Bayesian nonparametrics (BNPs) are poised to have a deep impact in the analysis of single molecule data as they provide posterior probabilities over entire models consistent with the supplied data, not just model parameters of one preferred model. Thus they provide an elegant and rigorous solution to the difficult problem encountered when selecting an appropriate candidate model. Nevertheless, BNPs’ flexibility to learn models and their associated parameters from experimental data is a double-edged sword. Most importantly, BNPs are prone to increasing the complexity of the estimated models due to artifactual features present in time traces. Thus, because of experimental challenges unique to single molecule methods, naive application of available BNP tools is not possible. Here we consider traces with time correlations and, as a specific example, we deal with force spectroscopy traces collected at high acquisition rates. While high acquisition rates are required in order to capture dwells in short-lived molecular states, in this setup, a slow response of the optical trap instrumentation (i.e., trapped beads, ambient fluid, and tethering handles) distorts the molecular signals introducing time correlations into the data that may be misinterpreted as true states by naive BNPs. Our adaptation of BNP tools explicitly takes into consideration these response dynamics, in addition to drift and noise, and makes unsupervised time series analysis of correlated single molecule force spectroscopy measurements possible, even at acquisition rates similar to or below the trap’s response times. Published by AIP Publishing. https://doi.org/10.1063/1.5008842

I. INTRODUCTION

Bayesian nonparametrics (BNPs) have recently been introduced to single molecule biophysics.1–4 Their unique characteristics make them ideal tools in modeling complex biomolecules as they offer clear physical interpretation and allow full posterior probabilities over molecular-level models to be inferred with minimal subjective choices.3,5

In particular, in single molecule analysis, it is typical to associate a (mathematical) model with a molecular state space of a specific size.1,4,6 For example, a molecule with 2 distinct biophysical states is represented by a model with 2 (mathematical) states and the corresponding number of state-specific parameters. Since traditional analysis methods can only infer parameters of particular models,7–9 but not populations of parameters, they require an a priori particular choice among all such possible models. In other words, a critical requirement is that the size of the molecular state space already be known, which is often not the case.

To this end, BNPs avoid the crucial step of having to select a single model out of many plausible ones as they can infer parameters while simultaneously inferring the appropriate model.1–5 Accordingly, BNPs have been used for the analysis of single molecule time series as natural extensions of well-established parametric methods,10–14 such as Hidden Markov Models (HMMs), without the need to impose additional model selection procedures.

The infinite Hidden Markov Model15 (iHMM) exploits BNPs and is the nonparametric extension of the routinely used HMM. The iHMM performs well in the analysis of decorrelated time series1,2 though it is challenged by correlated time series. Indeed, when time correlated traces are analyzed under the assumption of decorrelated data, iHMMs may give rise to a severe overestimate in the number of states because, unlike HMMs which have a fixed number of states, iHMMs recruit new states to model the data.

Time correlations are unavoidable experimentally. For instance, time correlations arise when data acquisition times are fast, as is required to capture rapid dynamics of molecules. In such cases, transitions between discrete molecular states may appear smooth or as artifacts of the data acquisition process.

For the sake of concreteness, we focus on traces obtained from single molecule force spectroscopy (FS) where fast measurements may introduce time correlations in the data because the observed dynamics of the molecule are heavily influenced by the response of the optical trap instrumentation. This response includes the motion of the associated bead, the settling of the ambient fluid, and the relaxation of the tethering DNA handles.16–18 In fact, measurements in FS probe the motion of the combined molecule-trap system which, at
fast time scales, may deviate substantially from the motion of the underlying molecule alone. As a result, sharp molecular transitions may be distorted or smeared out because the combined system, roughly idealized as an elastic spring, transforms from discrete into continuous events that are sampled over multiple time points during the molecule-trap settling time, Fig. 1.

Due to their ability to infer the size of the molecular state space simultaneously with the properties of the constituting states,3,5,10 iHMMs can over-interpret the distorted signals and overpopulate the state space with unphysical states in those regions of the time trace where transitions between states appear continuous. Inevitably, if the raw FS traces are directly used, the estimated number of states may be severely overestimated, Fig. 1.

To lessen such inaccuracies, one may naively choose to remove correlations by pre-processing the time trace. For instance, this can be achieved by artificially downsampling the experimental trace,20 where multiple consecutive data points are binned together and replaced by their average. Such preprocessing, while reducing correlations between successive measurements within a time series, has two main disadvantages: (i) it undermines the purpose of collecting data at high frequencies in the first place by reducing the measurement rate and leaving originally observed fast kinetics unresolved and (ii) it introduces artificial short-lived dwells due to aliasing.20 In turn, these artificial short-lived dwells may also be interpreted by iHMMs as artificial states leading to the same downfalls as when using the original, unprocessed, raw signal, Fig. 1.

In summary, iHMMs are advantageous because they do not assume a fixed number of states. Yet iHMMs fail when confronted with the, often unavoidable, problem of correlated time traces. A nonparametric method that can address this new challenge is needed in order to interpret high frequency measurements.

To address this limitation, we propose an extension of iHMM with continuous control ICON,2 a generalization of the iHMM method previously developed to deal with time traces with drift (such as FRET), that is appropriate for correlated traces as they appear in FS. With the extended formulation presented here, we account for slow instrumentation responses, noise, and also drift. Additionally, we introduce computational improvements that facilitate the analysis of long traces such as those in FS. A working implementation of the resulting method, termed xi-ICON, is available in the supplementary material.

II. METHODS

Consider a FS single molecule time series $\mathbf{z} = (z_1, z_2, \ldots, z_N)$, where $z_n$ is the extension measured at time $t_n$, with $t_1$ and $t_N$ denoting the onset and conclusion of the experimental time course, respectively. As the timing accuracy in a typical detector is far better than the measurement accuracy, for example, less than few ps, in this study we ignore timing errors and model $t_n$ as noiseless and known in advance.

To extract the molecular signature from $\mathbf{z}$, which, as explained in Sec. I, may be masked by slow responses, noise, and drift, we develop a detailed model of the whole biophysical/experimental system. Our formulation, which relies on ICON2 and the iHMM,1 is described in Sec. II A and depicted schematically in Fig. 2. In Sec. II B, we present the statistical machinery that is required to derive posterior probabilities over multiple models and the associated parameters, and in Sec. II C we describe the acquisition of the datasets used in the examples that follow in Sec. III. For quick reference, a summary of our notational conventions can be found in Appendix A.

