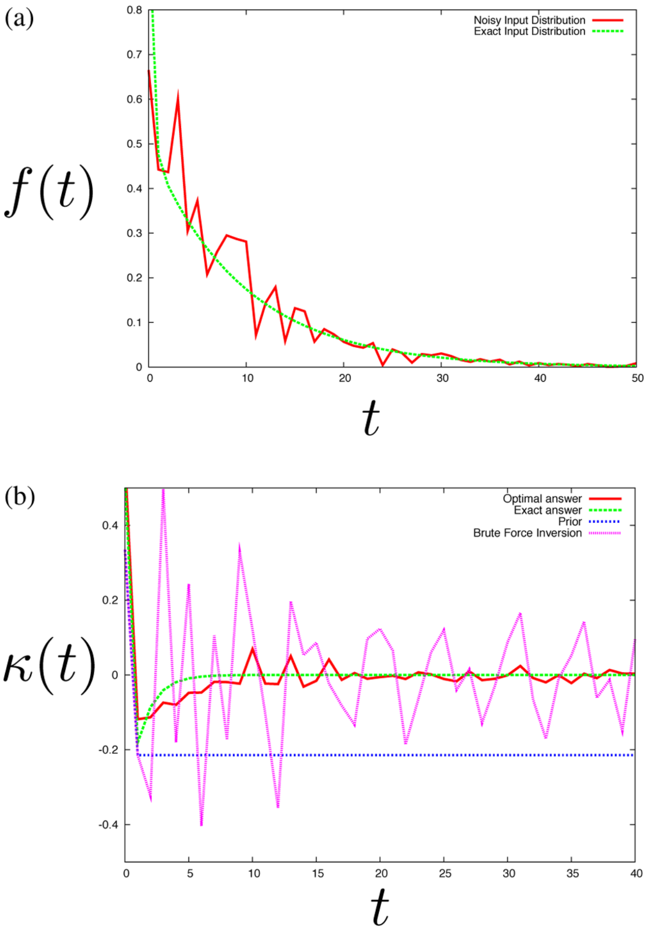


Figure 4. Extraction of the memory kernel from the data. (a) Time trace data, where the green curve is the double exponential function and the red curve is the fictitious data obtained by adding 40% noise. (b) The memory kernel. The green curve is the exact answer, the pink curve is the one obtained by brute force (i.e., direct) inversion, the blue curve is the prior, and the red curve is the solution obtained by image reconstruction (the optimal answer). See the text for details.

2.3. An Algorithm for Extracting the Memory Kernel.

As noted earlier, transitions between states in SM trajectories are rarely as clear as those shown in Figure 1a. Rather, transitions are obscured by noise. In this case, a dwell distribution like $f(t)$ would be obtained by using a model-independent change-point algorithm to detect transitions in the real noisy data.[31] That is, unlike HM models, the change-point algorithm would pick out significant transitions in the data without committing to a particular model from the onset. The output of this procedure is a “denoised time trace” like the one given in Figure 1a. The amount of time spent in each state is histogrammed. Because data is finite, the histograms themselves are noisy.

The numerical extraction of the memory kernel from noisy histograms can be mapped onto an adaptation of the method of image reconstruction.[15,26−28] Briefly, the goal of image reconstruction is to obtain an image, I, from data, D, where the data and image are related by a linear transform G. That is $D = G*I$. Direct inversion of the data, $(G^{*-1})D$, would be numerically unstable for noisy data; therefore, we use the variational procedure of image reconstruction to regularize the operation and obtain a reconstructed image. The variational procedure is based on the principle of MaxEnt.

The analogue of eq 1 in discrete time is

$$f_{j+1} - f_j = -\sum_{k=0}^{j} f_k \kappa_{j-k} \tag{7}$$

Here, the memory kernel plays the role of the image. The important difference between eq 7 and standard image reconstruction is the operator G. Our operator in eq 7 contains noisy data. Our goal is to derive a variational procedure for extracting the memory kernel with such an operator. We do so by defining an objective function that we will optimize with respect to each κ_j in order to obtain our reconstructed memory kernel.

