Generalizing HMMs to Continuous Time for Fast Kinetics: Hidden Markov Jump Processes

by Zeliha Kilic, Ioannis Sgouralis and Steve Pressé (last updated July 28, 2020)

1 Introduction **2**

2 Methods **4**
2.1 Model Description . . . 5
2.1.1 Dynamics . . . 5
2.1.2 Measurements . . . 6
2.1.3 Simulation . . . 7
2.2 Model Inference . . . 8
2.2.1 Model Inference via HMMs . . . 8
2.2.2 Model Inference via HMJPs . . . 10
2.2.3 Computational Inference . . . 11

3 Results **11**
3.1 Data Simulation . . . 11
3.2 Analysis with HMJPs . . . 12
3.3 Comparison of HMJPs with HMMs . . . 12

4 Discussion **17**

A Supporting Material **22**
A.1 Additional Analyses . . . 22
A.1.1 Comparison of HMJP with HMM for Slow Switching Rates . . . 22
A.1.2 Comparison of HMJP with HMM for Mixed Switching Rates . . . 22
A.1.3 Robustness Analysis for HMJP Posterior Rate Estimates for Fast and Mixed Kinetics . . . 26
A.1.4 Robustness Analysis with Respect to Data Set Length for HMM Posterior State Level and Transition Probability Estimates . . . 26
A.1.5 Robustness Analysis with Respect to Learning the Trajectory Simultaneously with Transition Probability Estimates for HMJPs . . . 31
A.1.6 Comparison of the Posterior Trajectory Estimates of HMM and HMJP . . . 31
A.2 Detailed Description of the Statistical Models . . . 34
A.2.1 Hidden Markov Jump Process . . . 34
A.2.1.1 Summary of Equations . . . 34
A.2.1.2 Description of the Computational Scheme . . . 34
A.2.2 Hidden Markov Model . . . 35
A.2.2.1 Summary of Equations . . . 35
A.2.2.2 Description of the Computational Scheme . . . 35
A.2.2.3 Overview of the Sampling Updates . . . 36
A.3 Notation . . . 41

Correspondence: *spresse@asu.edu*

Abstract

The hidden Markov model (HMM) is a framework for time series analysis widely applied to single molecule experiments. It has traditionally been used to interpret signals generated by systems, such as single molecules, evolving in a discrete state space observed at discrete time levels dictated by the data acquisition rate. Within the HMM framework, originally developed for applications outside the Natural Sciences, such as speech recognition, transitions between states, such as molecular conformational states, are modeled as occurring at the end of each data acquisition period and are described using transition probabilities. Yet, while measurements are often performed at discrete time levels in the Natural Sciences, physical systems evolve in continuous time according to transition rates. It then follows that the modeling assumptions underlying the HMM are justified if the transition rates of a physical process from state to state are small as compared to the data acquisition rate. In other words, HMMs apply to slow kinetics. The problem is, as the transition rates are unknown in principle, it is unclear, *a priori*, whether the HMM applies to a particular system. For this reason, we must generalize HMMs for physical systems, such as single molecules, as these switch between discrete states in *continuous time*. We do so by exploiting recent mathematical tools developed in the context of inferring Markov jump processes and propose the hidden Markov jump process (HMJP). We explicitly show in what limit the HMJP reduces to the HMM. Resolving the discrete time discrepancy of the HMM has clear implications: we no longer need to assume that processes, such as molecular events, must occur on timescales slower than data acquisition and can learn transition rates even if these are on the same timescale or otherwise exceed data acquisition rates.

1 Introduction

Hidden Markov models (HMMs) have been important tools of time series analysis for over half a century (1–3). Under some modeling assumptions, detailed shortly, HMMs have been used to self-consistently determine dynamics of physical systems under noise and the properties of the noise obscuring the system's dynamics itself.

Originally developed for applications in speech recognition (4, 5), the relevance of HMMs to single molecule time series analysis was quickly recognized (6–14). Since then, HMMs and variants have successfully been used in the interpretation of single ion channel patch clamp data (15–19); fluorescence resonant energy transfer (FRET) (20–28); force spectroscopy (29–31); amongst many other physical applications (4, 32, 33).

In order for HMMs to apply to single molecules and other physical systems, the assumptions underlying the HMMs must hold for such systems. There are several such assumptions worthy of consideration.

- HMMs assume that the system under study evolves in a discrete state space (whether physical of conformational). This is a reasonable approximation for biomolecules visiting different conformations (33–35) or fluorophores visiting different photo-states (33, 34, 36). Of parallel interest to this point is the notion that the number of discrete states is known (though the transition probabilities between states is unknown). The assumption of a known number of states has been lifted thanks to extensions of the HMM (34, 37–43) afforded by nonparametrics that we discuss elsewhere (33, 34, 38–40, 43).
- HMMs assume that measurements are obtained at discrete time levels. That is, successive measurements are reported at regular time levels separated by some fixed period Δt. For clarity, we call Δt the *data acquisition period*. This assumption is consistent with a number of experimental biophysical settings (11, 12, 38–40, 44–54).
- HMMs assume that physical systems transition between states in discrete time steps. Put differently, HMMs apply under the implicit assumption that the underlying system switches between states *rarely* as compared to the data acquisition period, Δt. This can only really be assured if the transition rates (as required in a continuous time description) are slow. This assumption is implicit to the very definition of

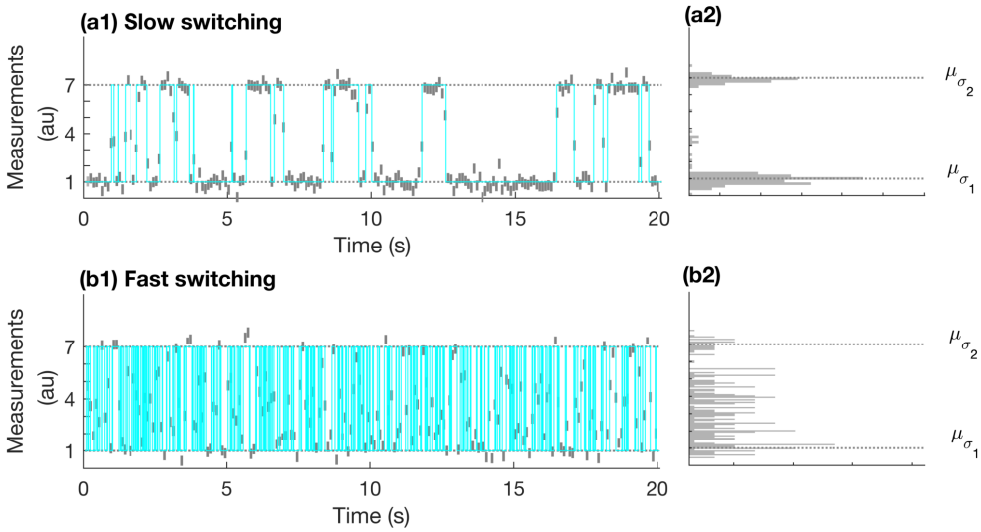

Fig. 1. ***A conceptual illustration of single molecule continuous time switching kinetics between discrete states probed in discrete time***. *For illustrative purposes only, the trajectory of a single molecule between two states (σ_1, σ_2) is shown in cyan in panel (a1) and (b1). For concreteness, we can think of these molecular states as conformational states. The state levels, i.e., signal level in the absence of noise, for these states is μ_{σ_1} and μ_{σ_2}, respectively, and shown by the horizontal gray dotted line in (a1) and (b1). This synthetic experiment starts at time $t_0 = 0.05$ s, and ends at time $t_N = 20$ s and the data acquisition period, Δt, is 0.1 s. Next, again only for illustration, we assume that the measurements are acquired by a detector that has a fixed integration period $\tau = 90$ ms (light gray) for each fixed data acquisition period shown in panels (a1) and (a2). As a molecule may switch between states during an integration period, the measurements represent the signal levels capturing the average of the amount of time spent at each state levels ($\mu_{\sigma_1}, \mu_{\sigma_2}$) visited (in addition to added noise). Panels (a1)-(a2) are associated with single molecule kinetics that are slower than the data acquisition rate. Instead, in panels (b1)-(b2) we show a single molecule trajectory when a molecule's kinetics are faster than the data acquisition rate. In panel (b1), slow kinetics result in well separated state occupancy histograms around the average state levels. In panel (b2), we don't have well separated histograms centered around the average state levels due to fast kinetics of the molecule.*

a HMM which requires that the system's switching occurs *precisely* at the end of the data acquisition periods (4, 38, 39, 55–57).

This last assumption is problematic and presents the following conundrum: on the one hand, the transition rates are unknown and their analog for discrete time processes (transition probabilities) are typically to be determined using HMMs. On the other hand, we must assume that these unknowns are slow as compared to the data acquisition rate. Even if, optimistically, transition rates are slow, molecular events themselves are stochastic and, as with all physical processes, occur in continuous time. As such, any one event has a probability of occurring on timescales faster than the data acquisition period.

As an example, Fig. 1 illustrates the types of dynamical measurements that we can and cannot analyze within the HMM paradigm. The top panel shows an example of single molecule measurements characterized by slow kinetics that can be analyzed within an HMM paradigm. In contradistinction to the above is the bottom panel illustrating an example of fast kinetics, as compared to the data acquisition rate, that cannot currently be analyzed within the HMM paradigm. The reason for this is simple: the fast kinetics give rise to a large number of apparent states that go beyond the two real states. This is because the measurements reported at each time point averages the molecular signal from all states visited in each acquisition period. Yet it is clear that the information on the transition rates between the rapidly switching molecular states is encoded in the time trace however uninterpretable. It is therefore natural to propose an analysis method that treats physical processes as they occur in *continuous* time in order to extract rates directly from traces with fast kinetics without relying on the artificial assumption that the physical processes involved occur on timescales much slower than the data acquisition period.

To do so, we must fundamentally upgrade both key ingredients of the HMM model: 1) the system dynamics must be in continuous time; and 2) the measurement output must realistically reflect an average over the dynamics of the system over the data acquisition period. The output is then understood to encode fast dynamics, that can be retrieved (54, 58–62).

It is indeed to address processes evolving in continuous time that continuous time Markov models, so-called Markov jump processes (MJPs), were developed (44, 63–65). MJPs describe continuous time events using rates (rather than transition probabilities) and recent advances in computational statistics (66–71), have made it possible to learn these rates given data. However, an important challenge remains for us to infer MJP rates under the assumption that the measurement process averages the probed signal over each measurement period. The nature of the measurement process and continuous dynamics therefore suggest a Hidden MJPs (HMJPs) framework that we put forward herewith.

In Section 2, below, we start with the formulation of our HMJP model and also, briefly, summarize the HMM. Next, in Section 3, we move on to the head-to-head comparison of HMMs and HMJPs (showing in what limit the HMM exactly reduces to the HMJP). We focus on their respective performance in learning molecular trajectories and transition probabilities. We show how HMJPs successfully outperform HMMs especially for kinetics occurring on timescales on the order of or exceeding the data acquisition period. Finally, in Section 4, we discuss the broader potential of HMJPs to Biophysics. Fine details on the implementation of these two methods can be found in Supporting Material (A).

2 Methods

In this section we describe a physical system that evolves in continuous time alongside a measurement model. We also discuss how to generate realistic synthetic data from such a model and subsequently analyze time traces reflecting both fast and slow dynamics. We analyze the traces using two different methods: HMMs, as they are broadly used across the literature, and our proposed HMJPs which we describe in detail. We compare the analyses in Section 3.

2.1 Model Description

Using the experimental data, and the model of the experiment that we will describe, our goal is to learn: 1) the switching rates between the states of the system; 2) the state of the system at any given time (that we call the trajectory of the system); 3) initial conditions of the system; and 4) parameters describing the measurement process, i.e., parameters of the emission distribution.

2.1.1 Dynamics

We start by defining the trajectory $\mathcal{T}(\cdot)$ that tracks the state of the system over time. Here $\mathcal{T}(t)$ is the state of the system at time t and, as such, it is a *function* over the time interval $[t_0, t_N]$. We adopt functional notation and distinguish between $\mathcal{T}(\cdot)$ and $\mathcal{T}(t)$ to avoid confusion with the entire trajectory and the value attained at particular time levels, critical to the ensuing presentation.

We label the states to which the system has access with σ_k and use the subscript $k = 1, \ldots, K$ to distinguish them. For example, σ_1/σ_2 may represent a protein in folded/unfolded conformation or an ion channel in on/off state. With this convention (borrowed from (68)) if the system is at σ_k at time t, then we write $\mathcal{T}(t) = \sigma_k$. An early point of distinction with the HMM is warranted: the switching times between states of the system, generally, differ from the measurement acquisition levels t_n.

As with most molecular systems (35, 38, 72, 73), the switching dynamics are faithfully modeled as memoryless. That is, the waiting time in a state of the system is exponentially distributed. Such memoryless systems are termed *Markov jump process* and below we present their mathematical formulation.

At the experiment's onset, we assume the state of the system $\mathcal{T}(t_0)$ is chosen stochastically among σ_k. We use ρ_{σ_k} to denote the probability of the system starting at σ_k and collect all initial probabilities in $\bar{\rho} = (\rho_{\sigma_1}, \rho_{\sigma_2}, ..., \rho_{\sigma_K})$, which is a probability vector (74–77).

Memoryless switching kinetics are described by *switching rates* between all possible state pairs. These switching rates are labeled with $\lambda_{\sigma_k \to \sigma_{k'}}$. By definition, all self-switching rates are zero $\lambda_{\sigma_k \to \sigma_k} = 0$, which, in general, allows for at most $K(K-1)$ non-zero rates (35). Although, the switching rates $\lambda_{\sigma_k \to \sigma_{k'}}$ fully describe the system's kinetics, as we will see shortly, it is mathematically more convenient to work with an alternative parametrization. In this alternative parametrization, we keep track of the *escape rates*

$$\lambda_{\sigma_k} = \sum_{k'=1}^{K} \lambda_{\sigma_k \to \sigma_{k'}} \tag{1}$$

which, for simplicity, we gather in $\bar{\lambda} = (\lambda_{\sigma_1}, \lambda_{\sigma_2}, \ldots, \lambda_{\sigma_K})$. Further, instead of keeping track of each rate $\lambda_{\sigma_k \to \sigma_{k'}}$, we keep track of the rates normalized by the escape rates, namely

$$\pi_{\sigma_k \to \sigma_{k'}} = \frac{\lambda_{\sigma_k \to \sigma_{k'}}}{\lambda_{\sigma_k}}. \tag{2}$$

Gathering all normalized rates out of the same state in $\bar{\pi}_{\sigma_k} = (\pi_{\sigma_k \to \sigma_1}, \pi_{\sigma_k \to \sigma_2}, \cdots, \pi_{\sigma_k \to \sigma_M})$, we see that each $\bar{\pi}_{\sigma_k}$ forms a probability vector (76).

In summary, instead of $K(K-1)$ switching rates $\lambda_{\sigma_k \to \sigma_{k'}}$, we describe the system's kinetics with K escape rates λ_{σ_k} and K switching probability vectors $\bar{\pi}_{\sigma_k}$. The latter have, by definition, $\pi_{\sigma_k \to \sigma_k} = 0$, and so, the total number of scalar parameters is the same in both parametrizations. Below, for simplicity, we gather all transition

probability vectors into a matrix

$$\bar{\bar{\pi}} = \begin{pmatrix} \bar{\pi}_{\sigma_1} \\ \bar{\pi}_{\sigma_2} \\ \vdots \\ \bar{\pi}_{\sigma_K} \end{pmatrix}. \tag{3}$$

2.1.2 Measurements

The overall input to our method consists of the measurements $\mathbf{x} = (x_1, x_2, ..., x_N)$ acquired in an experiment. Here, x_n indicates the nth measurement and, for clarity, we assume measurements are time ordered, so $n = 1$ labels the earliest acquired measurement and $n = N$ the latest. These measurements may be camera ADUs generated from photon detections, FRET efficiencies, derived inter-molecular extensions, or any other quantity determinable in an experiment.

Each x_n is reported at a time $t_n = t_{n-1} + \Delta t$, which is Δt later than the time t_{n-1} at which the previous measurement x_{n-1} was reported. For completeness, together with the time levels $t_1, t_2, \ldots, t_N$ at which a measurement is reported, we also consider an additional time level t_0, that marks the onset of the experiment, which is not associated with any measurement, Fig. 1.

The most common assumption made almost universally by HMMs is that the instantaneous state of the system at t_n determines x_n. Yet, for realistic detectors, the reported value x_n is influenced by the entire trajectory of our system during the nth integration period which we represent by the time window $[t_n - \tau, t_n]$. Here, τ is the duration of each integration time (such as an exposure period for optical experiments).