![FIG. 1. To resolve fast molecular dynamics, measurements are obtained at high acquisition rates. Because of the slow response time of the optical trap instrumentation used in FS, the molecular signal is masked (upper left). As a result, sharp molecular transitions are distorted or smeared out. Consequently, downsampling is often used to reduce (upper right) the induced correlations. Downsampling itself, however, introduces aliasing, which also distorts/smears out sharp transitions. In either case, naive application of BNP methods, such as iHMMs, leads to an overestimation of the number of molecular states (bottom), since artificial short-lived dwells (arising from either finite time trap responses or aliasing upon downsampling) are interpreted as visits to unphysical states.](image1)

![FIG. 2. For the analysis of the measurements, the molecule of interest may be represented as undergoing sudden changes. Here, the molecule (top) switches abruptly between a folded and an unfolded state. The signature of such spontaneous switching is contained in the extension time traces (bottom). The molecular signal (black) reflects the true molecular extension. Because of the slow response of the optical trap instrumentation, the molecular signal is distorted (red). In addition to the distortion, the measured signal (blue) is contaminated with unspecified drift components and corrupted by noise. In a typical analysis procedure, only the resulting signal (blue) is available and the main objective is the recovery of the underlying molecular signature (black).](image2)
A working implementation of the method described in this section can be found in the supplementary material.

A. Description

1. Molecular kinetics

At the single molecule level, similar to ICON\(^2\) and the initial iHMM,\(^1\) we assume that the system of interest alternates spontaneously between a number of distinct molecular states \(\sigma_1, \sigma_2, \ldots\), where \(\sigma_m\) is used to represent the biophysical states attainable over the course of the time trace. For example, for a protein undergoing conformation changes,\(^{16}\) as shown in Fig. 2, \(\sigma_1 = \text{“folded”} \quad \text{and} \quad \sigma_2 = \text{“unfolded”}.\) In general, the precise number of unique states \(\sigma_m\) available to the molecule is unknown and so we must also estimate it along the other quantities of the model that follow (see Sec. II B 1).

To facilitate the presentation, we let \(\bar{s} = (s_1, s_2, \ldots, s_N)\) be the sequence of successive states attained by the molecule during the experiment. Here, \(s_n\) denotes the state \(\sigma_m\) that the molecule attained at time \(t_n\). To model \(\bar{s}\), as in previous studies,\(^{1,10-12,21,22}\) we assume memoryless switching and adopt simple Markovian kinetics. In statistical notation, these kinetics can be formally expressed as \(s_{n+1}|s_n \sim \text{Categorical}(\pi_{\sigma_1 \rightarrow \sigma_n}, \pi_{\sigma_2 \rightarrow \sigma_n}, \ldots)\), where \(\pi_{\sigma_i \rightarrow \sigma_n}\) gathers the probabilities of the molecule switching from \(s_n\) to any \(\sigma_m\) at any time step.

The formulation we adopt here assumes that whenever the molecule departs from a state \(\sigma_m\) it arrives at a state \(\sigma_{m'}\) on average \(\pi_{\sigma_m \rightarrow \sigma_{m'}}\) of the times. So, \(\pi_{\sigma_m \rightarrow \sigma_{m'}} = 0\) indicates that \(\sigma_m\) never leads to \(\sigma_{m'}\), \(\pi_{\sigma_m \rightarrow \sigma_{m'}} = 1\) indicates that \(\sigma_m\) always leads to \(\sigma_{m'}\), while \(\pi_{\sigma_m \rightarrow \sigma_{m'}}\) describes self-transitions. The latter can be used to find\(^{23}\) the state’s mean dwell time \(T_{\sigma_m} = \Delta t/(1 - \pi_{\sigma_m \rightarrow \sigma_{m'}})\), where \(\Delta t\) is the time lapse between successive measurements. Similarly, transition rates\(^{23}\) between any pair of states can be found based on the probabilities \(\pi_{\sigma_m \rightarrow \sigma_{m'}}\).

2. Trap instrumentation response

Unlike the cases modeled by ICON and the iHMM,\(^1,2\) as the molecule switches from state to state producing \(\bar{s}\), it initiates the motion of the optical trap instrumentation that yields the trace \(\bar{t}\). To represent the driving signal, we consider \(\phi_{\sigma_m}\) to be the extension associated with state \(\sigma_m\) and the extension produced over the course of the experiment to be \(\bar{\phi} = (\phi_{s_1}, \phi_{s_2}, \ldots, \phi_{s_N})\).

To describe the dynamics of the trap system, we first consider the idealized case where noise is absent. For this case, we let \(d_n\) denote the distance that would be recorded by the trap at \(t_n\). For sufficiently fast trap systems, \(d_n\) is indistinguishable from the driving extension \(\phi_{s_n}\); however, for traps with response times comparable to the data acquisition rate, \(d_n\) may deviate considerably from \(\phi_{s_n}\) since the influence of past values \(d_{n-1}, d_{n-2}, \ldots\), etc., increases with increasing response times. For instance, for traps with exponentially decaying responses, such as those illustrated in Fig. 2, driving extensions and recorded extensions are related by \(d_n = \eta_0 \phi_{s_n} + \eta_1 d_{n-1}\), where \(\eta_0 = \Delta t/(t_{\text{resp}} + \Delta t)\) and \(\eta_1 = t_{\text{resp}}/(t_{\text{resp}} + \Delta t)\) with \(t_{\text{resp}}\) denoting the trap’s response time. In the case of traps with pronounced non-linearities in their response or traps incorporating feedback, generally more than one lag term, \(d_{n-1}, d_{n-2}, \ldots\) etc., is required in order to accurately represent \(d_n\). Since very little is typically known about the dynamics of the combined molecule/handles/bead/ambient fluid system, we model the response using a general autoregressive process\(^8\) having the form

\[
d_n = \eta_0 \phi_{s_n} + \eta_1 d_{n-1} + \eta_2 d_{n-2} + \cdots + \eta_K d_{n-K}
\]

and, to ensure that for static molecules the recorded extension is identical to the extension of the molecule, we set \(\eta_0 = 1 - \sum_{k=1}^{K} \eta_k\). We note that this condition also ensures “unit magnification” in the optical trap, that is, a change in the molecular extension is transmitted to a change in the recorded extension of the same magnitude, i.e., \(\Delta \phi = \Delta d\), and not to a constant multiple of it, i.e., \(\Delta \phi = c \Delta d\) for some \(c \neq 1\).

The assumption underlying Eq. (1) is that the extension \(d_n\) measured by the trap depends on (i) the current state of the molecule \(\phi_{s_n}\) and (ii) the extensions measured at the previous \(K\) time steps. A consequence is that \(d_n\) is determined not only by \(\phi_{s_n}\) but is also implicitly determined by the preceding extensions \(d_{n-1}, d_{n-2}, \ldots\). Therefore, the model trap inherits memory and the strength of the memory is ultimately determined by the values of the autoregressive coefficients \(\eta_1, \ldots, \eta_K\). In order to capture memory effects accurately, we learn the values of \(\eta_k\) directly from the data (see Sec. II B 1).