We begin by assuming the experimental input to be in the form of a dwell time histogram with error bars for each bin. That is $f_j^{\text{exp}} = f_j + \varepsilon_j$, where f_j is the theoretical value of the dwell time distribution at time point j. Brute force inversion of $f_{j+1}^{\text{exp}} - f_j^{\text{exp}} = -\sum_{k=0}^{j} f_k^{\text{exp}}\kappa_{j-k}$ to obtain $\{\kappa_j\}$ is numerically unstable because noise propagates quickly as we solve κ_1 in terms of κ_0, κ_2 in terms of κ_1, and so on. Image reconstruction is used as a regularizing procedure to overcome this problem. For this reason, using $f_j^{\text{exp}} = f_j + \varepsilon_j$, we write eq 7 as follows

$$\Delta f_{j+1,j}^{\text{exp}} + \sum_{k=0}^{j} f_k^{\text{exp}}\kappa_{j-k} = \varepsilon_{j+1} - \varepsilon_j + \sum_{k=0}^{j}\varepsilon_k\kappa_{j-k} \tag{8}$$

where $\Delta f_{j+1,j}^{\text{exp}} \equiv f_{j+1}^{\text{exp}} - f_j^{\text{exp}}$. The deviations of experiment and theory, the residuals, are on the right-hand side of eq 8. We assume $\langle\varepsilon_i\varepsilon_j\rangle = \sigma_j^2\delta_{ij}$ and $\langle\varepsilon_i\rangle = 0$. Squaring both sides of eq 8 and taking the average with respect to the noise, we find

$$\begin{aligned}\langle(\Delta f_{j+1,j}^{\text{exp}} + \sum_{k=0}^{j} f_k^{\text{exp}}\kappa_{j-k})^2\rangle \\ = \sigma_{j+1}^2 + \sigma_j^2 - 2\sigma_j^2\kappa_0 + \sum_{k=0}^{j}\sigma_k^2\kappa_{j-k}^2\end{aligned} \tag{9}$$

We define a χ^2 statistic as a sum over all time intervals

$$\chi^2 \equiv \sum_{j=0}^{N}\frac{(\Delta f_{j+1,j}^{\text{exp}} + \sum_{k=0}^{j} f_k^{\text{exp}}\kappa_{j-k})^2}{(\sigma_{j+1}^2 + \sigma_j^2 - 2\sigma_j^2\kappa_0 + \sum_{k=0}^{j}\sigma_k^2\kappa_{j-k}^2)} \tag{10}$$

where N here is the number of data points. For a finite sample, we may invoke the frequentist logic often used in image reconstruction,[15,26−28] and suppose that $\chi^2 = N$. That is, on average, the difference between each data point and its theoretical expected value differ by their standard deviation. We then follow the logic of MaxEnt, which is often invoked in image reconstruction, and ask that the memory kernel be as featureless as possible given $\chi^2 = N$ as a constraint on the data. In other words, we now maximize the objective function, $F(\theta, \{\kappa\})$, with respect to the set $\{\kappa_j\}$

$$\begin{aligned}F(\theta, \{\kappa\}) = -\alpha\sum_j(\kappa_j + \bar{\kappa})\log\left(\frac{\kappa_j + \bar{\kappa}}{\Lambda_j + \bar{\kappa}}\right) \\ - \frac{\beta}{2}(\chi^2 - N)\end{aligned} \tag{11}$$