We account for detector features in the generation of the measurements via characteristic state levels which we label with μ_{σ_k} and, for simplicity, gather these in $\bar{\mu} = (\mu_{\sigma_1}, \mu_{\sigma_2}, \ldots, \mu_{\sigma_K})$. In our formulation, each σ_k is associated with its own characteristic level μ_{σ_k}. If the system remains at a single state σ_k throughout an entire integration period $[t_n - \tau, t_n]$, then the detector is triggered by μ_{σ_k} and so, provided measurement noise is negligible, the reported measurement x_n is similar to $\mu_{\sigma_k}\tau$. However, if the system switches multiple states *during* an integration period, the detector is influenced by the levels of every state attained and the time spent in each state.

More specifically, the nth signal level triggering the detector during the nth integration period, $[t_n - \tau, t_n]$, is obtained from the time average of $\mu_{\mathcal{T}(\cdot)}$ over this integration period. Mathematically, this time average equals $\frac{1}{\tau}\int_{t_n-\tau}^{t_n} dt\, \mu_{\mathcal{T}(t)}$ and, provided measurement noise is negligible, the reported measurement x_n is similar to the value of this average.

In the presence of measurement noise, such as shot-noise (78–82), quantification noise (83–85), or other degrading effects common to detectors currently available, each measurement x_n depends *stochastically* upon the signal that triggers the detector (33, 86, 87). Of course, the precise relationship depends on the detector employed in the experiment and differs between the various types of cameras, single photon detectors or other devices used. To continue with our formulation, we assume that measurement noise is additive, which results in

$$x_n \big| \mathcal{T}(\cdot) \sim \mathbf{Normal}\left(\frac{1}{\tau}\int_{t_n-\tau}^{t_n} dt\, \mu_{\mathcal{T}(t)}, v\right). \tag{4}$$

The latter expression is a statistical shorthand for the following: the measurement x_n is a random variable that is sampled from a normal distribution whose mean is $\frac{1}{\tau}\int_{t_n-\tau}^{t_n} dt\, \mu_{\mathcal{T}(t)}$ and whose variance is v. For the Normal distribution, the variance is related to the detector's full-with-at-half-maximum (FWHM) by $v = \frac{\text{FWHM}^2}{8\log 2}$.

Our choice of a Normal distribution itself is incidental (and can be changed depending on detector type). However, this type of measurement model is general enough to capture the effect of the history of the system during the detector's integration time. For example, in an accompanying article (88), we adapt Eq. (4) to FRET measurements in separate donor and acceptor channels with shot-noise and background as follows

$$x_n^D \big| \mathcal{T}(\cdot) \sim \textbf{Poisson}\left(\mu_{\text{back}}^D \tau + \int_{t_n-\tau}^{t_n} dt\, \mu_{\mathcal{T}(t)}^D\right)$$
$$x_n^A \big| \mathcal{T}(\cdot) \sim \textbf{Poisson}\left(\mu_{\text{back}}^A \tau + \int_{t_n-\tau}^{t_n} dt\, \mu_{\mathcal{T}(t)}^A\right)$$

where $\mathbf{x}^D = \left(x_1^D, x_2^D, \ldots, x_N^D\right)$ and $\mathbf{x}^A = \left(x_1^A, x_2^A, \ldots, x_N^A\right)$ denote the measurements acquired in the donor's and the acceptor's channels, respectively. Here, $\mu_{\text{back}}^D, \mu_{\text{back}}^A$ denote the background and $\mu_{\sigma_k}^D, \mu_{\sigma_k}^A$ the characteristic state levels in the two channels.

2.1.3 Simulation

Given $\bar{\rho}$ and $\bar{\lambda}, \bar{\bar{\pi}}$, a trajectory $\mathcal{T}(\cdot)$ that mimics real systems may be simulated using the Gillespie algorithm (72) which we describe briefly here only in an effort to introduce necessary notation.

To begin, an initial state s_0 is chosen among $\sigma_1, \sigma_2, \cdots, \sigma_K$ with probability ρ_{σ_k}. Then, the period d_1 that the system spends in s_0 is chosen from the Exponential distribution with mean $1/\lambda_{s_0}$. Subsequently, the next state s_1 is chosen among $\sigma_1, \sigma_2, \cdots, \sigma_K$ with probability $\pi_{s_0 \to \sigma_k}$. As $\pi_{s_0 \to s_0} = 0$, any chosen s_1 is different from s_0, therefore the transition $s_0 \to s_1$ is a jump in the system's time course that occurs at time $t_0 + d_1$. Next, a new period d_2 is sampled from an Exponential distribution with mean $1/\lambda_{s_1}$ and a new state s_2 is chosen among σ_k with probability $\pi_{s_1 \to \sigma_k}$, and so on. These steps are repeated until the end of the experiment which, in our setup, is the same as the time t_N of the last measurement.

More formally, we summarize the sampling of a Gillespie trajectory as follows

$$s_0 \sim \textbf{Categorical}\left(\bar{\rho}\right) \tag{5}$$
$$d_m | s_m \sim \textbf{Exponential}\left(\lambda_{s_m}\right) \tag{6}$$
$$s_{m+1} | s_m \sim \textbf{Categorical}\left(\bar{\pi}_{s_m}\right) \tag{7}$$

for $m = 0, 1, 2, ..., M-1$ where $M-1$ is the lowest value such that

$$t_0 + \sum_{m=0}^{M-1} d_m \geq t_N. \tag{8}$$

The Categorical distribution we use here is the generalization of the Bernoulli distribution for which more than two outcomes are possible (55).

The successive states of the system $s_0, s_1, \cdots, s_{M-1}$ and the associated durations $d_0, d_1, d_2, \cdots, d_{M-1}$, which we term *holding states* and *holding times*, respectively, encode $\mathcal{T}(\cdot)$ throughout the experiment's time course $[t_0, t_N]$. Namely

$$\mathcal{T}(t) = \begin{cases} s_0 & \text{if} \quad t_0 \leq t < t_0 + d_0 \\ s_1 & \text{if} \quad t_0 + d_1 \leq t < t_0 + d_0 + d_1 \\ \vdots & \vdots \\ s_{M-1} & \text{if} \quad t_0 + d_0 + \cdots + d_{M-2} \leq t < t_0 + d_0 + d_1 + \cdots + d_{M-1}. \end{cases} \tag{9}$$

For convenience, we summarize the representation of $\mathcal{T}(\cdot)$ in a triplet $(\vec{S}, \vec{D}, M)$, where

$$\vec{S} = \{s_0, s_1, ..., s_{M-1}\}, \qquad \vec{D} = \{d_0, d_1, ..., d_{M-1}\}.$$

and M is the size of $\vec{S}$ and $\vec{D}$.

Once a trajectory is obtained through the Gillespie algorithm just described, the signal levels $\int_{t_n-\tau}^{t_n} dt\, \mu_{\mathcal{T}(t)}$ for each integration period can be computed. For instance, as the trajectory is piecewise constant, the integrals reduce to sums that can be easily calculated. Therefore, given an appropriate detector model, such as Eq. (4), and a trajectory's triplet $(\vec{S}, \vec{D}, M)$, we can obtain simulated measurements by adding noise according to the detector's distribution.

2.2 Model Inference

Using the data $\mathbf{x}$, and the model of the experiment that we have just described, our goal is now to learn initial probabilities ρ_{σ_k}, switching rates $\lambda_{\sigma_k \to \sigma_{k'}}$, and state levels μ_{σ_k} for all states as well as the trajectory of the system $\mathcal{T}(\cdot)$ throughout the experiment's time course $[t_0, t_N]$. Below we attempt to learn these using time series analysis with a HMM and then introduce a novel time series analysis relying on HMJPs.

2.2.1 Model Inference via HMMs

An HMM requires that each measurement x_n depends exclusively on the *instantaneous* state of the system, namely $\mathcal{T}(t_n)$. In view of Eq. (4), this is achieved by the trajectory of the system $\mathcal{T}(\cdot)$ remaining constant during the integration period $[t_n - \tau, t_n]$. To a sufficiently good approximation, this is satisfied provided

$$\tau \lambda_{\sigma_k} \ll 1 \tag{10}$$

for all σ_k. Thereby, the system rarely exhibits switching during periods that last shorter than τ. This approximation allows for $\frac{1}{\tau}\int_{t_n-\tau}^{t_n} dt \mu_{\mathcal{T}(t)} \approx \mu_{\mathcal{T}(t_n)}$ to be used. Accordingly, in a HMM, Eq. (4) is replaced with

$$x_n | \mathcal{T}(\cdot) \sim \mathbf{Normal}\left(\mu_{\mathcal{T}(t_n)}, v\right). \tag{11}$$

Again, as with Eq. (4), the exact choice of distribution (whether Normal or otherwise) is incidental. HMMs can treat any emission distribution *provided* x_n only depends on $\mathcal{T}(t_n)$ as opposed to the full history of the trajectory over the integration time.

With the measurements described by Eq. (11), we can use a HMM to learn the probabilities of the transitions $\mathcal{T}(t_{n-1}) \to \mathcal{T}(t_n) \to \mathcal{T}(t_{n+1})$. For clarity, from now on, we will use $c_n = \mathcal{T}(t_n)$ and denote these transitions with $c_{n-1} \to c_n \to c_{n+1}$. That is, c_n is the state of the system *precisely* at the time t_n.

For an HMM, transition probabilities are denoted with $P_{c_{n-1} \to c_n}$. Since the system can attain K different states σ_k, in general, a HMM possesses $K \times K$ transition probabilities $P_{\sigma_k \to \sigma_{k'}}$. Now, we gather the transition probabilities out of the same σ_k in a vector $\bar{P}_k = (P_{\sigma_k \to \sigma_1}, P_{\sigma_k \to \sigma_2}, \cdots, P_{\sigma_k \to \sigma_K})$ and, for clarity, gather all of the vectors in a matrix

$$\bar{\bar{P}} = \begin{pmatrix} \bar{P}_{\sigma_1} \\ \bar{P}_{\sigma_2} \\ \vdots \\ \bar{P}_{\sigma_K} \end{pmatrix}. \tag{12}$$

The matrix $\bar{\bar{P}}$ is related to the system's switching rates $\lambda_{\sigma_k \to \sigma_{k'}}$ and escape rates λ_{σ_k}. Specifically, if we gather them in

$$\bar{\bar{G}} = \begin{bmatrix} -\lambda_{\sigma_1} & \lambda_{\sigma_1 \to \sigma_2} & \dots & \lambda_{\sigma_1 \to \sigma_K} \\ \lambda_{\sigma_2 \to \sigma_1} & -\lambda_{\sigma_2} & \dots & \lambda_{\sigma_2 \to \sigma_K} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{\sigma_K \to \sigma_1} & \lambda_{\sigma_K \to \sigma_2} & \dots & -\lambda_{\sigma_K} \end{bmatrix} \tag{13}$$

which is termed *generator matrix* (76, 89), then $\bar{\bar{P}}$ is obtained by

$$\bar{\bar{P}} = \exp\left(\bar{\bar{G}}\Delta t\right) \tag{14}$$

where $\exp(\cdot)$ denotes the matrix exponential. We point out that $\bar{\bar{\pi}}$ and $\bar{\bar{P}}$ are both probability matrices, however, they assume quite *different* properties. For instance, $\pi_{\sigma_k \to \sigma_k} = 0$ while $P_{\sigma_k \to \sigma_k} > 0$.

Although knowing $\bar{\bar{G}}$ is sufficient to specify $\bar{\bar{P}}$, the inverse is *not* true: knowing $\bar{\bar{P}}$ does not necessarily lead us to a unique $\bar{\bar{G}}$ and so the switching rates *cannot* simply be inferred from $\bar{\bar{P}}$. This is a consequence of the multivalued nature of the logarithm. As such, one transition probability matrix may corresponds to multiple rate matrices (89). Instead, provided $\lambda_{\sigma_k}\Delta t \ll 1$ for all σ_k, we may approximate Eq. (14) by

$$\bar{\bar{P}} \approx \bar{\bar{I}} + \bar{\bar{G}}\Delta t \tag{15}$$

where $\bar{\bar{I}}$ is the identity matrix of size $K \times K$. Under this approximation, we can estimate transition rates by $\bar{\bar{G}} \approx (\bar{\bar{P}} - \bar{\bar{I}})/\Delta t$.

Below, we highlight the steps necessary to estimate the quantities of interest in a HMM. Specifically, a HMM relies on the statistical model

$$c_0 \sim \mathbf{Categorical}\left(\bar{\rho}\right) \tag{16}$$

$$c_{n+1}\big|c_n \sim \mathbf{Categorical}\left(\bar{P}_{c_n}\right) \tag{17}$$

$$x_n\big|c_n \sim \mathbf{Normal}\left(\mu_{c_n}, v\right). \tag{18}$$

To model the full distribution over the quantities of interest, e.g., initial probabilities, $\bar{\rho}$, transition probabilities $\bar{P}_{\sigma_k}$, state levels $\bar{\mu}$ and the trajectory of the system $\mathcal{T}(\cdot)$ which is encoded by $\vec{c} = (c_0, c_1, \dots, c_N)$, we follow the *Bayesian paradigm* (74, 90). Within this paradigm, we place prior distributions over the parameters and we discuss the appropriate choices next.

On the transition probabilities $\bar{P}_{\sigma_k}$, we place a Dirichlet prior with concentration parameter A

$$\bar{P}_{\sigma_1} \sim \mathbf{Dirichlet}\left(\frac{A}{K}, \frac{A}{K}, ..., \frac{A}{K}\right) \tag{19}$$

$$\bar{P}_{\sigma_2} \sim \mathbf{Dirichlet}\left(\frac{A}{K}, \frac{A}{K}, ..., \frac{A}{K}\right) \tag{20}$$

$$\vdots \tag{21}$$

$$\bar{P}_{\sigma_K} \sim \mathbf{Dirichlet}\left(\frac{A}{K}, \frac{A}{K}, ..., \frac{A}{K}\right) \tag{22}$$

which is conjugate to the Categorical distribution (38, 39, 41, 77). We consider a similar prior distribution, with concentration parameter α, also for the initial transition probability $\bar{\rho}$, namely

$$\bar{\rho} \sim \mathbf{Dirichlet}\left(\frac{\alpha}{K}, \frac{\alpha}{K}, ..., \frac{\alpha}{K}\right). \tag{23}$$

Subsequently, we place priors on the state levels $\bar{\mu} = (\mu_{\sigma_1}, \mu_{\sigma_2}, ..., \mu_{\sigma_K})$. The prior that we choose is the conjugate Normal prior

$$\mu_{\sigma_k} \sim \mathbf{Normal}\left(H, V\right) \tag{24}$$

with hyperparameters H, V.

Once the choices for the priors are made, we then form the posterior distribution (34, 38–43)

$$\mathbb{P}\left(\bar{\rho}, \bar{\bar{P}}, \bar{\mu}, \mathcal{T}\left(\cdot\right)\left|\mathbf{x}\right.\right) = \mathbb{P}\left(\bar{\rho}, \bar{\bar{P}}, \bar{\mu}, \vec{c}\left|\mathbf{x}\right.\right) \tag{25}$$

containing all unknown variables that we wish to learn. As the posterior distribution in Eq. (25) does not attain an analytical form, we develop a specialized computational scheme exploiting Markov Chain Monte Carlo (MCMC) to generate pseudorandom samples from this posterior. We explain the details of this scheme in Section 2.2.3.

2.2.2 Model Inference via HMJPs

HMJP apply directly on the formulation of Section 2.1.2 and, unlike with HMM (see Eq. (10)), no approximations are required on the system switching kinetics. Therefore, in order to proceed with inference, we need only provide appropriate prior distributions on the parameters, namely $\bar{\rho}, \bar{\bar{\pi}}, \bar{\lambda}, \bar{\mu}$.