3. Noise

In addition to the driving molecular signal and the induced response of the trap system, the actual recorded extension during the experiment may also be influenced by unspecified perturbations such as mechanical forces, unresolved dynamics, perturbative ambient flows, and thermal fluctuations. Thus, a realistic description of the trap dynamics requires a modification of the extension \(d_n\) in Eq. (1) which excludes them.

For a well-calibrated experimental apparatus, the perturbations are expected to be small relative to the typical distances between the molecular extensions \(\phi_{s_n}\). So, we model these perturbations collectively as random fluctuations \(w_n\) having the characteristics of white noise. Accordingly, in our formulation, the actual extension \(x_n\) recorded by the trap at \(t_n\) is given by a modification of Eq. (1) of the form

\[
x_n = \eta_0 \phi_{s_n} + \eta_1 \phi_{s_{n-1}} + \cdots + \eta_K \phi_{s_{n-K}} + w_n,
\]

where the perturbations \(w_n \sim \text{Normal}(0, \tau)\) have zero mean and variance equal to \(1/\tau\). As a side note, we point out that unlike \(d_n\) mentioned earlier, which describes the idealized situation of absent random perturbations, \(x_n\) explicitly includes them.

4. Drift

Similar to ICON,\(^2\) to account for drift in the measured time series, which, as shown previously, may compromise the analysis if ignored,\(^2\) we couple the molecular trace \(\bar{s}\) to


an additional process \( \tilde{y} \) that explicitly incorporates drift. To accommodate drift, we place a set of nodes \( \theta_j = (B_j, D_j) \) over the time series and use interpolation\(^{24} \) to link across them as shown in Fig. 3. In particular, we place the nodes over the available time series \( \bar{z} \) at positions \( B_j \) and heights \( D_j \). To allow the model drift to fit tightly even near sharp changes or discontinuities, we allow both \( D_j \) and \( B_j \) to shift. To achieve this, we restrict each \( B_j \) within regions \( [B_j^{\text{min}}, B_j^{\text{max}}] \) and choose \( B_j^{\text{min}} \) and \( B_j^{\text{max}} \) from a regular grid of time levels that spans the entire time series.

In summary, the drift contribution to the measured extension at \( t_n \) is evaluated as

\[
y_n = G_{\bar{y}}(t_n), \quad (3)
\]

where \( G_{\bar{y}}(t) \) is the interpolating function\(^{24} \) that uses the nodes \( \bar{\theta} = (\theta_1, \theta_2, \ldots, \theta_j) \). Depending on the specifics of the experimental setup at hand, \( G_{\bar{y}}(t) \) may utilize splines or more specialized basis functions. For example, for drift with severe discontinuities, Hermite polynomials\(^{25} \) might provide superior fits than splines.

For well-designed FS experiments, drift is expected to fluctuate on a slower time scale than the molecular transitions. As a result, assuming separation of times scales in \( \bar{x} \) and \( \tilde{y} \), a sufficient condition to avoid degeneracies of our formulation is \( J/N \ll 1 \). That is, to be able to accurately capture the shape of the drift, the number of nodes \( J \) used in the interpolation needs to be lower than the total number of time points in the supplied trace \( N \). Generally, a precise upper bound on the ratio \( J/N \) is hard to predict as it depends on (i) the length of the trace \( N \), (ii) the intermeasurement period \( \Delta t \), and (iii) the intrinsic time scale of the drift. As a rule of thumb, we suggest \( J/N \approx 0.01 \) or lower.

Finally, to resolve non-identifiability issues caused because the drift may move perpendicularly,\(^2 \) we restrict the allowed combinations of nodes \( \theta_j \) to those that yield either a zero global contribution, i.e., \( \sum_{n} y_n = 0 \), or a zero contribution locally at the experiment onset, i.e., \( y_1 = 0 \).

5. Measurements

Finally, having incorporated trap instrumentation responses Eq. (1), random fluctuations Eq. (2), and drift Eq. (3), we model any remaining noise as having little to no additional effect. Thus, our net observations are \( z_n = x_n + y_n \).

A graphical model summarizing the dependencies among the variables of the developed formulation, leading from molecular switching to observations, is shown in Fig. 4. In Sec. II B, we describe methods to obtain estimates.

B. Inference

The parameters that we can infer from a time trace—quantities such as \( \bar{\sigma}_{x_t}, \phi_{f_{x_t}}, \) etc., introduced above—are directly related to the molecular properties of the system, e.g., molecular extensions, kinetics, and dwell times, that we wish to learn. To infer parameter values, we would normally follow the Bayesian paradigm.\(^{7,8,13} \) That is, we would ascribe priors to each parameter, combine priors with the likelihood implied by the formulation developed in Sec. II A 5, and subsequently derive a posterior probability distribution.\(^{7,8} \) However, since the number of attained molecular states is a priori unknown, so is the number of parameter values that need to be inferred. To handle this more general case, we use a nonparametric prior\(^ {1,3,5,19} \) that allows us to estimate models and their associated parameters simultaneously.

Below we describe the inference procedure for our formulation. A graphical summary is shown in Fig. 4.

1. Specifying the nonparametric prior

To infer a posterior over models incorporating a varying number of states in the molecular state space, we use a (nonparametric) generalization of the Dirichlet distribution\(^ {1,15,26,27} \) that is routinely used in applications involving HMMs.\(^ {8,28,29} \) In this formulation, the states available to the molecule are infinite in number and our primary goal is to identify which and how many of them have been visited during the experiment. As it is apparent, the majority of these states will remain...

\[ \text{FIG. 3. For the representation of drift, nodes } \theta_j = (B_j, D_j) \text{ are placed over the time interval between the first and last available measurements and interpolation is used to link across them. Here, arrows indicate that each node is adjusted perpendicular as well as parallel to the time axis. This way, allowing neighboring nodes to approach each other results in better fits near drift discontinuities. To ensure an } a \text{ priori regular spacing of the nodes over the experimental time course, each node position } B_j \text{ is restricted to lie between preset times } B_j^{\text{min}} \text{ and } B_j^{\text{max}} \text{ that are kept fixed. Furthermore, weak priors } P(B_j) \text{ are used to influence placement of the nodes near the center of the designated intervals.} \]

\[ \text{FIG. 4. For the analysis of the experimental data, a detailed statistical formulation of the combined molecule-trap system is developed. Namely, the molecule’s time course is represented by a Markov chain } x_n; \text{ the response of the trap’s instrumentation is represented by an autoregressive chain } s_n \text{ that is driven by } x_n; \text{ and drift is represented by interpolation } y_n. \text{ The measurements } z_n \text{ are directly influenced by the trap system and drift and indirectly by the molecule and noise. Here, blue marks the nonparametric prior placed on the molecular state space, red marks the prior on the trap dynamics, and cyan marks the prior for the drift interpolation. For details, see Sec. II.} \]
unvisited since, for example, the molecule can visit at most $N$ of them. In practice, through a careful choice of the prior on the state transitions (see below), most of the states will be revisited multiple times and so the number of unique states in $\tilde{\mathcal{S}}$ will be much lower than $N$.