The entropy of the memory kernel is $-\sum_j(\kappa_j + \bar{\kappa})\log((\kappa_j + \bar{\kappa})/(\Lambda_j + \bar{\kappa}))$; $\{\alpha, \beta\} = \{\cos^2\theta, \sin^2\theta\}$ are the Lagrange multipliers that enforce the constraints on the data. (Only one Lagrange multiplier is independent because $\alpha + \beta = 1$. When we take a derivative of the objective function, eq 11, with respect to κ_j, we can divide through the entire expression by β. Then we are left with only one Lagrange multiplier, the ratio α/β.) Λ_j is the prior on κ_j; and $\overline{\kappa}$ is a constant positive parameter to ensure that $\kappa_j + \overline{\kappa}$ never becomes negative so that the entropy is well-defined. Our estimate of the set $\{\kappa_j\}$ in the absence of data is therefore $\kappa_j^0 = (\Lambda_j + \overline{\kappa})\mathrm{e}^{-1} - \overline{\kappa}$. By setting $\chi^2 = N$, the problem is underdetermined. We therefore need to use MaxEnt to select one of the many possible solutions for which $\chi^2 = N$. In other words, the Lagrange multipliers "tune" the balance between fitting the data points on the one hand and "smoothing" to achieve the model closest to the prior.

Now, we need to specify our prior. In the absence of any data, we should assume a simple Markov memory kernel. In our case, we set the first two points from our brute force memory kernel (j = 0 and 1), which are more reliably determined than later points, as the first two points of the prior, Λ_j. We take the rest of the prior to be flat. The idea is to take advantage of the structure of the memory kernel we know to be reliable to set our prior.

Next, the 95% confidence interval of the memory kernel is estimated assuming that the true memory kernel lies somewhere within the solution obtained by optimizing the objective function under the constraint $\chi^2 = \mu \pm 2\sigma$ rather than $\chi^2 = N$ when N is large (where the χ^2 distribution has $\mu = N$ and $\sigma = (2N)^{1/2}$). We should expect the memory kernel's lower bound (when $\chi^2 = \mu - 2\sigma$) to be close to the brute force inversion (where χ^2 is strictly zero) and its upper bound (when $\chi^2 = \mu + 2\sigma$) to be closer to the chosen prior (where χ^2 is very large).

As an aside, we choose the Shannon–Jaynes entropy in eq 11 because it worked for all examples that we have considered so far and it tries to minimize features in the memory kernel; we show only a small fraction of the many numerical examples that we tackled in the next section. It is quite possible that other entropies, or other regularizing procedures, would also serve as acceptable substitutes to the Shannon–Jaynes entropy.

3. RESULTS AND DISCUSSION: PROOF OF PRINCIPLE OF THE NMMK METHOD

Figure 4 shows an illustration of how NMMK methods work on a theoretical example. We first create a decay signal and then add noise to it. We imagine that such a dwell distribution originates from having used a change-point algorithm to detect significant transitions in a real time trace, as discussed earlier. Then, we ask how well NMMK can extract the correct underlying model. In this case, we added 40% noise to a biexponential dwell time distribution, $p_j = 0.5 \times 0.9^j + 0.5 \times 0.05^j$; the raw "data" made up in this way is shown in Figure 4a.

First, as a point of reference, Figure 4b shows the memory kernel that is obtained by brute force inversion of the raw data (pink line). It is too noisy to give a good approximation to the exact answer (green line) or even to register that there are two exponentials in the dwell distribution. The exact memory kernel (computed from eq 6) is not simply a delta function; it shows more structure, consistent with the presence of a second state. Now, in order to apply the NMMK method, we use only the first two points of the pink line to construct the prior (blue line). Using that prior, we then use the NMMK image reconstruction method. The NMMK predictions are shown as the red curve. The NMMK method clearly shows a peak followed by a dip, consistent with the correct two exponentials even when the noise level is high. This example shows that while the double exponential behavior is invisible in the time trace due to a high noise level (Figure 4a, red line), the double exponential is nevertheless detectable as the rising tail in the memory kernel obtained by the image reconstruction algorithm (Figure 4b, red curve). Hence, the NMMK method appears to be a sensitive model-building strategy that can detect structure in the memory kernel (or hidden intermediate states in the language of AM), without assuming that the process is Markovian.