We start with the prior distribution for the escape rates $\bar{\lambda} = (\lambda_{\sigma_1}, \lambda_{\sigma_2}, ..., \lambda_{\sigma_K})$. We put priors on each of the λ_{σ_k} for all $k = 1, 2, ..., K$. The prior we select is

$$\lambda_{\sigma_k} \sim \mathbf{Gamma}\left(\eta, \frac{b}{\eta}\right) \tag{26}$$

for all $k = 1, 2, ..., K$ with hyperparameters η, b. We note that this prior is conjugate to the exponential distribution given in Eq. (6). Next, we place a prior on $\bar{\pi}_{\sigma_k}$ for all $k = 1, 2, ..., K$. For this, we place independent conjugate Dirichlet priors with concentration parameter A such that $\pi_{\sigma_k \to \sigma_k} = 0$ holds for all $k = 1, 2, ..., K$

$$\bar{\pi}_{\sigma_1} \sim \mathbf{Dirichlet}\left(0, \frac{A}{K-1}, ..., \frac{A}{K-1}\right) \tag{27}$$

$$\bar{\pi}_{\sigma_2} \sim \mathbf{Dirichlet}\left(\frac{A}{K-1}, 0, ..., \frac{A}{K-1}\right) \tag{28}$$

$$\vdots$$

$$\bar{\pi}_{\sigma_K} \sim \mathbf{Dirichlet}\left(\frac{A}{K-1}, \frac{A}{K-1}, ..., 0\right). \tag{29}$$

Finally, on $\bar{\rho}$ and $\bar{\mu}$ we place the same prior distributions as in Eq. (23) and Eq. (24), respectively.

Once the choices for the priors are made, we then form the posterior distribution

$$\mathbb{P}\left(\bar{\rho}, \bar{\bar{\pi}}, \bar{\lambda}, \bar{\mu}, \mathcal{T}\left(\cdot\right)\left|\mathbf{x}\right.\right) = \mathbb{P}\left(\bar{\rho}, \bar{\bar{\pi}}, \bar{\lambda}, \bar{\mu}, (\vec{S}, \vec{D}, M)\left|\mathbf{x}\right.\right) \tag{30}$$

containing all unknown variables that we wish to learn. As, once more, the posterior distribution does not attain an analytical form, we develop a specialized computational scheme exploiting MCMC. We explain the details of this scheme in Section 2.2.3.

2.2.3 Computational Inference

We carry out the analyses, shown in Section 3, evaluating the associated posteriors with an MCMC scheme relying on Gibbs sampling (34, 38–43, 68). The overall sampling strategy, for either the HMM or the HMJP, is as follows:

1. Update the trajectory $\mathcal{T}(\cdot)$, that is $\vec{c}$ for HMM or $(\vec{S}, \vec{D}, M)$ for HMJP;
2. Update the kinetics, that is $\bar{P}_{\sigma_k}$ for HMM or $\bar{\pi}_{\sigma_k}$ and λ_{σ_k} for HMJP;
3. Update the initial probabilities $\bar{\rho}$;
4. Update state levels $\bar{\mu}$.

We repeat these updates to obtain a large number of samples. The end result is a sampling of the posterior $\mathbb{P}\left(\bar{\rho}, \bar{\bar{P}}, \bar{\mu}, \vec{c}|\mathbf{x}\right)$ for the HMM and $\mathbb{P}\left(\bar{\rho}, \bar{\bar{\pi}}, \bar{\lambda}, \bar{\mu}, (\vec{S}, \vec{D}, M)|\mathbf{x}\right)$ for the HMJP. Both samplings can be used to estimate switching rates $\lambda_{\sigma_k \to \sigma_{k'}}$; for example, HMJP by Eq. (2) and HMM by Eq. (15).

In the Supporting Material (A.2) we provide a thorough description of all steps. We also freely provide a working code through the authors' website.

3 Results

To demonstrate how HMJPs work and highlight their advantages over HMMs, in this section, we use synthetic data that mimic a single molecule experiment. Synthetic data are ideal for this purpose as they allow us to benchmark the results against the exact, readily available, "ground truth". We obtain such data from the Gillespie algorithm described in Section 2.1.3 and we explain our simulation choices below.

We focus on two datasets: one where the system exhibits slow kinetics and another where the system exhibits fast kinetics as compared with data acquisition, see Fig. 1. We provide the values for the hyperparameters in all analyses, as well as any other choices made, in Supporting Material (A.3). To be clear, we only assume to have access to the data, i.e., the gray dashes of panels (a1) and (b1) of Fig. 1. The cyan (ground truth) trajectories are assumed unknown and to be determined.

In our results, we first benchmark the HMJP on the easy (i.e., slow kinetics) case shown in Fig. 1 panel (a1); see Figs. 2 and 3. This is the regime where the HMM also works well and the expected (good) results for the HMM are relegated to the appendix; see Supporting Material (A.1). Next, we turn to the more complex case of fast kinetics. A sample time trace is shown in Fig. 1 panel (b1). The results for both the HMJP and HMM are shown in Figs. 4 and 5.

3.1 Data Simulation

To simulate the synthetic data, we assumed $K = 2$ distinct states, such as on/off or folded/unfolded states for illustrative purposes only. We assumed well separated state levels which we set at $\mu_{\sigma_1} = 1$ au and $\mu_{\sigma_2} = 7$ au where au denote arbitrary units. The prescribed detector FWHM was set at 0.25 au.

Additionally, for sake of concreteness only, we assumed an acquisition period of $\Delta t = 0.1$ s and consider long integration periods by setting τ equal to 90% of Δt. In terms familiar to microscopists, our setting corresponds to a frame rate of 10 Hz with exposure time of 90 ms and a dead time of 10 ms (40). The onset and concluding

time of the experiment are the same for all simulated measurements and set at $t_0 = 0.05$ s and $t_N = 20$ s, respectively.

To specify kinetics, we use the following structure for the switching rates $\lambda_{\sigma_1 \to \sigma_2}, \lambda_{\sigma_2 \to \sigma_1}$, with a parameter τ_f which sets the timescale of the system kinetics,

$$\lambda_{\sigma_1 \to \sigma_2} = \frac{1.1}{\tau_f}, \qquad \lambda_{\sigma_2 \to \sigma_1} = \frac{1.6}{\tau_f}. \tag{31}$$

We simulate a case with $\tau_f = 0.8$ s, which involves system kinetics that are slower than the data acquisition rate; and a case with $\tau_f = 0.067$ s, which involves system kinetics that are faster than the data acquisition rate.

3.2 Analysis with HMJPs

As a benchmark, we provide the results for the HMJP for those measurements shown on Fig. 1 panel (a1) associated with slow switching rates. These results include estimates of the trajectory $(\vec{S}, \vec{D}, M)$ (see Fig. 2), state levels $\bar{\mu}$ (see Fig. 3) and the switching rates $\bar{\lambda}$ (see Fig. 3). To obtain these estimates, we generate samples from the posterior distribution $\mathbb{P}(\bar{\rho}, \bar{\bar{\pi}}, \bar{\lambda}, \bar{\mu}, (\vec{S}, \vec{D}, M) | \mathbf{x})$ with the HMJP sampler of Section 2.2.3.

In Fig. 2 panel (a1), the ground truth trajectory is shown in cyan while the measurements are shown in gray. We showed the zoomed trajectory and observations in panel (b). We also provide the empirical histogram of the observation in panel (a2) highlighting the slow switching rates of the system. After determining the posteriors over the trajectories with HMJPs, for illustrative purposes, we only show the maximum a posteriori (MAP) trajectory in Fig. 2 panel (b). We observe that the HMJP MAP trajectory (magenta) captures most of the fast switches, shown in Fig. 2 panel (b), in the system trajectory. In Fig. 3, there are four panels. In these four panels, we provide the superposed posterior distributions over the two state levels and two rates estimated by the HMJP along with its associated 95% confidence interval and ground truths (dashed green lines).

In summary: HMJP performs well on this benchmark data. The same is true of the simpler HMM (as would be expected) whose results are shown in Figs. A.1 and A.2. An important bring home message for the HMJP however is the fact that even if state transitions occur midway through the integration time, the HMJP can discern when these occurred. The same is not true of the HMM that, as mentioned earlier, assumes by construction that state transitions must occur at the end of the data acquisition period.

3.3 Comparison of HMJPs with HMMs

We now present a comparison of HMJPs and HMMs on the analysis of the simulated measurements shown in Fig. 1 panel (b1). We expect the HMJP to outperform the HMM as we are now operating in a regime, with switching rates 2.5 times faster than earlier, where the HMM requirement, spelled out in Eq. (10), breaks down.

We used these measurements to estimate the posterior distribution over the trajectory, $\mathcal{T}(\cdot)$, initial and switching probabilities, $\bar{\rho}$ and $\bar{\bar{\pi}}$ or $\bar{\bar{P}}$, state levels, $\bar{\mu}$ and escape rates, $\bar{\lambda}$. To accomplish this we generate samples from the posterior distributions $\mathbb{P}(\bar{\rho}, \bar{\bar{\pi}}, \bar{\lambda}, \bar{\mu}, (\vec{S}, \vec{D}, M) | \mathbf{x})$ and $\mathbb{P}(\bar{\rho}, \bar{\bar{P}}, \bar{\mu}, \vec{c} | \mathbf{x})$ using the HMJP and HMM samplers of Section 2.2.3, respectively.

Following the pattern from the previous section on slow kinetics, we first show the trajectories inferred by HMJPs and HMMs in Fig. 4, then we show estimates of the state levels and transition probabilities in Fig. 5. In particular, the escape rates estimated from HMJPs are used in Eq. (15) to yield transition probabilities that we subsequently compared with the transition probabilities inferred by HMMs.

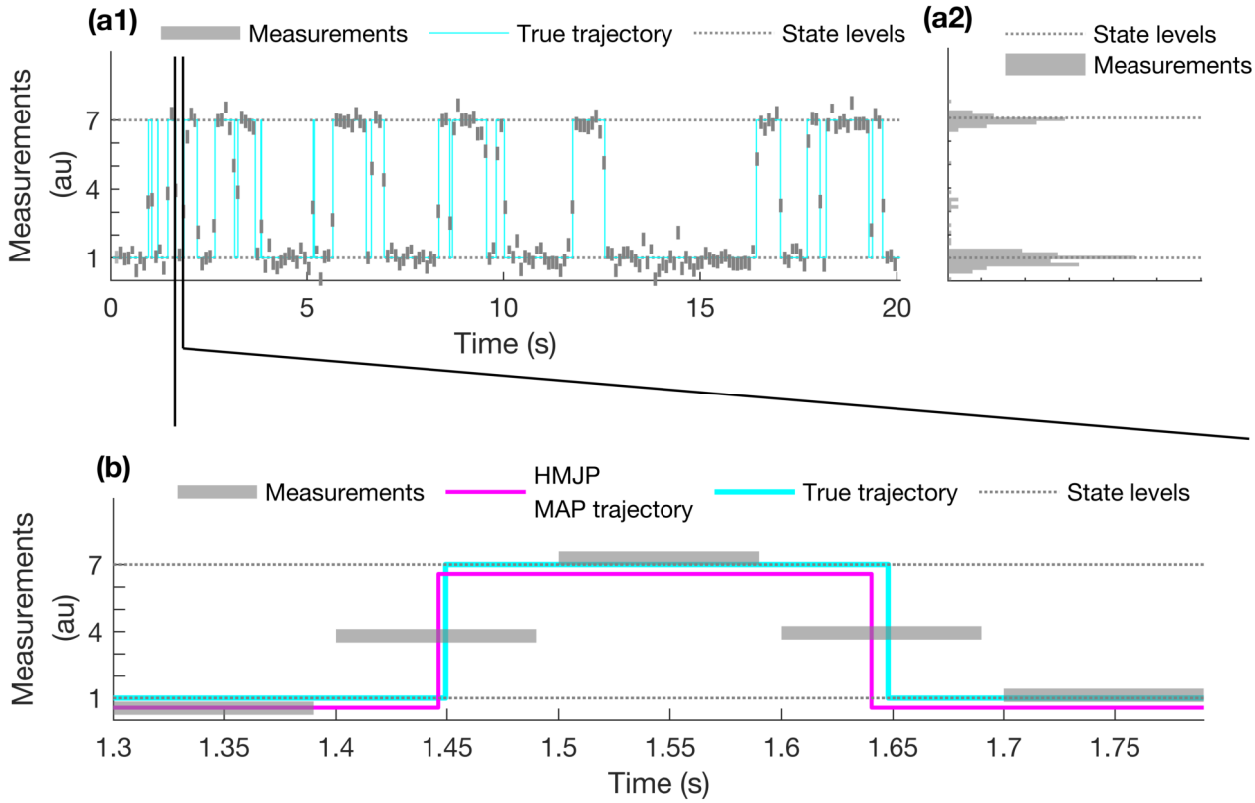

Fig. 2. ***HMJP trajectory estimates for slow state switching***. *Here we provide trajectory estimates obtained with the HMJP when the switching rate is slower than the data acquisition rate,* $1/\Delta t = 10$ *(1/s). In this figure's panel (a1), the measurements are shown as gray rectangles (the width of the rectangle coincides with the integration period as shown in Fig. 1) generated based on the description provided in Section 2.1. We superposed the true trajectory (cyan) with the measurements in panel (a). Next, in panel (a2), we provide the histogram of all measurements to visualize the system kinetics. For illustrative purposes, we only show the MAP estimates of the HMJP on a zoomed in region of panel (a1). Next, we provide that region of the panel (a1) in panel (b). In panel (b), we show the the MAP trajectory estimates of HMJP (magenta) that is superposed with the measurements and the true trajectory (cyan). For visual purposes only, we offset the HMJP MAP trajectory estimate by slightly shifting it downward. We observe that the HMJP MAP trajectory is able to capture switching occurring roughly in the middle of the integration time. This is not something that the HMM can capture. Here, simulated measurements are generated with* $\lambda_{\sigma_1 \to \sigma_2}, \lambda_{\sigma_2 \to \sigma_1}$ Eq. (31) *where the data acquisition happens at every* $\Delta t = 0.1$ *s with* $\tau_f = 0.8$ *s and* $\tau = 0.09$ *s starting at* $t_0 = 0.05$ *s until* $t_N = 20$ *s.*

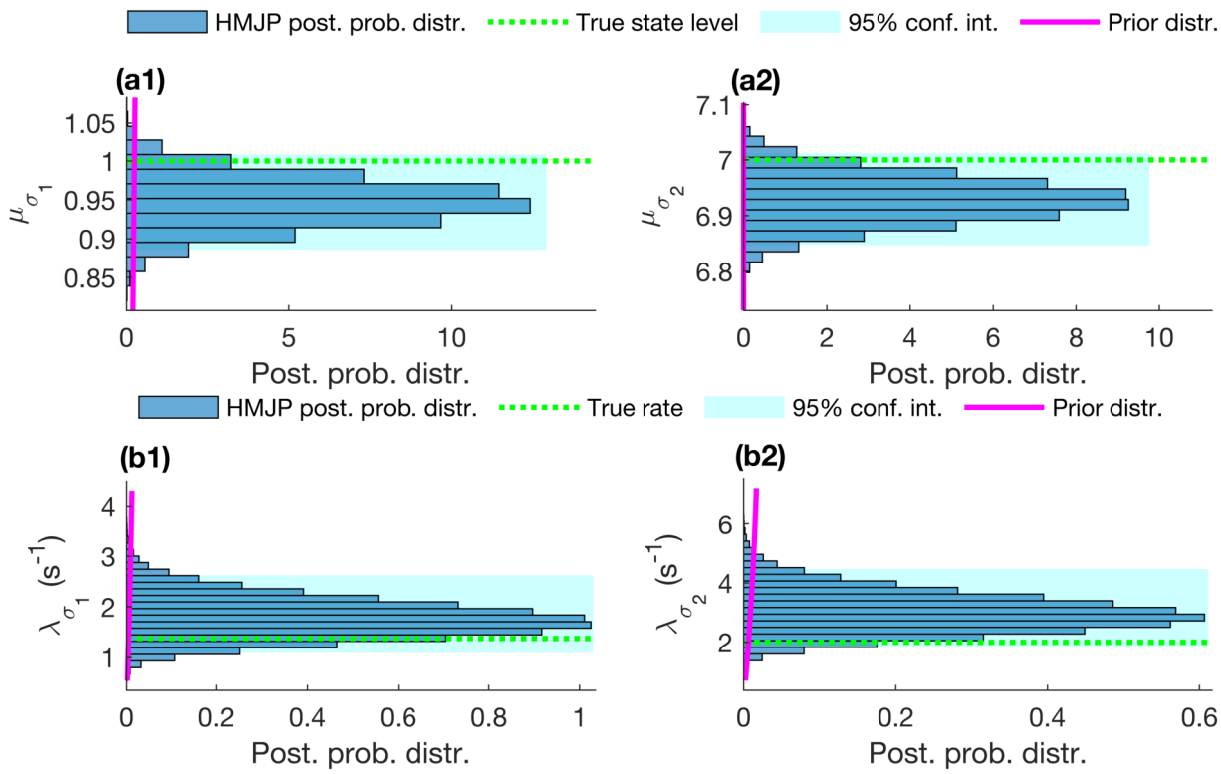

Fig. 3. ***HMJP state level and rate estimates for slow state switching***. *Here we provide posterior state level and rate estimates obtained with HMJP whose time trace we discussed in Fig. 2. We expect HMJPs to perform well in estimating the true state levels and rates when these rates are slower than the data acquisition rate. In all figure panels, we superposed the posterior distributions over state levels and rates for HMJP (blue) along with their* 95% *confidence intervals, the true state levels (dashed green lines) and the corresponding prior distributions (magenta lines). We start with the information in panels (a1) and (a2). We observe in these panels that the HMJP posterior distributions over state levels contain the true state levels within their* 95% *confidence intervals. Next we move to the panels (b1) and (b2) which show the posterior distributions over the rates labeled* $\lambda_{\sigma_k \to \sigma_{k'}}$ *for all* $k, k' = 1, 2$*. Again, the HMJP does quite well in estimating these rates as measured by the fact that the ground truth lies within the* 95% *confidence intervals of the posteriors. In this figure, the analyzed simulated measurements are generated with the same parameters as those provided in Fig. 2.*