Specifically, the prior we place on the molecular kinetics is a hierarchical Dirichlet process\textsuperscript{26} of the form

\begin{align}
\tilde{\beta} & \sim \text{GEM}(\gamma), \label{eq:beta_dist} \\
\tilde{\alpha}_{\sigma_m} | \tilde{\beta} & \sim \text{DP}(\xi, (1 - \rho)\tilde{\beta} + \rho \tilde{\delta}_{\sigma_m}), \label{eq:alpha_dist}
\end{align}

where $\tilde{\delta}_{\sigma_m}$ is the Dirac delta centered at $\sigma_m$ and $\gamma$, $\xi$, and $\rho$ are hyperparameters that, as explained below, determine the prior probability distributions on the number of molecular states and mean dwell times. In Eqs.\textsuperscript{(4)} and\textsuperscript{(5)}, GEM denotes the Griffiths-Engen-McCloskey stick breaking\textsuperscript{19,26,30} and DP denotes the Dirichlet process\textsuperscript{5,19,26,27,31,32} prior.

The unique characteristic of this prior is that when assigning molecular transitions, say from $s_n$ to $s_{n+1}$, it considers three possibilities: (i) the molecule stays in the same state, (ii) the molecule revisits one of the states visited before, and (iii) the molecule visits a state never visited before. Similar to ICON, it is because of the last possibility that, \textit{a priori}, the molecule has access to infinitely many states and it is up to the model’s likelihood to bias the transition toward options (i) and (ii) as more data are gathered, i.e., as $n$ moves to higher values. In other words, it is solely up to the likelihood to distinguish \textit{which and how many of the a priori} infinite states have been actually visited during the experiment. However, unlike the prior used in ICON, which places \textit{a priori} equal weights on these three possibilities, the present formulation places a higher weight on self-transitions as induced by the delta terms $\rho \tilde{\delta}_{\sigma_m}$ in Eq.\textsuperscript{(5)}. This makes the current formulation more appropriate in describing molecules exhibiting long dwells,\textsuperscript{33} for which multiple recurring self-transitions are required, while still maintaining an infinite pool of potential states.

In principle, the values of the hyperparameters $\gamma$, $\xi$, and $\rho$ can be inferred from the time series by invoking further hierarchical constructions.\textsuperscript{26,31,33,34} However, because prior information about the anticipated number of visited molecular states and characteristic time scales is often available—for example, either from previous experiments, computational modeling, or simply by intuition based on molecules of similar structure—such constructions may be unnecessary. Instead, below we describe how $\gamma$, $\xi$, and $\rho$ can be chosen to reflect the available information.

For a time series $\tilde{\mathcal{S}}$ containing $N$ measurements, provided $N$ is large as is typically the case in FS, it has been shown\textsuperscript{31} that the number of distinct molecular states to be visited, implied by the above prior, has a probability distribution that is approximately proportional to $\gamma^m (0.58 + \log N)^{m-1} (m - 1)!$. That is, when the only information about the molecule is that it has generated $N$ measurements, but the values of these measurements are unknown, the prior in Eqs.\textsuperscript{(4)} and\textsuperscript{(5)} implies that the number of unique states $\sigma_m$ in $\tilde{\mathcal{S}}$ is random and that its probability distribution is as given above. This relationship can be inverted to initialize a value for $\gamma = e^{\xi (M^*_{\text{mean}})} / (0.58 + \log N)$,

where $M^*_{\text{mean}}$ is the anticipated number of unique molecular states and $\psi$ is the digamma function.\textsuperscript{35}

Furthermore, the induced state mean dwell times $T_{\sigma_m}$, as shown in Appendix B, have a prior probability distribution that is illustrated in Fig.\textsuperscript{5}. This distribution has mean $T^*_{\text{mode}} = \Delta/(1 - \rho)$ and mode $T^*_{\text{mean}} = \Delta \xi / ((1 - \rho) \xi + 1)$, which may be used to choose appropriate values for the hyperparameters $\xi = T^*_{\text{mode}} T^*_{\text{mean}} / (T^*_{\text{mean}} - T^*_{\text{mode}}) / \Delta$ and $\rho = 1 - \Delta / T^*_{\text{mean}}$. More precisely, in this case, the mean sets the anticipated time scale of the molecular kinetics. In addition, a mode close to the mean indicates low uncertainty on the chosen time scale, as most of the prior probability is concentrated in a region around $T^*_{\text{mean}}$, while a mode distant from the mean indicates high uncertainty on the anticipated time scale, as the prior probability is spread over a wider region, Fig.\textsuperscript{5}.

At this point, it is worth noting that although the values of $M^*_{\text{mean}}$, $T^*_{\text{mean}}$, and $T^*_{\text{mode}}$ influence the prior, their effect on the resulting posterior—which incorporates information from the data introduced through the likelihood—decreases as the amount of data available increases or the noise level decreases.\textsuperscript{8,29}

For the remaining parameters, which may be treated within the parametric context, motivated by simplicity, we use normal priors for the molecular extensions $\phi_{\sigma_m} \sim \text{Normal}(\mu, \sigma^2_{\phi})$ and a gamma prior for the precision of the noise terms $\tau \sim \Gamma(\alpha/2, \alpha \sigma^2_{\tau}/2)$. Both choices are conjugate\textsuperscript{8} to their respective likelihoods. Consequently, their hyperparameters may be interpreted in terms of pseudo-observations,\textsuperscript{7,28} which in turn may be helpful in choosing appropriate values. Finally, we use normal priors for the autoregressive coefficients of the trap’s instrumentation response $\eta_k \sim \text{Normal}(\mu_\eta, \sigma^2_\eta)$, symmetric generalized beta priors for the placement of the drift nodes

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig5}
\caption{Anticipated time scales of the molecular kinetics can be represented from quantities appearing in the nonparametric prior. Here, the prior probability distribution of the mean dwell times $T_{\sigma_m}$ is shown for three choices of the hyperparameters $\xi$ and $\rho$. Vertical lines indicate the mean (solid) and mode (dashed) of the associated distributions. Choosing a large mean influences slower molecular switching (top), while choosing a smaller mean influences faster switching (middle) as most of the prior probability is located at lower values. Additionally, a mode close to the mean indicates strong beliefs about the chosen time scale (middle), while a mode distant from the mean indicates weak beliefs about the chosen time scale (bottom) as the prior probability is spread over a larger region.}
\end{figure}
$B_j \sim B^\text{min}_j + (B^\text{max}_j - B^\text{min}_j)\text{Beta}(\alpha_B, \alpha_B)$, and improper flat priors for the node heights $D_j \propto 1$. At this point, we want to mention that although these prior choices in our formulation are mainly motivated by modeling simplicity, due to the typical long length of FS traces, they have little influence on the resulting estimates though they remain necessary ingredients of Bayesian modeling.  