A good data-processing algorithm should display both sensitivity and specificity. Sensitivity means that small changes in the data originating from the underlying physics should be reflected in the corresponding model. Specificity means that irrelevant differences in the data due to noise should not be captured in the model. Figures 4–9 show our tests of sensitivity and specificity of the NMMK approach.

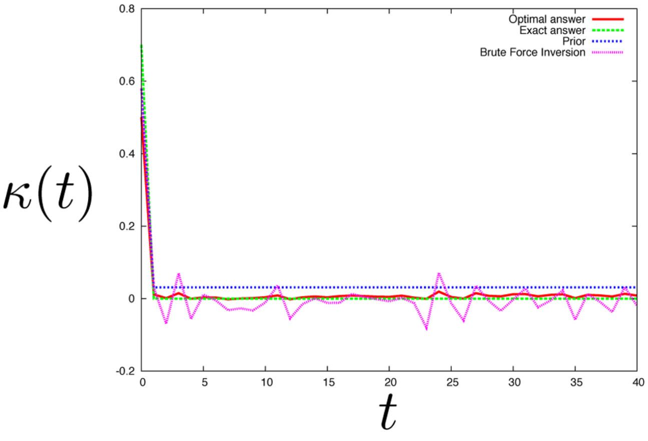


Figure 5. Extraction of the memory kernel from the data. The memory kernel as a function of time when the dwell time distribution coincides with a single exponential in the presence of 40% noise. Even at 40% noise, image reconstruction can reliably extract the memory kernel and pick out the delta function expected from the noise. Other features of the memory kernel apparent from the brute force curve are interpreted as noise, and the sharp peak at the origin is the only qualitatively significant feature apparent from the optimal solution.

In Figure 6, we show a memory kernel extracted from a biexponential dwell time histogram with 20% Gaussian noise. The noise is always proportional to the value of the histogram

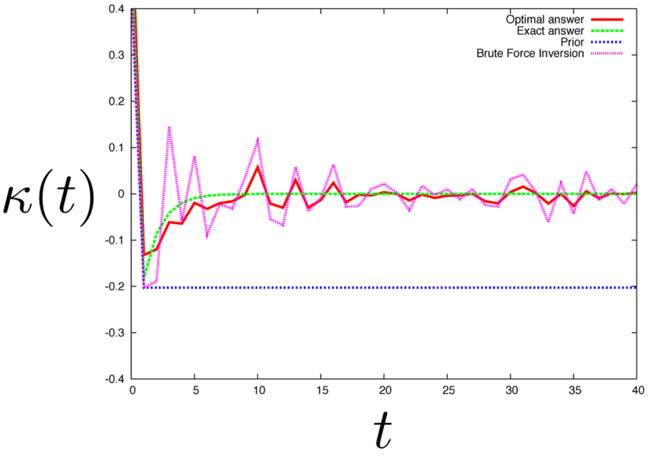


Figure 6. Extraction of the memory kernel from the data. Same as Figure 4 but now with 20% noise. As expected here, the optimal solution matches the theoretical solution much more closely.

bin for each bin. Compare this to the memory kernel extracted from a dwell histogram with 40% noise from Figure 4b. The biexponential nature of the dwell time distribution is apparent in both despite a large difference in the noise. Hence, NMMK shows specificity. Similarly, we show specificity for memory kernels extracted from a noisy monoexponential dwell time histogram in Figure 5 (with 40% noise) versus that in Figure 9 (with 60% noise). We can extract a memory kernel for a simpler model (like the monoexponential) at higher noise levels than we can for more complex models (like biexponentials). This is because the histograms of more complex models have more features that make the NMMK method less specific.

Comparing Figures 4b and 5, we find that even at 40% noise, the NMMK method can distinguish monoexponential versus biexponential models. Thus, NMMK is therefore also sensitive. In Figure 8, we show that if the two decay rates of the biexponential histogram are more similar than they are for Figure 4b, then the NMMK has difficulty picking out a reasonable memory kernel at 40% noise, though not at 20% noise (see Figure 7) for the biexpoential model considered.