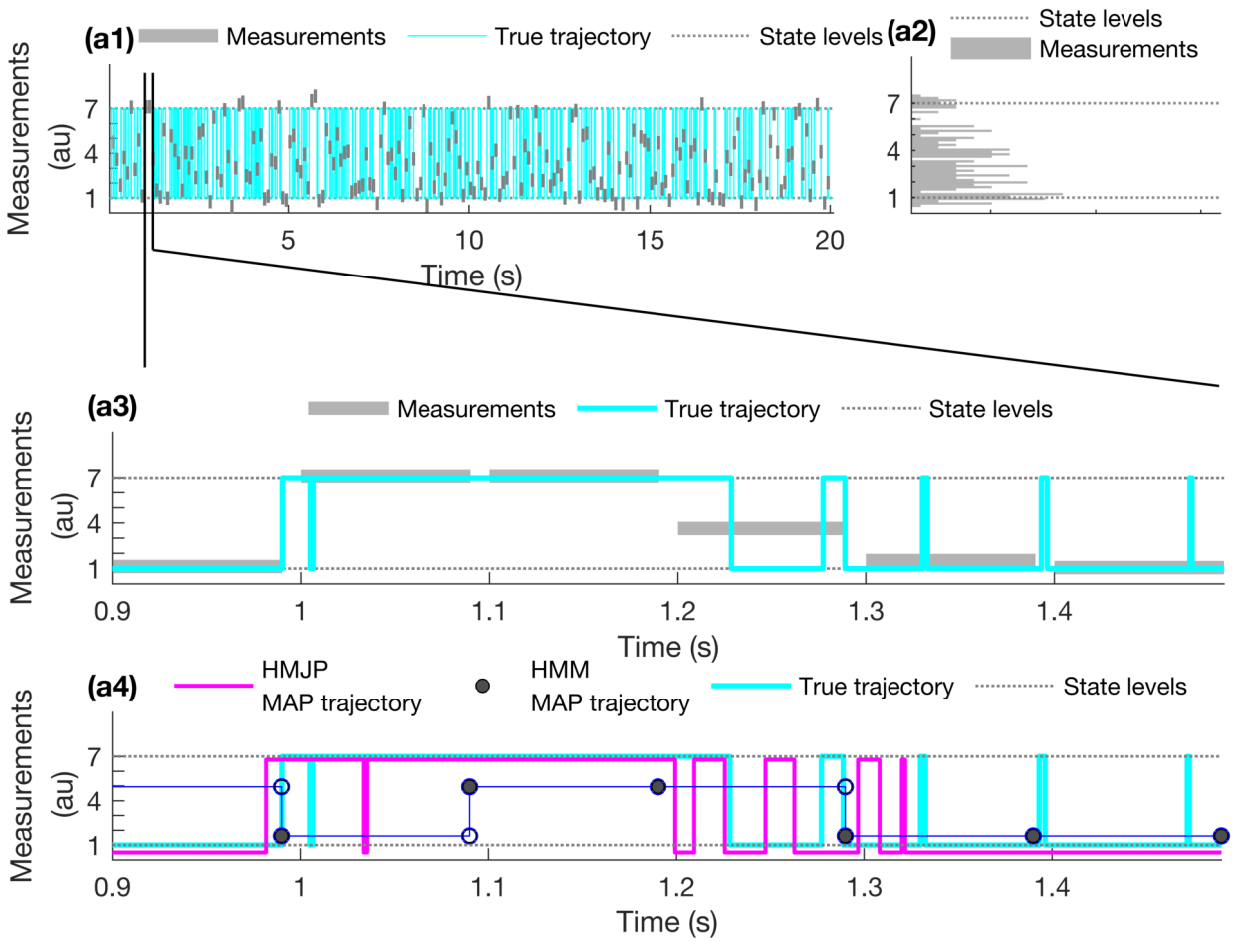

Fig. 4. ***HMJP with HMM trajectory estimates for fast state switching***. *Here we provide trajectory estimates obtained with HMJP and HMM when the switching rate is faster than the data acquisition rate,* $1/\Delta t = 10$ *(1/s). We expect HMMs to perform poorly in estimating the true trajectory when switching is fast. In this figure's panel (a1), the measurements are shown with gray rectangles (the width of the rectangle coincides with the integration period as shown in Fig. 1) that are generated based on the description provided in Section 2.1. We follow the same color scheme and layout as in Fig. 2 except for panel (a4) where we provide the MAP trajectory estimate provided by the HMJP as well as the HMM. In panel (a4), the magenta dashed line shows the HMJP MAP trajectory estimate and the blue line shows the HMM MAP trajectory estimate. For visual purposes, we offset the HMJP MAP trajectory estimate and HMM MAP trajectory estimate by shifting these downward. Here, simulated measurements are generated with* $\lambda_{\sigma_1\to\sigma_2}, \lambda_{\sigma_2\to\sigma_1}$ *(see* Eq. (31)*) where the data acquisition happens at every* $\Delta t = 0.1$ *s with* $\tau_f = 1/15$ *s and* $\tau = 0.09$ *s starting at* $t_0 = 0.05$ *s until* $t_N = 20$ *s.*

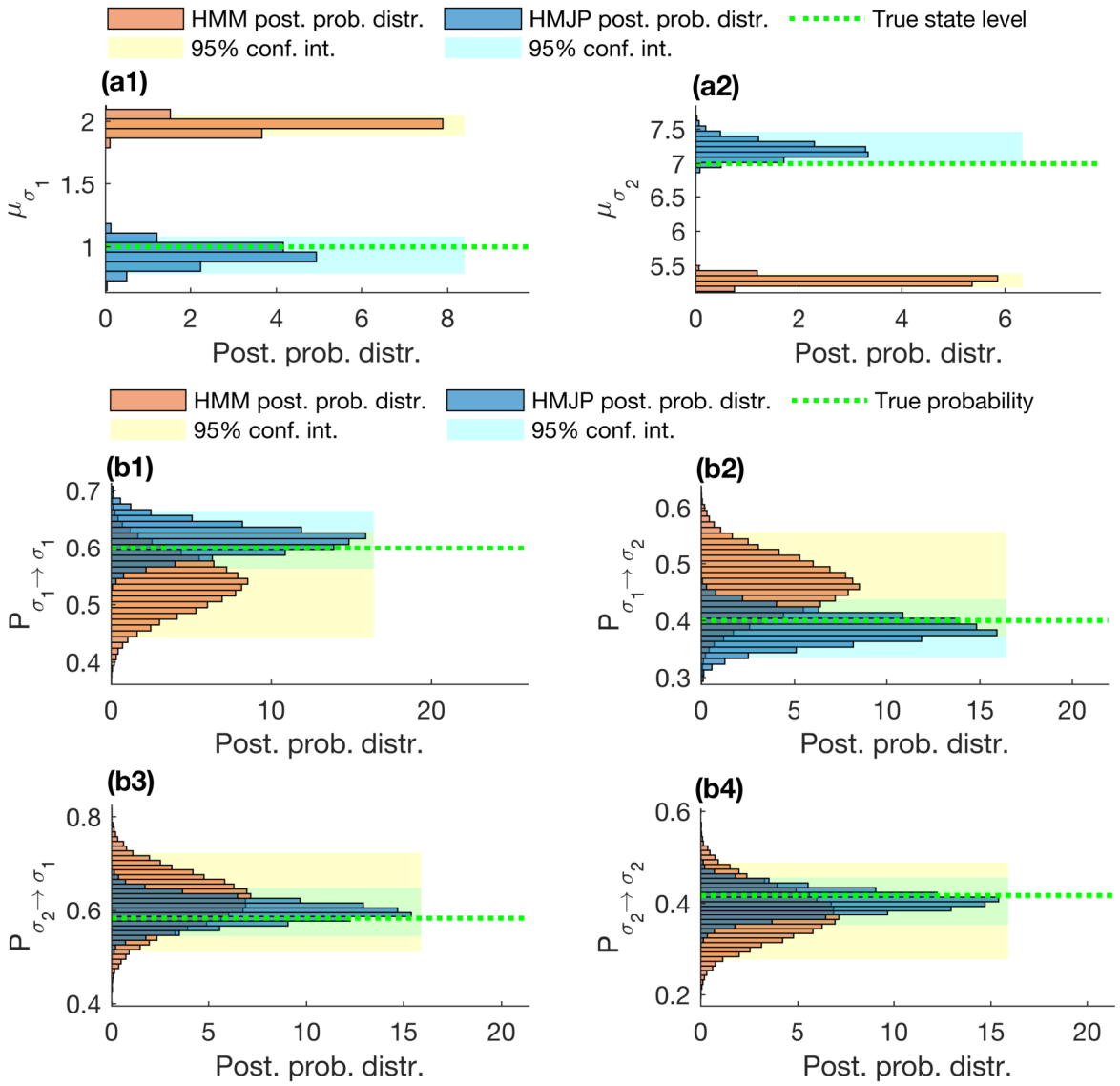

Fig. 5. ***HMJP with HMM state level and transition probability estimates for fast state switching***. *Here we provide posterior state level and transition probability estimates obtained with HMJP and HMM when the switching rate is faster than the data acquisition rate, $1/\Delta t = 10$ (1/s). We expect HMMs to perform poorly in estimating the true state levels and transition probabilities when the system switching is fast. In all of this figure's panels, we superposed the posterior distributions over state levels for both HMJP (blue) and HMM (orange) along with their 95% confidence intervals, the true state levels and true transition probabilities (dashed green lines). Next we move to transition probability estimates provided in panels (b1)-(b4). In these panels, we wish to test the performance of HMJPs and HMMs in estimating the transition probabilities. In this figure, each panel is corresponding to the posterior distribution of a transition probability labeled as $P_{\sigma_k \to \sigma_{k'}}$ for all $k, k' = 1, 2$. Here, simulated measurements are generated with the same parameters as those provided in Fig. 4.*

Predictably, the HMM performs poorly. For example, we see in Fig. 4 panel (a4) that the HMJP MAP trajectory captures many of the fast switches occurring during integration times. The HMM MAP trajectory is severely constrained to allowing switches at the end of the time period and, as such, cannot accommodate fast kinetics. While the trajectory inferred by HMJPs is not perfect, this ability to tease out many correct state switches in its MAP trajectory is sufficient for HMJPs to obtain estimates of the transition probabilities and state levels that lie within the 95% confidence interval; see panels (a1)-(a2) and (b1)-(b4) of Fig. 5. The same does not hold for HMMs where their inability to detect state switches now percolates down to the quality of their estimates for the state levels and transition probabilities. To wit, from panels (a1)-(a2) of Fig. 5 we see that the HMM grossly overestimates (by about 90%) μ_{σ_1} and underestimates (about 30%) μ_{σ_2}. What is more, as can be seen in panels (b1)-(b4) of Fig. 5, the HMM provides very wide posterior distributions over transition probabilities. This is by contrast to the much sharper posterior of the HMJP whose mode closely coincides with the ground truth; see panels (b1)-(b4) of Fig. 5.

An observation is warranted here. As the HMM cannot accommodate fast kinetics, it must ascribe the apparent spread around the $P_{\sigma_k \to \sigma_{k'}}$ histogram (see panels (b1)-(b4) of Fig. 5) to an increased variance in the posterior distribution of transition probabilities. So, while the breadth of the posteriors of the HMJP are primarily ascribed to the fact that finite data informs the posterior, the origin of the breadth of the histogram of the HMM is an artifact of its inability to accommodate fast kinetics.

In the appendix, we delve into finer details of the effect of finiteness of data on the HMM posterior distributions over transition probabilities for fast switching rates in Supporting Material (A), see Fig. A.6. In particular, we analyzed a sequence of 3 data sets using the HMM framework with the same fast switching rates as in Fig. 4 panel (a1) but with differing data set lengths. While more data eventually narrows the HMM's posterior over transition probabilities, the amount of data has almost no effect on the poor posterior distributions over state levels Fig. A.7. Further analysis on the performance of both HMJP and HMM is left for the Supporting Material (A) in particular cases where the rate from one state to another is fast and the other is slow. We also provide the HMJP rate estimates in Fig. A.5 panels (a1)-(a2) for the data set given in Fig. 1 panel (b1) as well as a comparison of HMJP posterior transition probability estimates associated with the data provided in Fig. 1 panel (a1) with and without learning the trajectory $\mathcal{T}(\cdot)$ simultaneously in Fig. A.8. Finally, we compare the posterior trajectory estimates of HMJP and HMM based on a metric that is the enclosed area under the learned trajectories in Fig. A.9.

4 Discussion

HMMs have been a hallmark of time series analysis in single molecule Biophysics (11, 12, 38–40, 44–54) but they have a critical limitation. HMMs apply only provided the temporal resolution of the experimental apparatus is faster than the system kinetics under study (31, 91–93). Otherwise, HMMs mistakenly ascribe the signal generated by fast dynamics to misassignments of signal levels, such as Förster resonance energy transfer distances in single molecule FRET (7, 12, 38, 39, 46), amongst other artefacts; see Figs. 5, A.2 and A.4. Fundamentally this limitation arises because HMMs treat the dynamics of the probed systems discretely while, in fact, they evolve continuously.

The HMJP we describe here is the continuous time analog of the HMM. It can be easily adapted to treat various single molecule experiment setups. To wit, in a companion article, we will use HMJPs for the analysis of single molecule FRET (20–27, 44).

The HMJP does have limitations. In the limit that state switching rate grows, the amount of data needed to ascertain a meaningful posterior over the transition kinetics also grows. In the trivial limit that the state switching is extremely fast, no method, whether HMJP or otherwise, would be able to tease out information on the transition kinetics from what appears as a uniform horizontal time trace with noise with no discernible transitions. Although, the quality of the data is not a limitation of HMJP, it is clear that the duration of the detector dead time affects

the performance of all methods of inference. Specifically, the longer the dead time, the worse the HMJP will perform. In the limit the dead time is as long as the exposure itself, the HMJP reduces to the HMM.

Of great interest is the possibility to learn the number of states within an HMJP framework. That is, to re-pitch the HMJP within a Bayesian nonparametric paradigm following in the footsteps of the HMM and its nonparametric realization, the infinite HMM (37–40, 94–96). Methods have been developed to report on point statistics as they pertain to infinite MJPs (65, 97). A natural extension for us would be to fully sample from a posterior with realistic measurement models relevant to single molecule Biophysics within the nonparametric paradigm.

ACKNOWLEDGEMENTS

SP acknowledges support from NSF CAREER grant MCB-1719537 and NIH NIGMS (R01GM134426). ASU cluster AGAVE and Saguaro are the main computational resources utilized in this study. ZK thanks Sina Jazani for his helpful suggestions on the manuscript.