2. Obtaining estimates

In Bayesian methods, it is often possible to obtain posterior probabilities in analytic form; however, due to the nonparametric prior of Eqs. (4) and (5), this is impossible in our case. Instead, to compute the posterior $P(\hat{s}, \hat{\beta}, \hat{\tau}, \hat{\theta} | z)$, we have to adopt a computational approach. For this reason, we have developed a specialized scheme described in detail in Appendix C. Using this scheme, we can generate samples from $P(\hat{s}, \hat{\beta}, \hat{\tau}, \hat{\theta} | z)$ and subsequently obtain full estimates, such as expected values and credible intervals, of the involved physical quantities. 

For most parameters of interest—for example, molecular kinetics $\hat{\tau}$, extensions $\hat{\phi}$, and drift $\hat{\gamma}$—the evaluation of the corresponding estimators is straightforward. Some difficulties may arise only when trying to estimate the most probable state sequence $\hat{s}$ since the popular Viterbi sequence, common in engineering applications, cannot be easily generalized to the nonparametric context. To that end, following existing approaches, here we propose as an alternative estimator for $\hat{s}$ the state sequence $\hat{s}$ that lies closest to the expectation $E(\hat{s} | z)$ of the marginal posterior probability distribution $P(\hat{s} | z)$. Similar to the expectation $E(\hat{s} | z)$, the proposed $\hat{s}$ considers the full extent of the model posterior, however, unlike $E(\hat{s} | z)$ which might violate the model kinetics, the proposed $\hat{s}$ is ensured to be in agreement as it is chosen exclusively among the feasible sequences $\hat{s}$.

C. Example time series

To test and validate our method and demonstrate its applicability in the analysis of single molecule data, we have used synthetic as well as experimental traces.

For the synthetic traces, i.e., Figs. 6, 8, and 9, we simulated the switching of a hypothetical biomolecule with 5 states of characteristic extensions $10, 20, 30, 40, 50$ nm and a common mean dwell time of 0.01 s. These traces are generated for a total duration of 5 s, with individual measurements every $15 \mu$s. The generated molecular signals are distorted by (i) the response of a trap system with linear dynamics, (ii) random fluctuations, and (iii) drift that consists of linear and sinusoidal terms. Intentionally, for these traces, we used exceedingly slow response times $\tau_{\text{resp}} = 150 \mu$s, so the effect of the induced correlations may be clearly visualized. Random fluctuations introduced in these examples simulate a single noise source that combines every noise process occurring experimentally. Such noise processes include but are not limited to thermal fluctuations, unspecified mechanical perturbations, detection shot-noise, read-out noise, background signals, etc. In more detail, we have added fluctuations as described by Eq. (2) with perturbations $\nu_n$ that are normally distributed as one would expect in cases of noise collectively obeying the law of large numbers. Finally, to achieve a comparison with real-data, we have selected the variance (i.e., 1/$\tau$) of $\nu_n$ such that the signal-to-noise ratio in the resulting traces is similar to (Figs. 6 and 8) or much lower than (Fig. 9) the experimental trace used in this study (see below). The precise values of $\tau$ used are

![Image](https://via.placeholder.com/150)

FIG. 6. xl-ICON is used to analyze a synthetic single molecule trace (upper left) intended to represent real force spectroscopy measurements. To facilitate visual inspection, the synthetic signal intentionally contains long trap instrumentation responses (response time of 150 $\mu$s) and low noise levels ($\tau = 4$ nm$^2$s$^{-1}$). We show the estimated sequence of molecular extensions and drift in the entire trace (lower left) and the estimated trap response (upper and lower right). For clarity, we omit showing true molecular and drift traces since they overlap with the estimated ones. Note that axes are drawn to different scales.
4 nm$^2$ and 0.04 nm$^2$ for the high and low signal-to-noise cases, respectively.

The optical trap instrument used to obtain the experimental trace, i.e., Fig. 10, was a home-built timeshared dual optical trap constructed and operated generally as described in previous studies.\textsuperscript{44,45} In brief, an acousto-optic modulator rapidly deflects a 1064 nm laser between two positions to create two independent optical traps (66.7 kHz modulation rate). Time traces were obtained with the traps held at fixed positions. A pair of beads was held in the dual traps and tethered together \textit{in situ} by a single molecule consisting of an 89-bp dsDNA hairpin flanked by two 1.5-kbp dsDNA “handles.” The construct and hairpin sequence are the exact same as “hairpin 1” described earlier.\textsuperscript{45,46} The change in the positions of the pair of trapped beads is the readout of the changing tether length and thus the hairpin folded/unfolded configuration. The tether tension was chosen so as to observe the long DNA hairpin reversibly unfold and refold in multiple steps (four clear steps are observed in this case). Hairpins such as these are often used as test substrates for investigations of helicase protein machines. Bead position data were recorded unfiltered at 66.7 kHz sampling rate (i.e., one data point per individual trap on cycle or every 15 \textmu s) for a total of 5 s. Because the measured trace appeared void of drift, we artificially contaminated it with the addition of drift similar to the synthetic traces.

III. RESULTS

We implemented xl-ICON as described in Appendix C and used it to analyze the single molecule time series described above. The full set of options used in these analyses is listed in Appendix D.

A. Synthetic measurements

Figure 6 shows a sample synthetic trace with low noise as well as a collection of characteristic estimates. In this trace, the simulated trap instrumentation response consists of a single exponential decay. However, because in the experimental setting the trap’s response is generally uncharacterized, we have used a broader 3 order autoregressive model, i.e., $K = 3$, to represent it. Despite the intentional misspecification, the estimated coefficients $\eta_1 = 0.91, \eta_2 \approx 0, \eta_3 \approx 0$ indicate that the model accurately identifies the correct autoregressive order. The corresponding estimate of the response time is $t_{\text{resp}} = 150.8 \pm 1.1 \text{ s}$, which is in agreement with the simulated one of 150 s. Furthermore, xl-ICON also accurately identifies and removes drift which is also shown in Fig. 6. Overall, in the estimated state trace $\hat{\gamma}$, xl-ICON misses or miss-identifies less than 1.5% of the transitions (missed transitions are not shown for clarity).