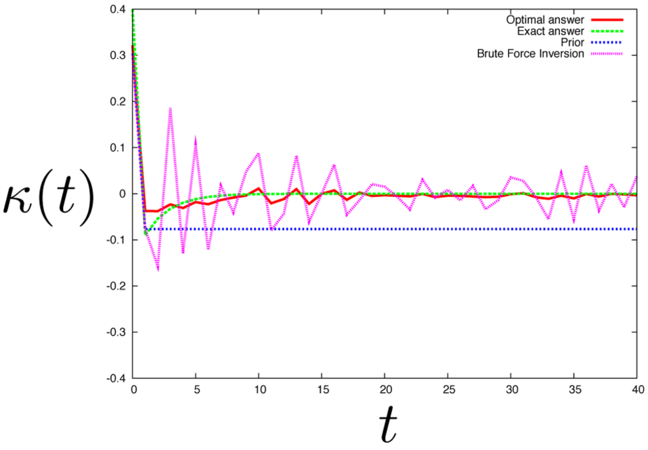


Figure 7. Extraction of the memory kernel from the data. Same as Figure 6 but with a double exponential decay described by the eq 0.5 × 0.9^j + 0.5 × 0.3^j. As the two decay constants here are more similar than those in Figure 6, the negative dip expected theoretically is less pronounced and is consequently more difficult to pick out at 20%.

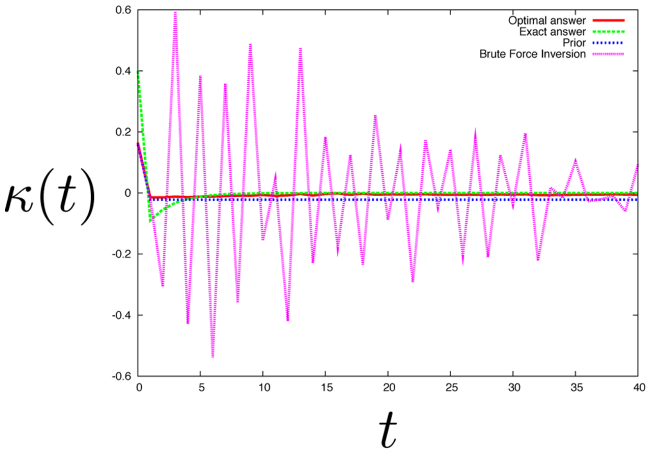


Figure 8. Extraction of the memory kernel from the data. Same as Figure 7, but now, the noise is raised to 40%. For sufficiently similar rate constants, the negative dip is now impossible pick out at 40% noise. This figure should be compared to Figure 4b, where the negative dip is more pronounced when the decay constants are more dissimilar. In that case, the negative dip is clearly noticeable even at 40% noise.

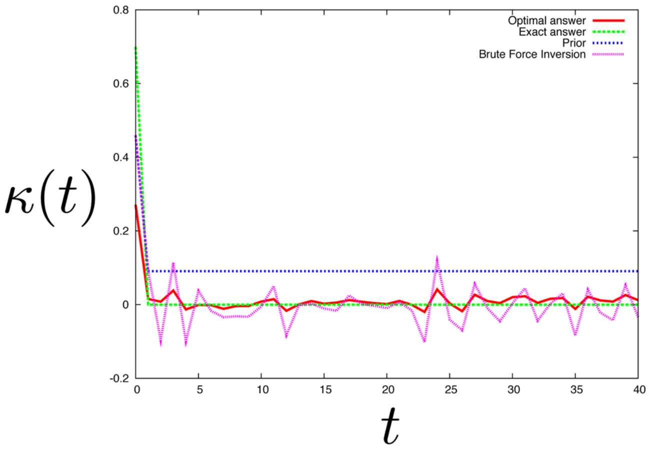


Figure 9. Extraction of the memory kernel from the data. Same as Figure 5, the monoexponential case, but now, the noise is level is raised to 60%. These histograms have fewer features than their biexponential counterparts; therefore, it is easier to extract the memory kernel from such histograms at higher noise levels. The most prominent feature of the memory kernel is, somewhat surprisingly, still the sharp peak at the origin, which is correct. However, at such high noise levels, the magnitude of the peak recorded at the origin is far from the theoretical exact answer.