1. L. E. Baum, T. Petrie, *et al.*, "The annals of mathematical statistics," *Statistical inference for probabilistic functions of finite state markov chains*, vol. 37, pp. 1554–1563, 1966.
2. L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains," *The annals of mathematical statistics*, vol. 41, pp. 164–171, 1970.
3. T. Petrie, "Probabilistic functions of finite state markov chains," *The Annals of Mathematical Statistics*, vol. 40, pp. 97–115, 1969.
4. L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, pp. 257–286, 1989.
5. S. E. Levinson, L. R. Rabiner, and M. M. Sondhi, "An introduction to the application of the theory of probabilistic functions of a markov process to automatic speech recognition," *Bell System Technical Journal*, vol. 62, pp. 1035–1074, 1983.
6. D. Kelly, M. Dillingham, A. Hudson, and K. Wiesner, "A new method for inferring hidden markov models from noisy time sequences," *PloS one*, vol. 7, 2012.
7. Y. Liu, J. Park, K. A. Dahmen, Y. R. Chemla, and T. Ha, "A comparative study of multivariate and univariate hidden markov modelings in time-binned single-molecule fret data analysis," *The Journal of Physical Chemistry B*, vol. 114, pp. 5386–5403, 2010.
8. N. Zarrabi, S. Ernst, B. Verhalen, S. Wilkens, and M. Börsch, "Analyzing conformational dynamics of single p-glycoprotein transporters by förster resonance energy transfer using hidden markov models," *Methods*, vol. 66, pp. 168–179, 2014.
9. D. A. Smith, W. Steffen, R. M. Simmons, and J. Sleep, "Hidden-markov methods for the analysis of single-molecule actomyosin displacement data: the variance-hidden-markov method," *Biophysical journal*, vol. 81, pp. 2795–2816, 2001.
10. N. Zarrabi, M. Düser, R. Reuter, S. Dunn, J. Wrachtrup, and M. Börsch, "Detecting substeps in the rotary motors of fof1-atp synthase by hidden markov models," in *Ultrasensitive and Single-Molecule Detection Technologies II*, vol. 6444, p. 64440E, 2007.
11. T.-H. Lee, "Extracting kinetics information from single-molecule fluorescence resonance energy transfer data using hidden markov models," *The Journal of Physical Chemistry B*, vol. 113, pp. 11535–11542, 2009.
12. S. A. McKinney, C. Joo, and T. Ha, "Analysis of single-molecule fret trajectories using hidden markov modeling," *Biophysical journal*, vol. 91, pp. 1941–1951, 2006.
13. B. G. Keller, A. Kobitski, A. Jäschke, G. U. Nienhaus, and F. Noé, "Complex rna folding kinetics revealed by single-molecule fret and hidden markov models," *Journal of the American Chemical Society*, vol. 136, pp. 4534–4543, 2014.
14. M. Andrec, R. M. Levy, and D. S. Talaga, "Direct determination of kinetic rates from single-molecule photon arrival trajectories using hidden markov models," *The Journal of Physical Chemistry A*, vol. 107, pp. 7454–7464, 2003.
15. G. Baumann and G. S. Easton, "Modeling state-dependent sodium conductance data by a memoryless random process," *Mathematical Biosciences*, vol. 60, pp. 265–276, 1982.
16. R. Edeson, G. Yeo, R. Milne, and B. Madsen, "Graphs, random sums, and sojourn time distributions, with application to ion-channel modeling," *Mathematical biosciences*, vol. 102, pp. 75–104, 1990.
17. G. F. Yeo, R. Milne, R. Edeson, and B. Madsen, "Statistical inference from single channel records: two-state markov model with limited time resolution," *Proceedings of the Royal society of London. Series B. Biological sciences*, vol. 235, pp. 63–94, 1988.

18. R. Milne, G. Yeo, B. Madsen, and R. Edeson, "Estimation of single channel kinetic parameters from data subject to limited time resolution," *Biophysical journal*, vol. 55, pp. 673–676, 1989.
19. R. Milne, G. Yeo, R. Edeson, and B. Madsen, "Stochastic modelling of a single ion channel: an alternating renewal approach with application to limited time resolution," *Proceedings of the Royal society of London. Series B. Biological sciences*, vol. 233, pp. 247–292, 1988.
20. M. R. Wasserman, J. L. Alejo, R. B. Altman, and S. C. Blanchard, "Multiperspective smfret reveals rate-determining late intermediates of ribosomal translocation," *Nature structural & molecular biology*, vol. 23, p. 333, 2016.
21. M. Lu, X. Ma, L. R. Castillo-Menendez, J. Gorman, N. Alsahafi, U. Ermel, D. S. Terry, M. Chambers, D. Peng, B. Zhang, *et al.*, "Associating hiv-1 envelope glycoprotein structures with states on the virus observed by smfret," *Nature*, vol. 568, pp. 415–419, 2019.
22. A. Mazouchi, Z. Zhang, A. Bahram, G.-N. Gomes, H. Lin, J. Song, H. S. Chan, J. D. Forman-Kay, and C. C. Gradinaru, "Conformations of a metastable sh3 domain characterized by smfret and an excluded-volume polymer model," *Biophysical journal*, vol. 110, pp. 1510–1522, 2016.
23. R. Roy, S. Hohng, and T. Ha, "A practical guide to single-molecule fret," *Nature methods*, vol. 5, pp. 507–516, 2008.
24. Y. Zhu, L. He, Y. Liu, Y. Zhao, and X. C. Zhang, "smfret probing reveals substrate-dependent conformational dynamics of e. coli multidrug mdfa," *Biophysical journal*, vol. 116, pp. 2296–2303, 2019.
25. E. Lerner, T. Cordes, A. Ingargiola, Y. Alhadid, S. Chung, X. Michalet, and S. Weiss, "Toward dynamic structural biology: Two decades of single-molecule förster resonance energy transfer," *Science*, vol. 359, p. eaan1133, 2018.
26. A. Krishnamoorti, R. C. Cheng, V. Berka, and M. Maduke, "Clc conformational landscape as studied by smfret," *Biophysical Journal*, vol. 116, p. 555a, 2019.
27. S. Hohng, R. Zhou, M. K. Nahas, J. Yu, K. Schulten, D. M. Lilley, and T. Ha, "Fluorescence-force spectroscopy maps two-dimensional reaction landscape of the holliday junction," *Science*, vol. 318, pp. 279–283, 2007.
28. B. Schuler, "Single-molecule fret of protein structure and dynamics-a primer," *Journal of nanobiotechnology*, vol. 11, p. S2, 2013.
29. M. Kruithof and J. van Noort, "Hidden markov analysis of nucleosome unwrapping under force," *Biophysical journal*, vol. 96, pp. 3708–3715, 2009.
30. P. J. Elms, J. D. Chodera, C. J. Bustamante, and S. Marqusee, "Limitations of constant-force-feedback experiments," *Biophysical journal*, vol. 103, pp. 1490–1499, 2012.
31. Y. Zhang, J. Jiao, and A. A. Rebane, "Hidden markov modeling with detailed balance and its application to single protein folding," *Biophysical journal*, vol. 111, pp. 2110–2124, 2016.
32. F. G. Ball and J. A. Rice, "Stochastic models for ion channels: Introduction and bibliography," *Mathematical Biosciences*, vol. 112, pp. 189 – 206, 1992.
33. A. Lee, K. Tsekouras, C. Calderon, C. Bustamante, and S. Pressé, "Unraveling the thousand word picture: an introduction to super-resolution data analysis," *Chemical reviews*, vol. 117, pp. 7276–7330, 2017.
34. M. Tavakoli, J. N. Taylor, C.-B. Li, T. Komatsuzaki, and S. Pressé, "Single molecule data analysis: An introduction," *arXiv preprint arXiv:1606.00403*, 2016.
35. N. G. Van Kampen, *Stochastic processes in physics and chemistry*, vol. 1. 1992.
36. M. Levitus and S. Ranjit, "Cyanine dyes in biophysical research: the photophysics of polymethine fluorescent dyes in biomolecular environments," *Quarterly reviews of biophysics*, vol. 44, pp. 123–151, 2011.
37. J. Van Gael, Y. Saatci, Y. W. Teh, and Z. Ghahramani, "Beam sampling for the infinite hidden markov model," in *Proceedings of the 25th international conference on Machine learning*, pp. 1088–1095, 2008.
38. I. Sgouralis and S. Pressé, "An introduction to infinite hmms for single-molecule data analysis," *Biophysical journal*, vol. 112, pp. 2021–2029, 2017.
39. I. Sgouralis and S. Pressé, "Icon: an adaptation of infinite hmms for time traces with drift," *Biophysical journal*, vol. 112, pp. 2117–2126, 2017.
40. I. Sgouralis, S. Madaan, F. Djutanta, R. Kha, R. F. Hariadi, and S. Pressé, "A bayesian nonparametric approach to single molecule förster resonance energy transfer," *The Journal of Physical Chemistry B*, vol. 123, pp. 675–688, 2018.
41. S. Jazani, I. Sgouralis, O. M. Shafraz, M. Levitus, S. Sivasankar, and S. Pressé, "An alternative framework for fluorescence correlation spectroscopy," *Nature communications*, vol. 10, pp. 1–10, 2019.
42. S. Jazani, I. Sgouralis, and S. Pressé, "A method for single molecule tracking using a conventional single-focus confocal setup," *The Journal of Chemical Physics*, vol. 150, p. 114108, 2019.
43. I. Sgouralis, M. Whitmore, L. Lapidus, M. J. Comstock, and S. Pressé, "Single molecule force spectroscopy at high data acquisition: A bayesian nonparametric analysis," *The Journal of chemical physics*, vol. 148, p. 123320, 2018.
44. T. Burzykowski, J. P. Szubiakowski, and T. Ryden, "Statistical analysis of data from single molecule experiment," in *IV Workshop on Atomic and Molecular Physics*, vol. 5258, pp. 171–177, 2003.
45. M. Pirchi, R. Tsukanov, R. Khamis, T. E. Tomov, Y. Berger, D. C. Khara, H. Volkov, G. Haran, and E. Nir, "Photon-by-photon

hidden markov model analysis for microsecond single-molecule fret kinetics," *The Journal of Physical Chemistry B*, vol. 120, pp. 13065–13075, 2016.
46. J. Lee, S. Lee, K. Ragunathan, C. Joo, T. Ha, and S. Hohng, "Single-molecule four-color fret," *Angewandte Chemie International Edition*, vol. 49, pp. 9922–9925, 2010.
47. H. Mazal and G. Haran, "Single-molecule fret methods to study the dynamics of proteins at work," *Current Opinion in Biomedical Engineering*, vol. 12, pp. 8–17, 2019.
48. K. Shen, S. Arslan, D. Akopian, T. Ha, and S.-o. Shan, "Activated gtpase movement on an rna scaffold drives co-translational protein targeting," *Nature*, vol. 492, pp. 271–275, 2012.
49. C. Joo, S. A. McKinney, M. Nakamura, I. Rasnik, S. Myong, and T. Ha, "Real-time observation of reca filament dynamics with single monomer resolution," *Cell*, vol. 126, pp. 515–527, 2006.
50. C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, and T. Ha, "Advances in single-molecule fluorescence methods for molecular biology," *Annu. Rev. Biochem.*, vol. 77, pp. 51–76, 2008.
51. P. V. Cornish, D. N. Ermolenko, D. W. Staple, L. Hoang, R. P. Hickerson, H. F. Noller, and T. Ha, "Following movement of the l1 stalk between three functional states in single ribosomes," *Proceedings of the National Academy of Sciences*, vol. 106, pp. 2571–2576, 2009.
52. H. Li and H. Yang, "Statistical learning of discrete states in time series," *The Journal of Physical Chemistry B*, vol. 123, pp. 689–701, 2018.
53. I. V. Gopich, "Likelihood functions for the analysis of single-molecule binned photon sequences," *Chemical physics*, vol. 396, pp. 53–60, 2012.
54. H. S. Chung and I. V. Gopich, "Fast single-molecule fret spectroscopy: theory and experiment," *Physical Chemistry Chemical Physics*, vol. 16, pp. 18644–18657, 2014.
55. C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.
56. L. Rabiner and B. Juang, "An introduction to hidden markov models," *ieee assp magazine*, vol. 3, pp. 4–16, 1986.
57. L. R. Rabiner, *Multirate digital signal processing*. 1996.
58. J. E. Bronson, J. Fei, J. M. Hofman, R. L. Gonzalez Jr, and C. H. Wiggins, "Learning rates and states from biophysical time series: a bayesian approach to model selection and single-molecule fret data," *Biophysical journal*, vol. 97, pp. 3196–3205, 2009.
59. B. Shuang, D. Cooper, J. N. Taylor, L. Kisley, J. Chen, W. Wang, C. B. Li, T. Komatsuzaki, and C. F. Landes, "Fast step transition and state identification (stasi) for discrete single-molecule data analysis," *The journal of physical chemistry letters*, vol. 5, pp. 3157–3161, 2014.
60. J. Chen, N. K. Poddar, L. J. Tauzin, D. Cooper, A. B. Kolomeisky, and C. F. Landes, "Single-molecule fret studies of hiv tar–dna hairpin unfolding dynamics," *The Journal of Physical Chemistry B*, vol. 118, pp. 12130–12139, 2014.
61. W. Koopmans, A. Brehm, C. Logie, T. Schmidt, and J. Van Noort, "Single-pair fret microscopy reveals mononucleosome dynamics," *Journal of fluorescence*, vol. 17, pp. 785–795, 2007.
62. S. Preus, S. L. Noer, L. L. Hildebrandt, D. Gudnason, and V. Birkedal, "isms: single-molecule fret microscopy software," *nature methods*, vol. 12, pp. 593–594, 2015.
63. P. Metzner, E. Dittmer, T. Jahnke, and C. Schütte, "Generator estimation of markov jump processes," *Journal of Computational Physics*, vol. 227, pp. 353–375, 2007.
64. A. Hobolth and E. A. Stone, "Simulation from endpoint-conditioned, continuous-time markov chains on a finite state space, with applications to molecular evolution," *The annals of applied statistics*, vol. 3, p. 1204, 2009.
65. J. H. Huggins, K. Narasimhan, A. Saeedi, and V. K. Mansinghka, "Jump-means: Small-variance asymptotics for markov jump processes," *arXiv preprint arXiv:1503.00332*, 2015.
66. B. Zhang and V. Rao, "Efficient parameter sampling for markov jump processes," *Journal of Computational and Graphical Statistics*, vol. 0, pp. 1–18, 2020.
67. C. H. L. Beentjes and R. E. Baker, "Uniformization techniques for stochastic simulation of chemical reaction networks," *The Journal of Chemical Physics*, vol. 150, p. 154107, 2019.
68. V. Rao and Y. W. Teh, "Fast mcmc sampling for markov jump processes and extensions," *The Journal of Machine Learning Research*, vol. 14, pp. 3295–3320, 2013.
69. N. M. Van Dijk, "Uniformization for nonhomogeneous markov chains," *Operations research letters*, vol. 12, pp. 283–291, 1992.
70. J. D. Diener, W. H. Sanders, H. William, and Z. Ers, "Empirical comparison of uniformization methods for continuous-time markov chains," 1994.
71. A. P. Van Moorsel and K. Wolter, "Numerical solution of non-homogeneous markov processes through uniformization.," in *ESM*, pp. 710–717, 1998.
72. D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of computational physics*, vol. 22, pp. 403–434, 1976.
73. D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The journal of physical chemistry*, vol. 81,