Figure 7 shows the combinations of state extensions $\phi_{\sigma_m}$ and mean dwell times $T_{\sigma_m}$ estimated through the posterior probability distribution. Here, color coded is the average fraction of time that the molecule is estimated to occupy a state $\sigma_m$ with the corresponding extension and mean dwell time. For example, a fraction of 10% indicates that the molecule produced in total 10% of the measurements in the time trace while occupying state $\sigma_m$; further, a null fraction indicates that the molecule never occupied the corresponding combination during the experiment. As expected, the estimated occupation fraction clusters around the simulated combinations of extensions of 10, 20, 30, 40, 50 nm and dwell time 0.01 s indicating that xl-ICON successfully identifies and localizes the correct states.

We emphasize that the success of xl-ICON in clearly identifying 5 states at the correct extensions, despite the severe correlations in the time series analyzed, relies heavily on the explicit incorporation of the trap instrumentation response (Sec. II A 2) that is the novelty of our method. To demonstrate how critical this is, we analyzed the same trace without considering the response. This is achieved, for example, by keeping the coefficients $\eta_1, \ldots, \eta_K$ in Eqs. (1) and (2) fixed at zero. Figure 8 demonstrates how the estimates deteriorate in this case. By not accounting for the induced correlations, which is effectively equivalent to applying ICON or a plain iHMM, the method estimates an unrealistic total of $\approx$40 states with characteristic extensions spread throughout the full range of the data.

To facilitate the presentation, for the examples shown so far, we have intentionally utilized low noise levels. Nevertheless, it is worth noting that xl-ICON provides robust estimates even at exceedingly high noise. For instance, Fig. 9 shows corresponding estimates obtained from a considerably noisier trace than those used earlier. As can be seen, the estimated molecular extensions and drift remain in excellent agreement with the true ones despite the significant overlap in the state emissions. Further, good agreement is also maintained with the estimated molecular trace $\hat{\gamma}$. In the latter, missed transitions account for less than 10% of the entire trace. Although this is already a low error rate, we want to point out that the molecular traces shown throughout are only point estimates $\hat{\gamma}$ (see Sec. II B 2) that represent the single best choices out of every possible trace $\bar{\gamma}$ in the sense that they simultaneously
optimize the number of occupied states, state kinetics, state extensions, as well as drift, trap response, and noise globally over the entire time series. So a non-zero error rate is expected to occur when the supplied data are noisy. Such errors typically disappear once the estimator $\hat{s}$ is enclosed within the appropriate credible region, $i.e.$, traces that engulf a significant portion of the posterior probability $P(\tilde{s} | \bar{z})$. In principle, this region may be obtained, nonetheless, in a computationally intensive manner. $^{38}$ We avoid the required computations here, since most commonly the main task in the analysis of single molecule measurements is the quantification of the molecular properties, such as state extensions $\phi_{m}$ and mean dwell times $T_{m}$, rather than the precise quantification of the molecular trace $\tilde{s}$ itself.

**B. Experimental measurements**

Having validated our model using synthetic data with known ground truth in Sec. III A, we now turn to real FS data where the ground truth is unknown. Given that the assumptions underlying our method (see Sec. II) are expected to hold in the experimental context—for example, Markovian molecular kinetics and white noise perturbations of trap’s recorded extension—it is anticipated that xl-ICON performs equally well on experimental traces as well.

Figure 10 shows the experimental trace and selected estimates. During the acquisition of the trace shown, the optical trap has been configured such as the underlying DNA hairpin folds/unfolds in discrete steps that can be observed in the recorded extension (for further details, see Sec. II C). Similar to the synthetic traces shown earlier, the method successfully isolates the underlying molecular transitions from the drift and trap responses, despite the high noise levels and the correlations in the raw trace. As an extra layer of validation, we note here that since the drift in this time series is known, as it is introduced artificially, we verify the accuracy of its estimate, Fig. 10.

Figure 11 illustrates the estimated state extensions and mean dwell times. As can be seen, xl-ICON clearly identifies 4 distinct molecular states with extensions varying between $-25$ and $10$ nm. Unlike the synthetic trace in Fig. 7, which ascribed similar kinetics to every state, here one state exhibits relatively stable kinetics (mean dwell time of $\approx 0.1$ s), while the remaining states are more transient (mean dwell times between $1$ ms and $0.1$ ms). To facilitate the interpretation of these estimates, in Fig. 11 we also show a simple histogram of the experimental data (without including drift). A visual inspection of this histogram reveals the presence of at least three states. All of them have been accurately identified by xl-ICON. Further, since xl-ICON considers also kinetic properties which are lost when the data are collapsed to produce a histogram, it also reveals the presence of a fourth one at $\approx 4.5$ nm. This state is barely discernible from the histogram; however, its presence in the experimental trace can be
verified from the blow ups of the time courses presented in Fig. 10.

At this point, a comparison of the estimates obtained through xl-ICON with those that may be obtained visually through simple downsampling of the raw data is appropriate. In Fig. 12, we show histograms of the experimental trace at full rate 15 µs (left most panel) and successively downsampled prior to binning by a factor of 10 until a rate of 15 ms. The state located near 6 µm, due to its slow kinetics (≈100 ms) that exceed by several orders the inter-measurement times (15 µs – 15 ms) and the trap response (≈60 µs), is sharpened as expected. However, the remaining states, due to their transient kinetics (≈0.1–1 ms), contain only a small fraction of the points in the entire time series. As a result, a mild downsampling may enhance them (for example, at 0.15 ms, the lower state is enhanced, and also the middle two states can be visualized as the histogram’s middle peak spreads asymmetrically suggesting at least two states). Nevertheless, with further
downsampling these transient states are completely eliminated as they merge into a single one located at around $-12$ nm. Besides causing transient states to merge, downsampling also introduces artifactual states (for example, states at extensions between $-10$ nm and $10$ nm that may be seen at $15$ ms). These stem from the mixing (aliasing) of the transient states with the stable one. For all these cases, \textbf{xl-ICON} provides superior estimates as (i) it considers state kinetics that are lost when the trace is binned and thus resolves noisy states robustly, (ii) it resolves fast molecular kinetics that are comparable to the experimental acquisition rate without downsampling, and (iii) it avoids aliasing.

IV. DISCUSSION

High data acquisition rates make it possible to capture increasingly faster dynamics of biomolecules. However, the interpretation of those datasets presents unique challenges that are not fully met by existing analysis methods both because such data sets are larger in size but also because the data sets themselves are noisier and more complex.