(We extract our memory kernel from dwell histograms. Because we assume that the noise around each bin value of our dwell histogram is Gaussian, doubling our histogram bin size reduces the noise by a factor of $\sqrt{2}$. Selecting bin sizes that are too large can make complex memory kernels look like delta functions. The optimal bin size therefore depends on the model.).

We have illustrated how the NMMK method is used to extract memory kernels here from simple mono- and biexponential histograms. We have used sums of exponentials because they are common and important in dynamical processes and simple to illustrate. However, the NMMK is not limited in any way to exponentials. NMMK can also treat processes involving a continuum of states.

4. CONCLUSION

Stochastic trajectories are the common form of data from SM and few-particle experiments. Such processes are often treated using AM models, where it is assumed that states are clustered together into aggregated states that can interconvert using Markov processes. To go beyond the limitations of such models, we propose an alternative, the non-Markov memory kernel (NMMK) model. We show that the NMMK model has the following advantages: for given data, it gives a unique model (with error bars); it does not require inputting knowledge of the number of underlying states; it is readily combined with current first-principles-based image reconstruction methods to provide a stable numerical recipe; and it can be used even when the underlying physical states are not discrete and transitions between these are not Markovian. When the data warrant a simple Markov model, the memory kernel reduces to a delta function. The NMMK is a way to harness the entire experimental output from kinetics experiments to reconstruct kinetic models with error bars. In addition, it complements the method of image reconstruction for the determination of rate distributions though NMMK is not intrinsically tied to exponential decay forms. We have only explicitly considered extracting memory kernels from dwell histograms where the time ordering information of dwells is lost. However, our method readily generalizes to the analysis of conditional and higher-order histograms if such data are available. It is worth investigating whether our NMMK formulation could be adapted to treat time-ordered data explicitly rather than processed data like histograms.

APPENDIX: DERIVING DWELL TIME DISTRIBUTIONS FOR AGGREGATED MARKOV (AM) MODELS

We first summarize the necessary theory and notation from AM models.[9,10,12,29] We consider Markov states that are aggregated into s distinguishable groups, which we simply call aggregates. We define the Markovian rate matrix $\mathbf{Q}$, whose rows sum to zero. The submatrix $\mathbf{Q}_{\alpha\beta}$ is the rate matrix connecting the states in aggregate α to those in β. The individual matrix elements $Q^{kl}_{\alpha\beta}$ within $\mathbf{Q}_{\alpha\beta}$ for $\alpha \neq \beta$ are the physical values for the rates connecting the kth state in aggregate α to the lth state in aggregate β.

Aggregate α has n_α underlying states, where $\sum_\alpha n_\alpha = N$. The general joint probability density for the dwell time t_1 in the α_1 aggregate followed by a dwell of t_2 in α_2, and so forth, denoted as $f_{\alpha 1...\alpha r}(t_1...t_r)$, is[9]

$$f_{\alpha_1...\alpha_r}(t_1...t_r) = \pi_{\alpha_1}\exp(\mathbf{Q}_{\alpha_1\alpha_1}t_1)\mathbf{Q}_{\alpha_1\alpha_2} \times \exp(\mathbf{Q}_{\alpha_2\alpha_2}t_2) \\ ...\exp(\mathbf{Q}_{\alpha_r\alpha_r}t_r)\sum_\beta Q_{\alpha_r\beta}\mathbf{u}_\beta \tag{12}$$

where $\mathbf{u}_\beta$ is a column vector of length n_β whose elements are all equal to 1 and π_α is the row vector of length n_α whose jth element is the probability that a dwell in the aggregate α starts in the jth state of that aggregate. By diagonalizing $\mathbf{Q}_{\alpha\alpha}$ as $\mathbf{S}_\alpha^{-1}\lambda_\alpha\mathbf{S}_\alpha$ with a diagonal matrix λ_α, we obtain

$$\exp(\mathbf{Q}_{\alpha\beta}t) = \sum_{i=1}^{n_\alpha}\mathcal{P}^i_\alpha\exp(-\lambda^i_\alpha t) \tag{13}$$