pp. 2340–2361, 1977.
74. D. Sivia and J. Skilling, *Data analysis: a Bayesian tutorial*. 2006.
75. B. Buelens, P. Daas, J. Burger, M. Puts, and J. van den Brakel, *Selectivity of Big data*. 2014.
76. A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*. 2002.
77. C. Robert and G. Casella, *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.
78. E. Nir, X. Michalet, K. M. Hamadani, T. A. Laurence, D. Neuhauser, Y. Kovchegov, and S. Weiss, "Shot-noise limited single-molecule fret histograms: comparison between theory and experiments," *The Journal of Physical Chemistry B*, vol. 110, pp. 22103–22124, 2006.
79. S. Kalinin, E. Sisamakis, S. W. Magennis, S. Felekyan, and C. A. Seidel, "On the origin of broadening of single-molecule fret efficiency distributions beyond shot noise limits," *The journal of physical chemistry B*, vol. 114, pp. 6197–6206, 2010.
80. K. I. Mortensen, L. S. Churchman, J. A. Spudich, and H. Flyvbjerg, "Optimized localization analysis for single-molecule tracking and super-resolution microscopy," *Nature methods*, vol. 7, p. 377, 2010.
81. F. Charrière, T. Colomb, F. Montfort, E. Cuche, P. Marquet, and C. Depeursinge, "Shot-noise influence on the reconstructed phase image signal-to-noise ratio in digital holographic microscopy," *Applied optics*, vol. 45, pp. 7667–7673, 2006.
82. M. Gross, P. Goy, and M. Al-Koussa, "Shot-noise detection of ultrasound-tagged photons in ultrasound-modulated optical imaging," *Optics letters*, vol. 28, pp. 2482–2484, 2003.
83. F. Huang, T. M. Hartwich, F. E. Rivera-Molina, Y. Lin, W. C. Duim, J. J. Long, P. D. Uchil, J. R. Myers, M. A. Baird, W. Mothes, *et al.*, "Video-rate nanoscopy using scmos camera–specific single-molecule localization algorithms," *Nature methods*, vol. 10, pp. 653–658, 2013.
84. V. Krishnaswami, C. J. Van Noorden, E. M. Manders, and R. A. Hoebe, "Towards digital photon counting cameras for single-molecule optical nanoscopy," *Optical Nanoscopy*, vol. 3, p. 1, 2014.
85. Y. Lin, J. J. Long, F. Huang, W. C. Duim, S. Kirschbaum, Y. Zhang, L. K. Schroeder, A. A. Rebane, M. G. M. Velasco, A. Virrueta, *et al.*, "Quantifying and optimizing single-molecule switching nanoscopy at high speeds," *PloS one*, vol. 10, 2015.
86. M. A. Little, *Machine Learning for Signal Processing: Data Science, Algorithms, and Computational Statistics*. 2019.
87. B. A. Hamilton, "The field guide to data science," 2015.
88. S. I. H. W. I. K. T. T. Kilic, Zeliha and S. Pressé, "Beating the single molecule nyquist limit," *Submitted*, 2020.
89. R. B. Israel, J. S. Rosenthal, and J. Z. Wei, "Finding generators for markov chains via empirical transition matrices, with applications to credit ratings," *Mathematical finance*, vol. 11, pp. 245–265, 2001.
90. E. T. Lee and J. Wang, *Statistical methods for survival data analysis*, vol. 476. 2003.
91. J.-W. van de Meent, J. E. Bronson, C. H. Wiggins, and R. L. Gonzalez Jr, "Empirical bayes methods enable advanced population-level analyses of single-molecule fret experiments," *Biophysical journal*, vol. 106, pp. 1327–1337, 2014.
92. W. Whitt, "Continuous-time markov chains," *Dept. of Industrial Engineering and Operations Research, Columbia University, New York*, 2006.
93. C. D. Kinz-Thompson, N. A. Bailey, and R. L. Gonzalez Jr, "Precisely and accurately inferring single-molecule rate constants," vol. 581, pp. 187–225, 2016.
94. M. J. Beal, Z. Ghahramani, and C. E. Rasmussen, "The infinite hidden markov model," in *Advances in neural information processing systems*, pp. 577–584, 2002.
95. M. Nakano, J. Le Roux, H. Kameoka, T. Nakamura, N. Ono, and S. Sagayama, "Bayesian nonparametric spectrogram modeling based on infinite factorial infinite hidden markov model," in *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 325–328, 2011.
96. J. M. Maheu and Q. Yang, "An infinite hidden markov model for short-term interest rates," *Journal of Empirical Finance*, vol. 38, pp. 202–220, 2016.
97. A. Georgoulas, J. Hillston, and G. Sanguinetti, "Unbiased bayesian inference for population markov jump processes via random truncations," *Statistics and computing*, vol. 27, pp. 991–1002, 2017.
98. J. Pan, V. A. Rao, P. K. Agarwal, and A. E. Gelfand, "Markov-modulated marked poisson processes for check-in data," in *ICML*, 2016.
99. A. Jensen, "Markoff chains as an aid in the study of markoff processes," *Scandinavian Actuarial Journal*, vol. 1953, pp. 87–91, 1953.
100. B. Miasojedow, W. Niemiro, *et al.*, "Geometric ergodicity of rao and teh's algorithm for markov jump processes and ctbns," *Electronic Journal of Statistics*, vol. 11, pp. 4629–4648, 2017.

A Supporting Material

In this supplement we provide: 1) additional time series analysis to test the robustness of HMJP and HMM referenced in the main text, see Supporting Material (A.1); 2) a complete summary of the theory underlying the HMJP and HMM frameworks, see Supporting Material (A.2); 3) a summary of notational conventions and parameter choices used, see Supporting Material (A.3).

A.1 Additional Analyses

Additional results are provided in the following order: we investigated a) the performance of HMJP and HMM in estimating the trajectory, state levels and transition probabilities on the analysis of the simulated measurements generated with slow switching rates and mixed switching rates (i.e., $\lambda_{\sigma_2 \to \sigma_1}$ fast and $\lambda_{\sigma_1 \to \sigma_2}$ slow); b) the posterior rate distributions of HMJP for fast and mixed switching cases; c) how the finiteness of the data affect HMM posterior transition probabilities; d) the source of bias in negatively skewed HMJP posterior transition probability estimates; e) the assessment of the mean of the are of posterior trajectory estimates of HMJP and HMM.

A.1.1 Comparison of HMJP with HMM for Slow Switching Rates

In the text in Figs. 2 and 3, we omitted the results of the slow kinetics for the HMM analysis (as this is a regime where we expect the HMM to do well). We use the same simulation parameters that were used to generate in Fig. 1 panel (a1). Predictably, in the slow kinetics regime, we expect both HMJP and HMM to do well and this is what we find.

In Fig. A.1 panel (a3), we observe that there is at most one switch during one integration period. We see in Fig. A.1 panel (a4) that both HMJP and HMM MAP trajectory estimates are capable of capturing these switches in the true trajectory.

Next, we move to the posterior probability distribution over state levels shown in Fig. A.2 panels (a1)-(a2). In panel (a1), we observe that the HMM grossly overestimates (by about 30%) μ_{σ_1} and in panel (a2), HMM underestimates (about 3%) μ_{σ_2}. By contrast, the HMJP posterior probability distribution for the state levels provide very good state level estimates for both μ_{σ_1} and μ_{σ_2}. Put differently, true $\mu_{\sigma_1}, \mu_{\sigma_2}$ values occur in the 95% confidence intervals of the HMJP posterior distributions over state levels.
Finally, we discuss the posterior distribution over transition probabilities obtained from HMJP and HMM frameworks in Fig. A.4 panels (b1)-(b4). We expect both HMM and HMJP to provide very good estimates for probabilities of switching among the states as these are associated with slow switching rates prescribed with $\lambda_{\sigma_1 \to \sigma_2}, \lambda_{\sigma_2 \to \sigma_1}$ (see Eq. (31)).
In Fig. A.2 panels (b1)-(b4), we observe that HMM and HMJP perform quite well in estimating transition probabilities based on its narrow 95% confidence intervals.

Now we move on to the results for mixed switching rates. We define the values for the parameters τ_f, τ as we proceed.

A.1.2 Comparison of HMJP with HMM for Mixed Switching Rates

In this section, we test the performance of HMJPs and HMMs in estimating trajectory Fig. A.3, state levels Fig. A.4 and transition probabilities Fig. A.4 on the analysis of the simulated measurements associated with the switching

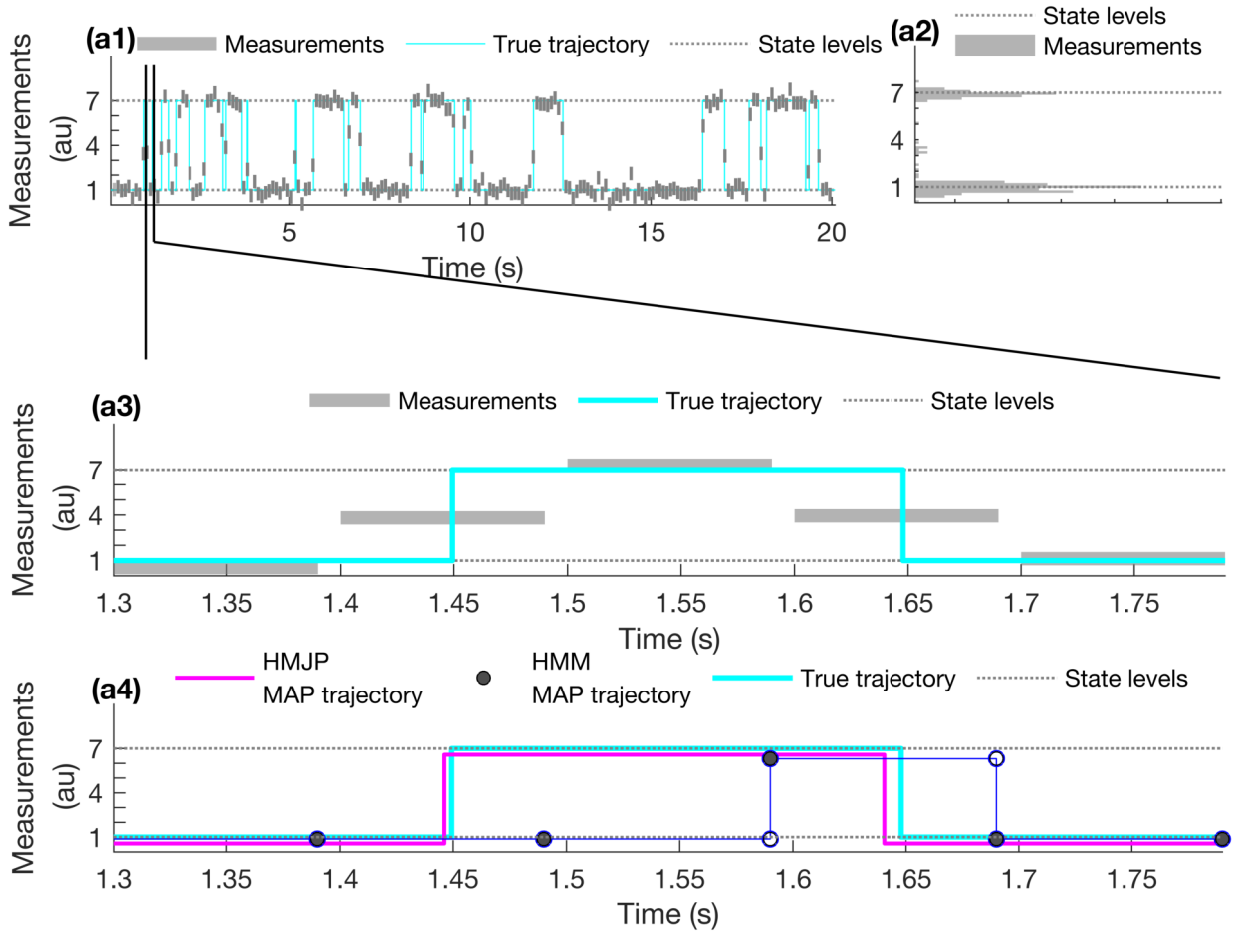

Fig. A.1. ***HMJP with HMM trajectory estimates for slow switching***. *Here we provide trajectory estimates obtained with HMJPs and HMMs. The results for the HMJP were previously provided in Figs. 2 and 3 but here we add the results for the HMM in panel (a4). We follow identical coloring schemes and convention as with Fig. 2 of the main. The parameters used are identical to those of Fig. 2.*

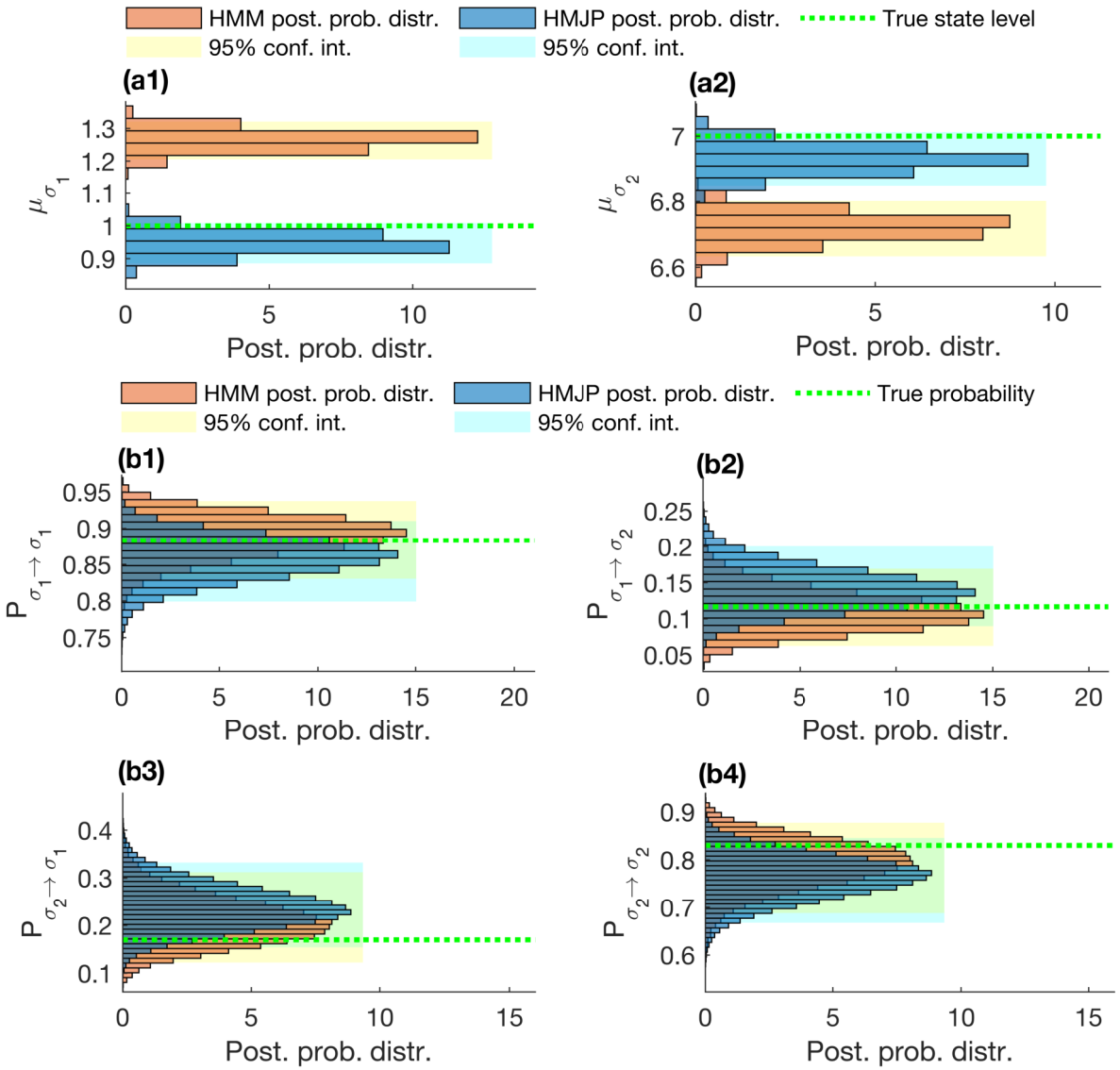

Fig. A.2. ***HMJP with HMM state level and transition probability estimates for slow switching***. *Here we provide posterior state level and transition probability estimates obtained with HMJPs and HMMs. The results for the HMJP on posterior state levels were previously provided in Fig. 3 but here we add the results for the posterior transition probabilities for both HMJP and HMM in panels (b1)-(b4) and posterior state level estimaets for HMM in panels (a1)-(a2). We follow identical coloring schemes and convention as with Fig. 3 of the main. The parameters used are identical to those of Fig. 2.*