For the sake of concreteness, in this study, we focused on force spectroscopy (FS) datasets collected at high rates that fall within the low kHz timeframe. Although these datasets present a unique opportunity to observe molecular dynamics at faster scales, their interpretation is challenging because fast molecular motions are distorted by continuous aberrations, such as drift and correlation, which pose a severe limitation not only for traditional (parametric) approaches which specify models ahead of time but also for naive nonparametric methods.\textsuperscript{1–3} For this reason, the experimental measurements are often averaged down or otherwise artificially reduced to render them analyzable. However, this strategy is appropriate only for measurements from biomolecules with relatively slow kinetics that could be also captured in datasets obtained at lower acquisition rates. However, fundamentally, averaging or reducing the collected datasets undermines the whole reason behind collecting data at higher rates in the first place.

In part to resolve emergent modeling challenges and in part to maintain a feasible computational cost (as high rate datasets may be vastly larger), existing analysis methods often require crucial user dependent choices that yield subjective or partial conclusions. In the extreme example of the hidden Markov model (HMM), the workhorse of biophysical time series data analysis,\textsuperscript{10–14} the model itself (i.e., the number of attained states) is commonly specified by hand and extensive data pre-processing is used to remove drift and correlations. Subsequently, to discriminate between plausible models, different model candidates are compared using additional criteria.\textsuperscript{4,6} The challenge here is that pre-processing (e.g., to remove drift) and/or post-processing (e.g., to select the number of states) yields an output that strongly depends on multiple independent steps in the analysis procedure.\textsuperscript{1,6}

To that end, Bayesian nonparametrics (BNPs)\textsuperscript{1–4} offer powerful and feasible alternatives to obtaining globally optimal model and parameter estimates. BNPs do so by determining model parameters and models themselves simultaneously thereby relaxing many of the theoretical requirements of traditional (parametric) approaches. More specifically, infinite hidden Markov models (iHMMs) can relax the requirements of the traditional HMMs to a bare minimum,\textsuperscript{5,5,19} though they remain more computationally costly than HMMs. Nevertheless, current research into efficient algorithmic alternatives to the slower Monte Carlo techniques typically employed, for example, variational methods,\textsuperscript{47–36} may reduce computational cost in the years to come. As an example, \textbf{xl-ICON}, which utilizes the same computational tools\textsuperscript{33,34} as the iHMMs, requires $\approx 2$ days of computational time, in an average desktop computer, to analyze a trace of $5$ s collected at $66.7$ kHz (for example, similar to the trace of Fig. 10). Although, such computational requirements are certainly within the capabilities of modern labs, anticipated computational improvements could reduce it by several orders in the near future.

Our analysis method directly addresses challenges presented by data collected at high rates by exploiting BNPs. By utilizing BNPs, we present the first strategy, to the best of our knowledge, in the single molecule literature toward making raw experimental datasets collected at high rates amenable to rigorous analysis. The resulting method represents the entire experimental system encountered in force spectroscopy and allows clear interpretation of the involved processes, for example, molecular transitions, trap instrumentation responses, and drift. As a result, our formulation of \textbf{xl-ICON}—but also the logic behind our formulation—may be used for unsupervised time series analysis of raw single molecule data or used as a building block for models of larger biophysical/experimental setups.

SUPPLEMENTARY MATERIAL

See supplementary material for a software implementation of \textbf{xl-ICON}.

ACKNOWLEDGMENTS

S.P. acknowledges support from NSF CAREER Grant No. MCB-1719537. M.J.C. acknowledges support from NSF Grant No. MCB-1514706.

APPENDIX A: NOTATION

Table I summarizes the notation used. Throughout this study, $n = 1, 2, \ldots, N$ indexes time levels; $m = 1, 2, \ldots, M$ indexes molecular states; $k = 1, 2, \ldots, K$ indexes autoregressive coefficients; and $j = 1, 2, \ldots, J$ indexes drift nodes. To group similar variables, we use bars to denote lists over time, e.g., $\bar{z} = (z_1, z_2, \ldots, z_N)$; tildes to denote lists over the molecular state space, e.g., $\tilde{\phi} = (\phi_{r_1}, \phi_{r_2}, \ldots, \phi_{r_M})$; double tildes to denote lists of lists over the molecular state space, e.g., $\tilde{\tilde{\phi}}$; right arrows to gather the trap’s autoregressive coefficients, e.g., $\tilde{\eta} = (\eta_1, \eta_2, \ldots, \eta_K)$; and left arrows to gather the drift nodes, e.g., $\tilde{\theta} = (\theta_1, \theta_2, \ldots, \theta_J)$. Occasionally, we use carets to denote estimators, e.g., $\hat{\phi}$.

The notation “$X \sim F$” where $X$ is a random variable and $F$ is a probability distribution, that is used extensively throughout this study, indicates\textsuperscript{7–9,36} that $X$ is sampled from
\[
\Delta t \quad \text{Measurement period (i.e., inverse acquisition rate)}
\]

\[
t_{\text{resp}} \quad \text{Response time of trap’s instrumentation}
\]

\[
t_n \quad \text{Time levels of individual measurements}
\]

\[
z_n \quad \text{Extension measured at } t_n
\]

\[
y_n \quad \text{Drift at } t_n
\]

\[
x_n \quad \text{Recorded position at } t_n \text{ (i.e., with noise)}
\]

\[
y_d \quad \text{Idealized position at } t_d \text{ (i.e., without noise)}
\]

\[
u_n \quad \text{Trap perturbation at } t_n
\]

\[
s_n \quad \text{Molecular state at } t_n
\]

\[
\sigma_m \quad \text{Distinct molecular state}
\]

\[
\phi_{\sigma_m} \quad \text{Extension associated with } \sigma_m
\]

\[
T_{\sigma_m} \quad \text{Mean dwell time of } \sigma_m
\]

\[
\pi_{\sigma_m \rightarrow \sigma_m'} \quad \text{Transition probability from } \sigma_m \text{ to } \sigma_m'
\]

\[
\eta_k \quad \text{Trap response autoregressive coefficient}
\]

\[
\theta_i \quad \text{Drift node}
\]

\[
\beta_i \quad \text{Time position of drift node } \theta_i
\]

\[
D_j \quad \text{Extension height of drift node } \theta_j
\]

\[
G_d(i) \quad \text{Drift interpolating function}
\]

\[F \text{ or simply that } X \text{ follows the probability measure associated with } F.\]

**APPENDIX B: PRIORS FOR STATE MEAN DWELL TIMES**

According to the prior in Eq. (5), the probabilities of a self-transition, i.e., \( \pi_{\sigma_m \rightarrow \sigma_m} \), and a non-self transition, i.e., \( \pi_{\sigma_m \rightarrow \sigma_m'} \), are jointly distributed according to

\[
(\pi_{\sigma_m \rightarrow \sigma_m}, \pi_{\sigma_m \rightarrow \sigma_m'}) \sim \text{Dir}(\xi(1-\rho), \xi(1-\rho)(1-\beta_{\sigma_m})) \quad (B1)
\]

since \( \sum_{m' \neq m} \beta_{\sigma_m} = 1 - \beta_{\sigma_m} \). Here, Dir denotes the Dirichlet distribution.\(^8,9,29\) Given that \( \pi_{\sigma_m \rightarrow \sigma_m} = 1 - \pi_{\sigma_m \rightarrow \sigma_m} \), the above joint probability may be reduced to

\[
\pi_{\sigma_m \rightarrow \sigma_m} \sim \text{Beta}(\xi(1-\rho), \xi(1-\rho)) \quad (B2)
\]

where Beta denotes the beta distribution.