λ^i_α are the eigenvalues of $\mathbf{Q}_{\alpha\alpha}$ and also the ith diagonal element of λ_α, and the projection matrix $\mathcal{P}^i_\alpha$ is defined by its elements

$$[\mathcal{P}^i_\alpha]^{kl} = (S^{-1}_\alpha)^{ki}S^{il}_\alpha \tag{14}$$

Using eq 13, eq 12 is rewritten as

$$f_{\alpha_1...\alpha_r}(t_1...t_r) = \sum_{i_1=1}^{n_{\alpha_1}}\sum_{i_2=1}^{n_{\alpha_2}}...\sum_{i_r=1}^{n_{\alpha_r}}a^{i_1...i_r}_{\alpha_1...\alpha_r}\exp(-\sum_{k=1}^{r}\lambda^{i_k}_{\alpha_k}t_k) \tag{15}$$

and

$$a^{i_1...i_r}_{\alpha_1...\alpha_r} = \pi_{\alpha_1}\mathcal{P}^{i_1}_{\alpha_1}\mathbf{Q}_{\alpha_1\alpha_2}...\mathcal{P}^{i_r}_{\alpha_r}\sum_\beta\mathbf{Q}_{\alpha_r\alpha_\beta}\mathbf{u}_\beta \tag{16}$$

For AM processes, the two-time joint probability density $f_{\alpha\beta}(t,\tau)$ for all label pairs α and β or, alternatively, the pair $f_{\alpha\beta}(t|\tau)$ and $f_\alpha(t)$ captures the full statistics of the n-time density given by eq 12.[9]

$$f_{\alpha\beta}(t,\tau) = \sum_{i,j}a^{ij}_{\alpha\beta}\exp(-\lambda^i_\alpha t - \lambda^j_\beta\tau) \tag{17}$$

An AM model is fully characterized by the combination of the model topology and the rates that are captured in $f_\alpha(t)$ and $f_{\alpha\beta}(t|\tau)$. As we noted in the main body, a set of densities $f_\alpha(t)$ and $f_{\alpha\beta}(t|\tau)$ for all α and β can correspond to many potential topologies and rates. AM models are therefore not unique in general.[12]

AUTHOR INFORMATION

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We appreciate the support of NIH Grant GM34993 and NIH/NIGMS Grant GM09025. S.P. acknowledges the FQRNT for its support and G. J. Peterson, E. Zimanyi, K. Ghosh, A. Szabo, and R. J. Silbey for their insights.