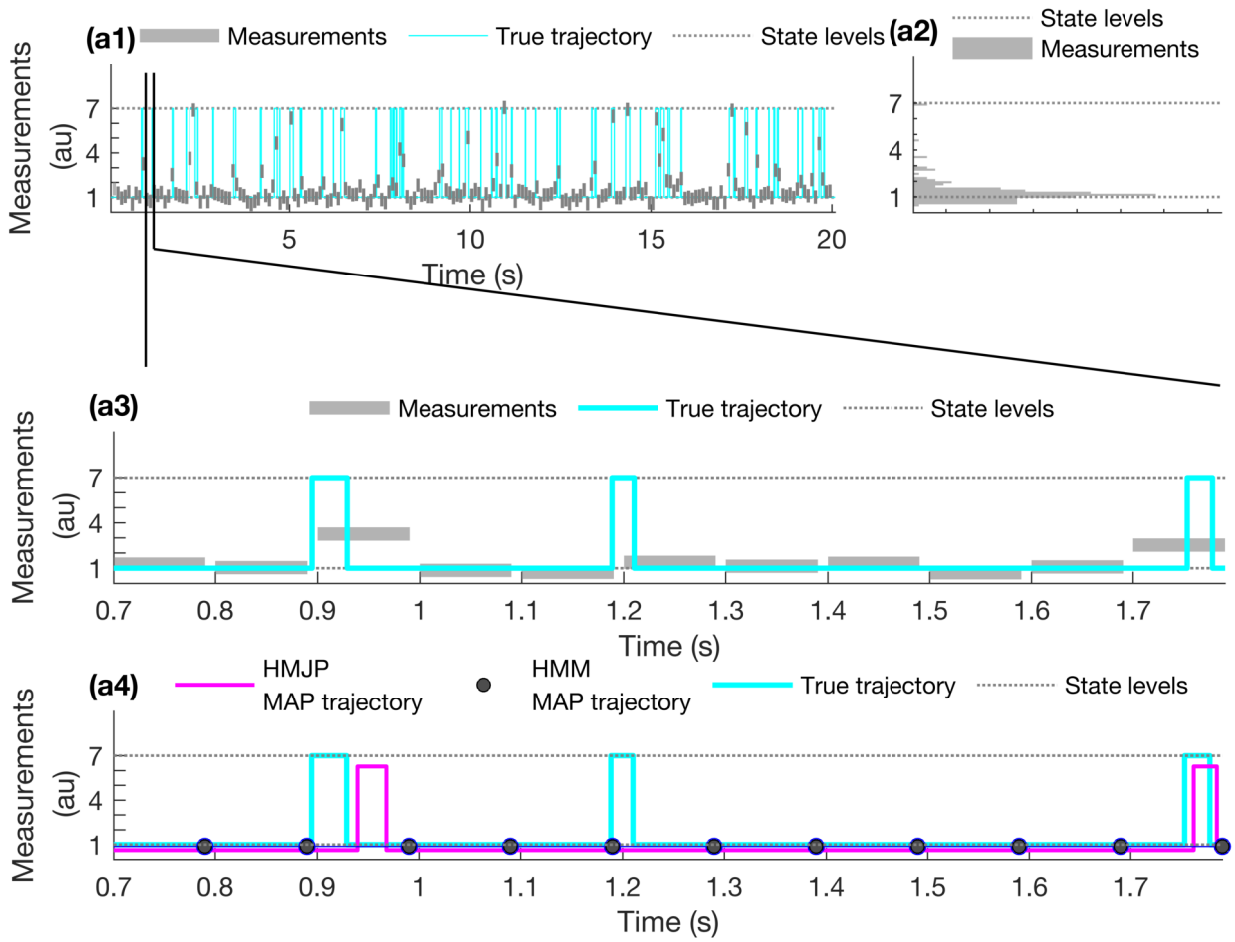

Fig. A.3. ***HMJP with HMM trajectory estimates for mixed switching***. *Here we provide trajectory estimates obtained with the HMJP and HMM when switching rates are slower or faster than the data acquisition rate,* $1/\Delta t = 10$ *(1/s). We follow identical coloring schemes and convention as in Fig. A.1. We see that the HMJP MAP trajectory is able to capture most of the fast switches between the states* σ_1 *and* σ_2*. Simulated measurements are generated with* $\lambda_{\sigma_1 \to \sigma_2}, \lambda_{\sigma_2 \to \sigma_1}$ *(see* Eq. (A.1)*) where the data acquisition happens every* $\Delta t = 0.1$ *s with* $\tau_f = 1/15$ *s and* $\tau = 0.09$ *s starting at* $t_0 = 0.05$ *s until* $t_N = 20$ *s.*

rates $\lambda_{\sigma_1 \to \sigma_2}, \lambda_{\sigma_2 \to \sigma_1}$, with a parameter τ_f which sets the timescale of the system kinetics,

$$\lambda_{\sigma_1 \to \sigma_2} = \frac{0.2}{\tau_f}, \qquad \lambda_{\sigma_2 \to \sigma_1} = \frac{1.6}{\tau_f} \tag{A.1}$$

where $\tau_f = 1/15$ s and data acquisition rate, 10 (1/s). We start with Fig. A.3 panel (a3) where we observe that there are multiple fast switching events during one integration period. We see in Fig. A.3 panel (a4) that the HMJP MAP trajectory estimate is capable of capturing most of the fast switches occurring in the true trajectory while the HMM MAP trajectory estimate provides a poor trajectory estimate as compared to ground truth.

The HMM's poor MAP trajectory estimate can be explained with the posterior probability distribution over state levels shown in Fig. A.4 panels (a1)-(a2). In Fig. A.4, we provide the superposed posterior distributions for HMJP and HMM over state levels in panels (a1)-(a2) and transition probabilities in panels (b1)-(b4). In panel (a1), we observe that the HMM grossly overestimates (by about 30%) μ_{σ_1}. In panel (a2), the HMM grossly underestimates (by about 20%) μ_{σ_2}. By contrast, the HMJP posterior probability distribution for the state levels provide very good state level estimates for both μ_{σ_1} and μ_{σ_2}. Put differently, true $\mu_{\sigma_1}, \mu_{\sigma_2}$ values take place at the 95% confidence intervals of the HMJP posterior distributions over state levels. Finally, we discuss the posterior distribution over transition probabilities obtained from the HMJP and HMM frameworks Fig. A.4 panels (b1)-(b4). We expect to see HMMs provide very good estimates for the transition probabilities from state σ_1 to itself and σ_2 as these are associated with slow switching rates prescribed with $\lambda_{\sigma_1 \to \sigma_2}$ (see Eq. (A.1)). By contrast, we expect wide posterior distributions over transition probabilities from conformational state σ_2 to itself and σ_1 as $P_{\sigma_2 \to \sigma_1}$ and $P_{\sigma_2 \to \sigma_2}$ are associated to the fast switching rates defined by $\lambda_{\sigma_2 \to \sigma_1}$ (see Eq. (A.1)). In Fig. A.4 panels (b1)-(b4), we observe that the HMJP still performs quite well in estimating transition probabilities based on its narrow 95% confidence intervals. Next, we provide the posterior distributions over rates obtained by the HMJP for fast and mixed switching kinetics.

A.1.3 Robustness Analysis for HMJP Posterior Rate Estimates for Fast and Mixed Kinetics

In the main text, for sake of comparison between HMJPs and HMMs, we compared their posterior estimates for the transition probability. However, HMJPs are also capable of providing estimates for the transition rates, not just transition kinetics.

Using parameter values to simulate the results of Fig. 4 for the fast kinetics and Fig. A.3 for the mixed rates we show rate posteriors below in Fig. A.5.

In Fig. A.5, we observe that the HMJP is capable of providing very good estimates for the analyzed data generated with slow, fast and mixed switching rates. For all simulations, we always have the same shape (η) and scale ($\frac{b}{\eta}$) hyperparameters for the gamma prior distributions and we set $\eta = 2$ and $b = 300$. Now we move on to the rest of the additional results. We define the values for the parameters τ_f, τ as we proceed.

A.1.4 Robustness Analysis with Respect to Data Set Length for HMM Posterior State Level and Transition Probability Estimates

Here we show how our posterior estimates over quantities including MAP trajectory estimate, state level and transition probabilities vary as a function of variable data set length, see Figs. A.6 and A.7. .

We focused on the case of fast kinetics because we want to demonstrate that the poor estimates are not due to finiteness of data. The parameters used are identical to those used to generate the data for Fig. 1 panel (b1) except for the final time that is $t_N = 60$ s.

We provide a quantitative metric (area under the curve (98); see Eq. (A.2)) for deviation from ground truth

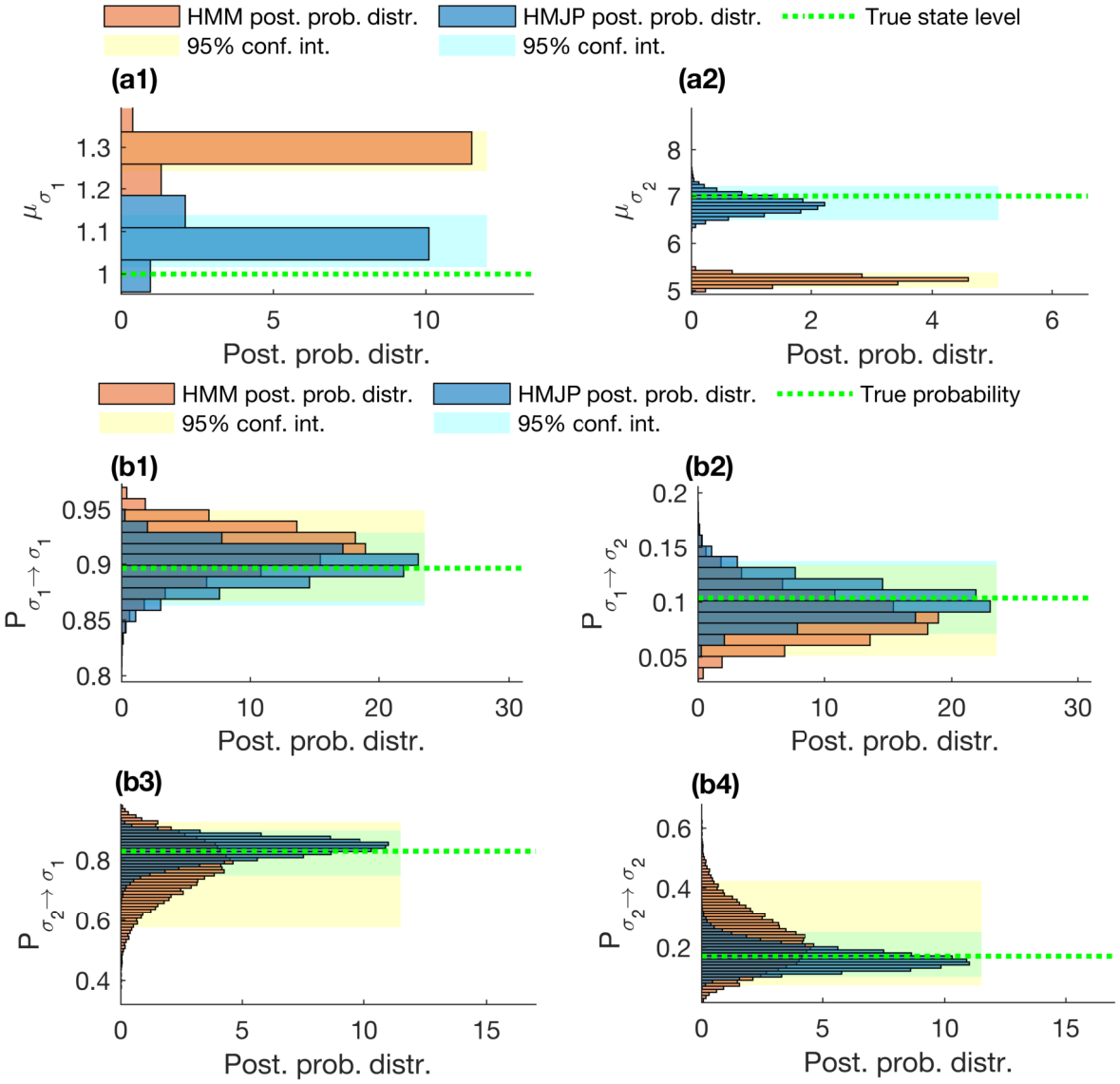

Fig. A.4. ***HMJP with HMM state level and transition probability estimates for mixed switching***. *Here we provide posterior state level and transition probability estimates obtained with the HMJP and HMM when the switching rate is slower or faster than the data acquisition rate, The results for the HMJP and HMM MAP trajectory estimates are provided in Fig. A.3. We show the superposed posterior state level estimates for the HMJP and HMM in panels (a1)-(a2) and the superposed results for the posterior transition probabilities for both HMJP and HMM in panels (b1)-(b4). We follow identical coloring schemes and convention as with Fig. 3 of the main text. Here, simulated measurements are generated with the same parameters as those provided in Fig. A.3.*

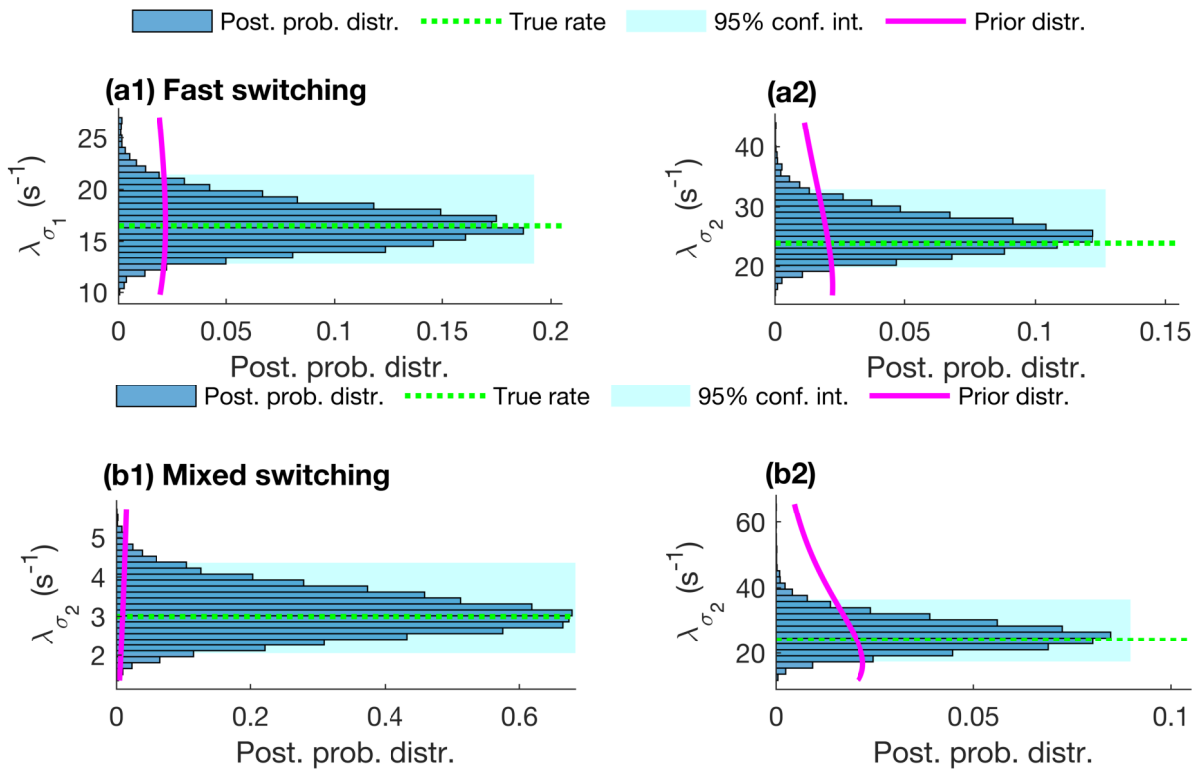

Fig. A.5. ***HMJP rate estimates for fast and mixed switching rates***. *Here we provide posterior rate estimates with the HMJP for the simulated measurements provided in Figs. 4 and A.3. We follow identical coloring schemes and convention as with Fig. 3 of the main. Here, simulated measurements associated with estimates given in panels (a1)-(a2) and (b1)-(b2) are generated with the same parameters used in Figs. 2 and 4.*