Finally, due to the Markov property of the state transitions,\(^9\) the number of dwell time steps \( n_{\sigma_m} \) in state \( \sigma_m \) follows a geometric probability distribution with mean \( N_{\sigma_m} = 1/(1 - \pi_{\sigma_m \rightarrow \sigma_m}) \). In turn, according to Eq. (B2), \( N_{\sigma_m} \) follows an inverse beta distribution

\[
N_{\sigma_m} \sim \text{Beta}^{-1}(\xi(1-\rho), \xi(1-\rho)) \quad (B3)
\]

Therefore, the mean dwell time \( T_{\sigma_m} = N_{\sigma_m} \Delta t \) in state \( \sigma_m \) follows a scaled inverse beta distribution and the precise probability density is

\[
p(T_{\sigma_m} | \tilde{\beta}) = \frac{\Delta t}{T_{\sigma_m}} \frac{\xi(1-\rho) + 1}{\text{B}(\xi(1-\rho), \xi(1-\rho))} \frac{1 - \Delta t / T_{\sigma_m}}{\xi(1-\rho)} \quad (B4)
\]

where B denotes the beta function.\(^35\)

**APPENDIX C: COMPUTATIONAL SCHEME**

A fully working version of xl-ICON, in Matlab source code and GUI formats, is available in the supplementary material. The implementation utilizes the sampler briefly mentioned in Sec. II B 2 and generates samples from the model’s augmented posterior probability distribution \( P(\tilde{z}, \tilde{\beta}, \tilde{\kappa}, \tilde{\phi}, \tau, \tilde{\eta}, \tilde{\theta}) \) through an Markov Chain Monte Carlo (MCMC) scheme\(^51\) that utilizes Gibbs sampling with embedded Metropolis updates\(^31\).

Specifically for the nonparametric components related to the molecular state space, as in previous studies,\(^33,34,49\) we use a weak limit approximation that replaces Eq. (4) with

\[
\tilde{\beta} \sim \text{Dir}_{\sigma_m} \left( \frac{\gamma}{M^\sigma_1}, \frac{\gamma}{M^\sigma_2}, \ldots, \frac{\gamma}{M^\sigma_M} \right) \quad (C1)
\]

Results from the theory of statistics\(^32,34,49\) can be invoked to ensure that for sufficiently large \( M \), Eq. (C1) yields estimates that are indistinguishable from Eq. (4). The main advantage of adopting this approximation is that we can invoke efficient machine learning algorithms, such as forward filtering\(^1,29\) without modifications,\(^33\) in order to minimize the associated computational cost. Additionally, sampling from Eq. (C1), compared to sampling from Eq. (4), for example, through the popular beam sampler\(^52\) as in previous implementations,\(^1,2\) leads to faster mixing of the MCMC chain.\(^33,34\) The latter is crucial in the analysis of large traces such as those encountered in force spectroscopy.

In detail, our implementation generates the outlined samples by iterating the following steps:

**Step 1:** Update the molecular states \( \tilde{z} \) using forward filtering-backward sampling.\(^1\) Briefly, in a forward pass of the trace, compute recursively the state probabilities \( A_n(\sigma_m) = P(s_n = \sigma_m | z_1, \ldots, z_n) \) according to the filtering relation \( A_n(\sigma_m) = P(s_n | \sigma_m) \sum_{m'} A_{n-1}(\sigma_{m'}) \pi_{\sigma_{m'} \rightarrow \sigma_m} \), while in a backward pass of the trace, sample states according to \( \tilde{s}_n | s_{n+1} \sim \pi_{\sigma_{n+1} \rightarrow \sigma_n} A_n(\tilde{s}_n) \). The resulting state trace \( \tilde{z} = (s_1, \ldots, s_N) \) can be shown\(^1,28,52\) to follow the conditional probability \( P(\tilde{z} | \tilde{\beta}, \tilde{\kappa}, \tilde{\phi}, \tau, \tilde{\eta}, \tilde{\theta}, \tilde{z}) \) that is required by the overall Gibbs sampling scheme.\(^51\) A detailed exposition of the algorithm, which lies beyond the scope of this study, can be found in the description of ICON.\(^1,2\)

**Step 2:** Update the nonparametric base \( \tilde{\beta} \), molecular kinetics \( \tilde{\kappa} \), and hyperparameters \( \gamma, \rho, \) and \( \xi \) (if necessary), similar to the blocked sampler of the sticky-HDP-HMM.\(^53\)

**Step 3:** Update the molecular extensions \( \tilde{\phi} \).

**Step 4:** Update the noise precision \( \tau \).

**Step 5:** Update the autoregressive extensions \( \tilde{\eta} \) using a Metropolis-within-Gibbs update.\(^51\)

**Step 6:** Update the drift nodes \( \tilde{\theta} \) using a Metropolis-within-Gibbs update similar to ICON’s sampler.\(^2\)

Note that since \( \tilde{\phi} \) and \( \tau \) utilize conjugate prior distributions,\(^8\) sampling in steps 3 and 4 is achieved directly. Also, note that to improve mixing\(^51\) (i) steps 3–6 are swept in random order, (ii) additional repetitions of steps 4 and 5 are embedded
within steps 3 and 6, and (iii) steps 2–6 are repeated several times in each iteration. Finally, the proposals for the Metropolis updates of steps 5 and 6 are adjusted during burn-in to yield individual acceptance rates of ≈25%.

Fine details of the outlined implementation can be found in the xl-ICON’s source code, available in the supplementary material.

APPENDIX D: EXAMPLES

Table II summarizes the options used in Sec. III. For the computations, a total of 45 000 MCMC samples are generated for each example. From these, the initial 1000 are subtracted as burn-in, while, from the remaining, only 1 out of every 64 is used for the results shown in Figs. 6–11.