REFERENCES

(1) Watkins, L.; Chang, H.; Yang, H. *J. Phys. Chem. A* **2006**, *110*, 5191−5203.
(2) Mickler, M.; Hessling, M.; Ratzke, C.; Buchner, J.; Hugel, T. *Nat. Struct. Mol. Biol.* **2009**, *16*, 281−286.
(3) Hamill, O. P.; Marty, A.; Neher, E.; Sakmann, B.; Sigworth, F. J. *Pfluegers Arch.* **1981**, *391*, 85−100.
(4) Shank, E. A.; Cecconi, C.; Dill, J. W.; Marqusee, S.; Bustamante, C. *Nature* **2010**, *465*, 637−640.
(5) Neuweiler, H.; Johnson, C. M.; Fersht, A. R. *Proc. Natl. Acad. Sci., U.S.A.* **2010**, *106*, 18569−18574.
(6) Osváth, S.; Sabelko, J. J.; Gruebele, M. *J. Mol. Biol.* **2003**, *333*, 187−199.
(7) Bronson, J. E.; Fei, J.; Hofman, J. M.; Gonzalez, R. L., Jr.; Wiggins, C. H. *Biophys. J.* **2009**, *97*, 3196−3205.
(8) Pirchi, M.; Ziv, G.; Riven, I.; Cohen, S. S.; Zohar, N.; Barak, Y.; Haran, G. *Nat. Commun.* **2011**, *2*, 493.
(9) Fredkin, D. R.; Rice, J. A. *J. Appl. Prob.* **1986**, *23*, 208−214.
(10) Colquhoun, D.; Hawkes, A. G. *Proc. R. Soc. London, Ser. B* **1981**, *211*, 205−235.
(11) Qin, F.; Auerbach, A.; Sachs, F. *Proc. R. Soc. London, Ser. B* **1997**, *264*, 375−383.
(12) Kienker, P. *Pr. R. Soc. London, Ser. B* **1989**, *236*, 269−309.
(13) Xie, L.-H.; John, S. A.; Ribalet, B.; Weiss, J. N. *J. Physiol.* **2008**, *586*, 1833−1848.
(14) Milescu, L. S.; Akk, G.; Sachs, F. *Biophys. J.* **2005**, *88*, 2494−2515.
(15) Steinbach, P. J.; Chu, K.; Frauenfelder, H.; Johnson, J. B.; Lamb, D. C.; Nienhaus, G. U.; Sauke, T. B.; Young, R. D. *Biophys. J.* **1992**, *61*, 235−245.
(16) Steinbach, P. J.; Ionescu, R.; Matthews, C. R. *Biophys. J.* **2002**, *82*, 2244−2255.
(17) Voelz, V.; Pande, V. *Proteins: Struct., Funcy., Bioinf.* **2011**, *80*, 342−351.
(18) Zhou, Y.; Zhuang, X. *J. Phys. Chem. B* **2007**, *111*, 13600−13610.
(19) Floyd, D. L.; Harrison, S. C.; van Oijen, A. M. *Biophys. J.* **2010**, *99*, 360−366.
(20) Yang, H.; Luo, G.; Karnchanaphanurach, P.; Louie, T.-M.; Rech, I.; Cova, E.; Xun, L.; Xie, X. S. *Science* **2003**, *302*, 262−266.
(21) Siwy, Z.; Ausloos, M.; Ivanova, K. *Phys. Rev. E* **2002**, *65*, 031907.
(22) Yang, S.; Cao, J. *J. Chem. Phys.* **2002**, *117*, 10996−11009.
(23) Li, C.-B.; Yang, H.; Komatsuzaki, T. *Proc. Natl. Acad. Sci., U.S.A.* **2008**, *105*, 536−541.
(24) Min, W.; English, B. P.; Luo, G.; Cherayil, B. J.; Kou, S. C.; Xie, X. S. *Acc. Chem. Res.* **2005**, *38*, 923−931.
(25) Qin, F.; Auerbach, A.; Sachs, F. *Biophys. J.* **2000**, *79*, 1929−1944.
(26) Bryan, R. K. *Eur. Biophys. J.* **1990**, *18*, 165−174.
(27) Gull, S. F.; Daniell, G. J. *Nature* **1978**, *272*, 686−690.
(28) Skilling, J.; Gull, S. F. *Monogr. Ser., Spatial Stat. Imaging* **1991**, *20*, 341−367.

(29) Fredkin, D. R.; Montal, M.; Rice, J. A. *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*; Le Cam, L. M., Ohlsen, R. A., Eds.; Wadsworth: Belmont, CA, 1986; pp 269−289.

(30) Berezhkovskii, A. M. *J. Chem. Phys.* **2011**, *134*, 074114.

(31) Kalafut, B.; Visscher, K. *Comput. Phys. Commun.* **2008**, *179*, 716−723.