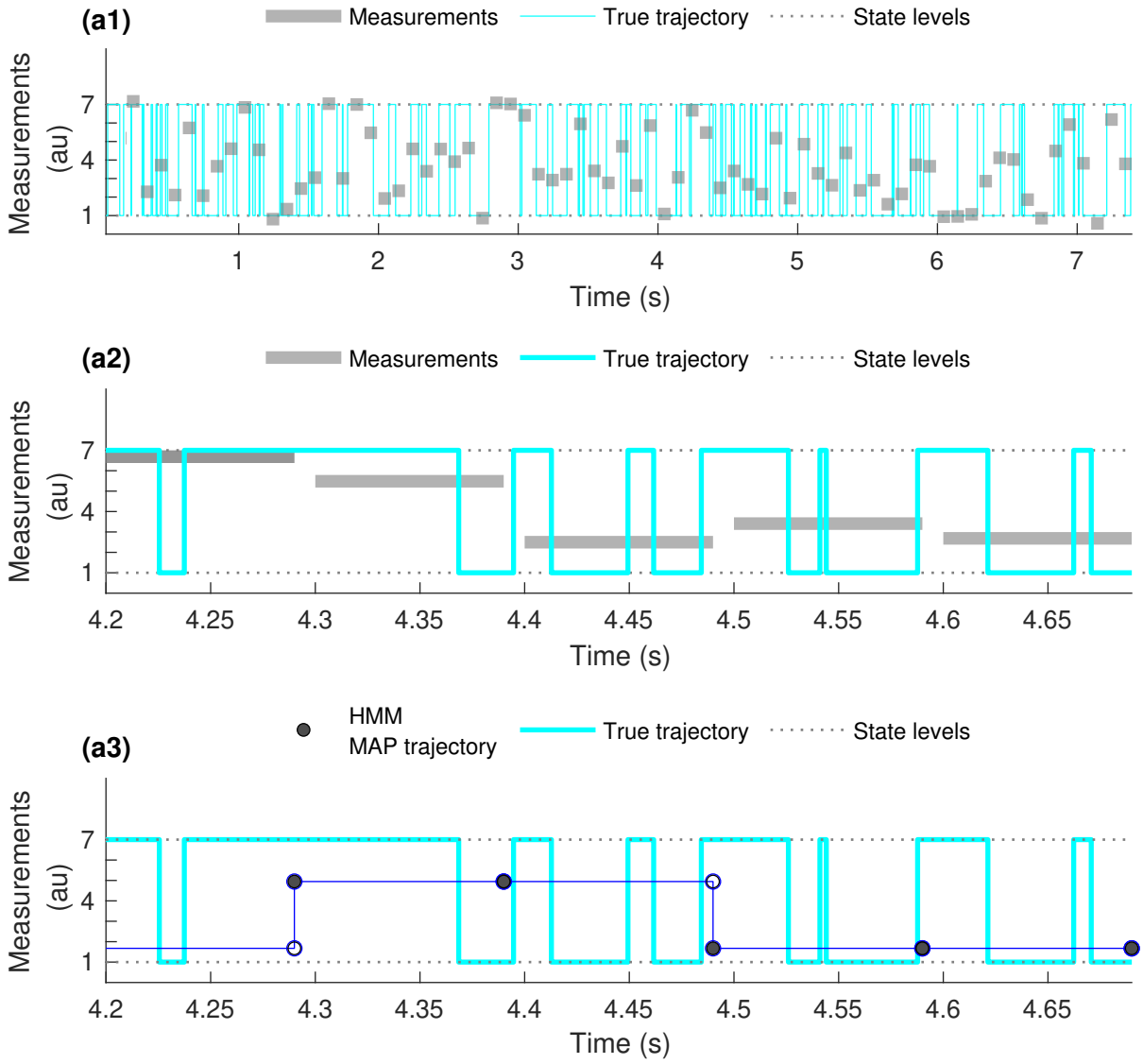

Fig. A.6. ***HMM trajectory estimates for fast state switching***. *Here we provide trajectory estimates obtained with the HMM when the switching rates are faster than the data acquisition rate,* $1/\Delta t = 10$ *(1/s). We follow identical coloring schemes and conventions as we do in Fig. A.1. We demonstrate that the HMM MAP trajectory poorly estimates the true trajectory.*

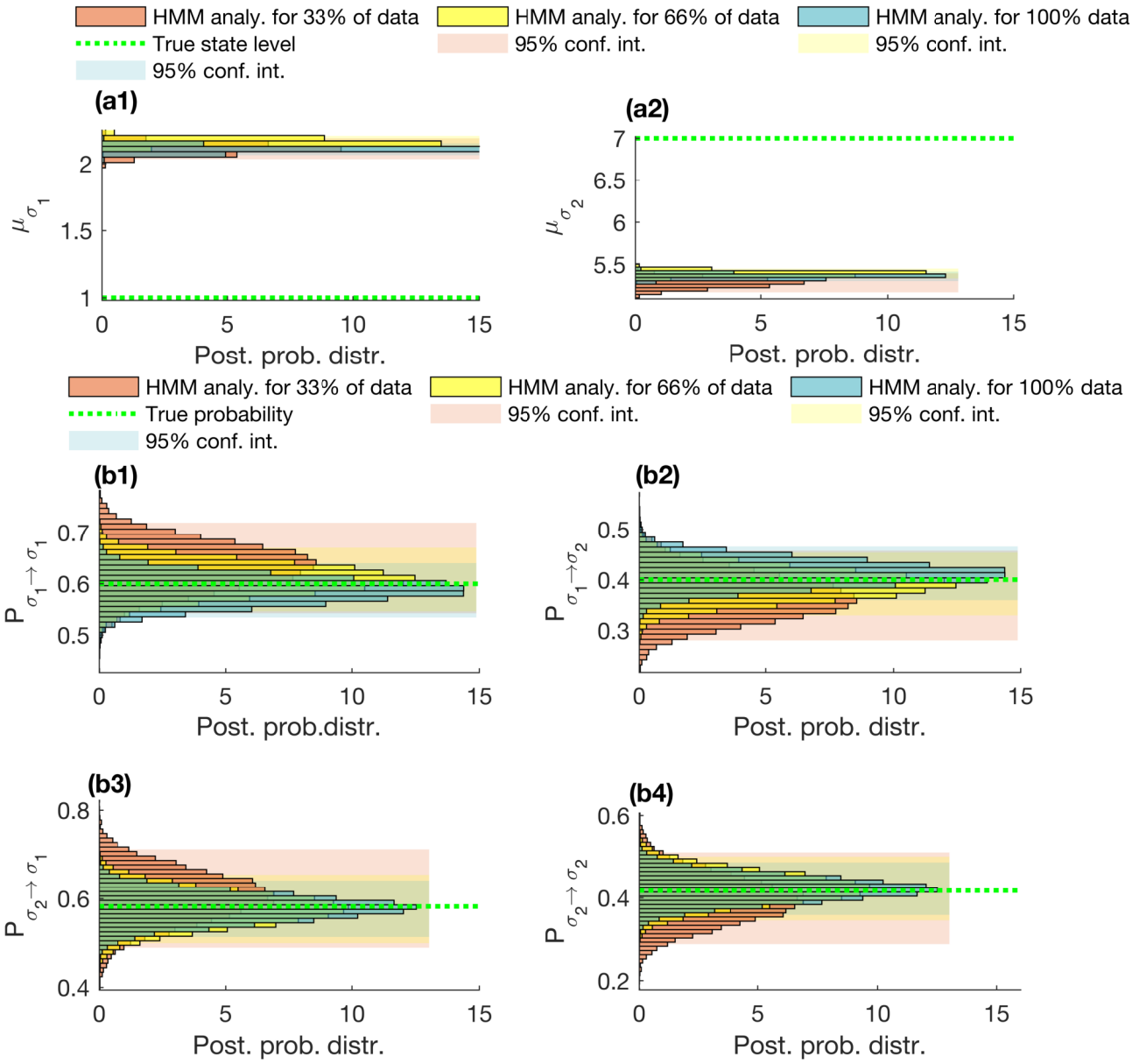

Fig. A.7. ***Finiteness of data analysis for HMM state level and transition probability estimates in the presence of fast state switching***. *Here, we would like to show the effect of set data set length on the state level and transition probability estimates provided by HMMs on the simulated measurements shown in Fig. A.6. To begin, in all panels, we see the superposed posterior distributions of the state levels and transition probabilities obtained from HMMs for* 33% *(orange),* 66% *(yellow),* 100% *(blue) of the entire simulated measurements along with their associated* 95% *confidence intervals, the true state levels and transition probabilities (green dashed line).*

in Fig. A.9.

We performed our investigation for the HMM in 3 parts. First, we analyzed 33% of the data, next 66% of the data and lastly we analyzed the entire data set with the HMM. We showed the simulated trajectory for the first 12% of the entire data set in Fig. A.6. In Fig. A.6 panel (a1), we show the true trajectory (cyan) and the measurements (gray rectangles). Next, in Fig. A.6 panel (a2) we showed the zoomed true trajectory and measurements. Finally, in panel (a3), we provide the superposed true trajectory with HMM MAP trajectory estimate. We observe that the HMM MAP trajectory estimate is a poor trajectory estimate. This poor trajectory estimates of HMM can be explained by the poorly estimated state levels, see Fig. A.7 panels (a1)-(a2). In Fig. A.7, we observe in both panels (a1)-(a2) that more data does not provide better state level estimates for HMMs. Subsequently, in Fig. A.7 panels (b1)-(b4), we demonstrate that the posterior distributions over transition probabilities obtained with HMM for all 3 data sets do not differ greatly from each other. Next we investigate the HMJP's performance in estimating transition probabilities with and without pre-specifying the ground truth trajectory.

A.1.5 Robustness Analysis with Respect to Learning the Trajectory Simultaneously with Transition Probability Estimates for HMJPs

In this section, we test the effect of estimating trajectories simultaneously with transition probability estimates using HMJPs. With sufficient data, we expect HMJPs to perform similarly well when: 1) the trajectory is assumed to be known; and 2) the trajectory is not assumed to be known and thus to be estimated as well. We demonstrated that posterior transition probabilities look similar when the trajectory is assumed to be known or not known in Fig. A.8.

Now we compare the performance of HMJP and HMM in estimating trajectories based on the average area under their posterior trajectory estimates.

A.1.6 Comparison of the Posterior Trajectory Estimates of HMM and HMJP

In this section, we provide a metric to compare the quality of the trajectory determination by HMMs and HMJPs (98). As we have a posterior distribution over trajectories $\mathcal{T}(.)$ obtained by HMMs and HMJPs, we also have posterior distributions over areas

$$\text{area under trajectories} = \int_{t_0}^{t_N} dt \mathcal{T}(t) \,. \tag{A.2}$$

In Fig. A.9, we compare the mean point statistic of *area under trajectories* for both HMM and HMJP.

We carried out this analysis for the simulated measurements investigated in the main text generated with slow switching rates Fig. 2 and fast switching rates Fig. 4.

We observe that, the average areas under HMJP posterior trajectory estimates are very closed to the areas under the true trajectories for slow and fast switching rate cases.

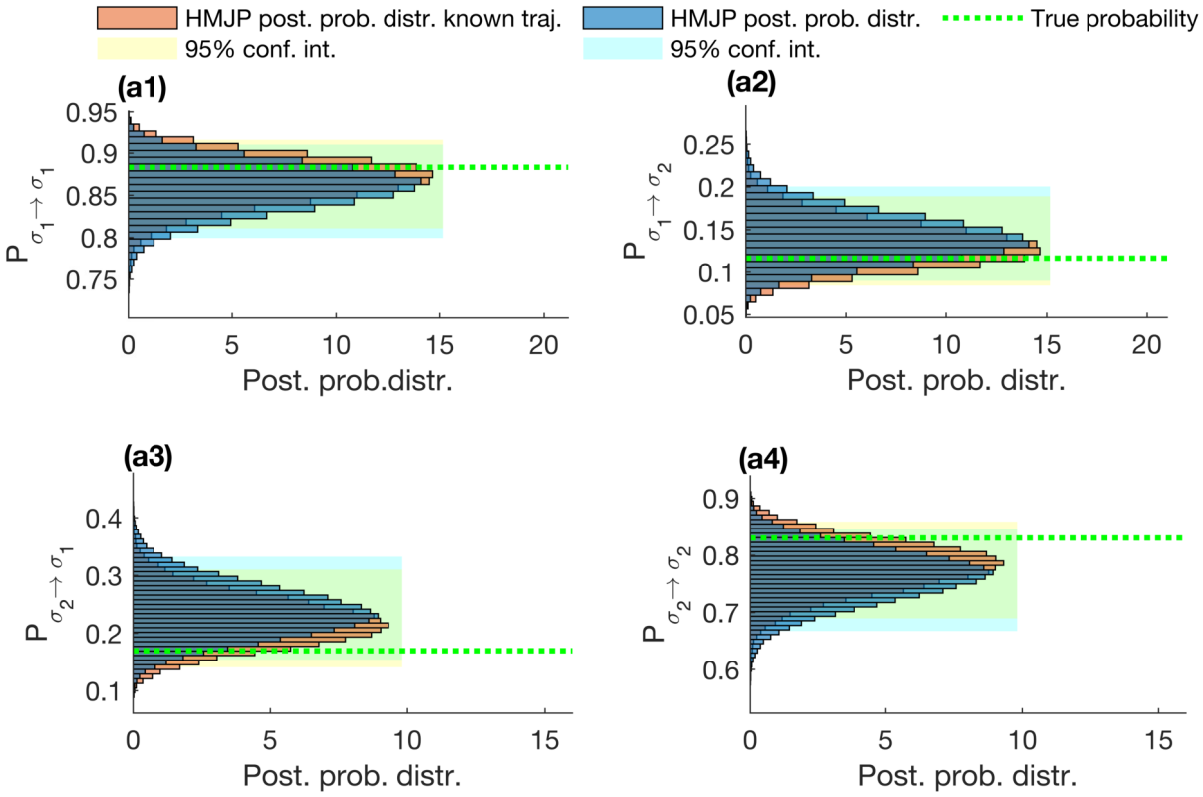

Fig. A.8. ***HMJP transition probability estimates with and without trajectory information***. *Here, we provide the analysis of the simulated measurements provided in Fig. 2 panel (a1). In this figure, we illustrate the HMJP's performance in estimating transition probabilities with and without learning the trajectory. To do so, in each figure panel we provide the superposed posterior distributions over transition probabilities labeled as* $P_{\sigma_k \to \sigma_{k'}}$ *for all* $k, k' = 1, 2$ *obtained by HMJP with unknown system trajectory (blue), with known system trajectory (orange) along with their associated* 95% *confidence intervals for the estimates and the true transition probabilities (dashed green lines). We observe that there is very limited change in the posterior transition probabilities based on their* 95% *confidence intervals when the trajectory is known or, alternatively, learned simultaneously.*

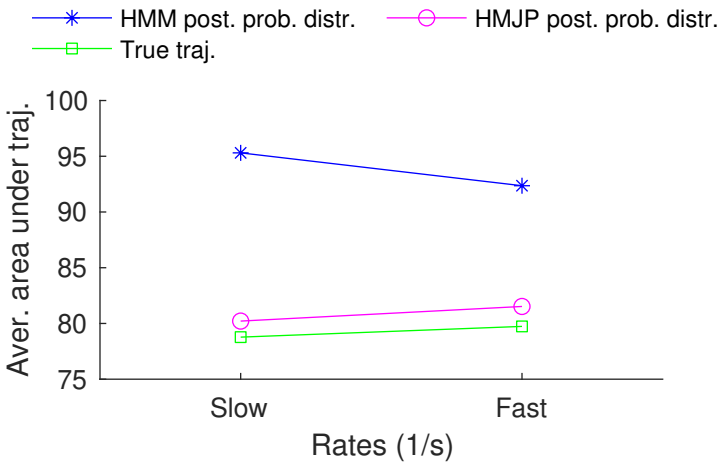

Fig. A.9. ***HMJP with HMM average area under the trajectory estimates***. *Here we compare the performance of HMJPs and HMMs in estimating trajectories based on the average area under their posterior trajectory estimate for the simulated measurements generated with slow switching rates (denoted with "Slow") , see Fig. A.1 and fast switching rates (denoted with "Fast"), see Fig. 4. In this figure, we provide the average areas under the HMM (blue star) and HMJP (magenta circle) posterior trajectory estimates and the area under the true trajectory (green circle) for both simulated measurements generated with slow switching rates and fast switching rates.*

A.2 Detailed Description of the Statistical Models

A.2.1 Hidden Markov Jump Process

A.2.1.1 Summary of Equations

$$\bar{\rho} \sim \textbf{Dirichlet}\left(\frac{\alpha}{2}, \frac{\alpha}{2}\right) \tag{A.3}$$

$$\lambda_{\sigma_1} \sim \textbf{Gamma}\left(\eta, \frac{b}{\eta}\right) \tag{A.4}$$

$$\lambda_{\sigma_2} \sim \textbf{Gamma}\left(\eta, \frac{b}{\eta}\right) \tag{A.5}$$

$$\mu_{\sigma_1} \sim \textbf{Normal}\left(H, V\right) \tag{A.6}$$

$$\mu_{\sigma_2} \sim \textbf{Normal}\left(H, V\right) \tag{A.7}$$

$$\bar{\pi}_1 \sim \textbf{Dirichlet}\left(0, A\right) \tag{A.8}$$

$$\bar{\pi}_2 \sim \textbf{Dirichlet}\left(A, 0\right) \tag{A.9}$$

$$s_0 \big| \bar{\rho} \sim \textbf{Categorical}\left(\tilde{\pi}_0\right) \tag{A.10}$$

$$d_m | s_m, \lambda_{\sigma_1}, \lambda_{\sigma_2} \sim \textbf{Exponential}\left(\bar{\lambda}_{s_m}\right) \qquad \text{for} \quad m = 0, 1, 2, ..., M-1 \tag{A.11}$$

$$s_{m+1} | s_m, \bar{\pi}_{\sigma_1}, \bar{\pi}_{\sigma_2} \sim \textbf{Categorical}\left(\tilde{P}_{s_m}\right) \qquad \text{for} \quad m = 0, 1, 2, ..., M-1 \tag{A.12}$$

$$x_m | \mathcal{T}\left(\cdot\right), \mu_{\sigma_1}, \mu_{\sigma_2} \sim \textbf{Normal}\left(\frac{1}{\tau} \int_{t_m - \tau}^{t_m} dt \mu_{\mathcal{T}(t)}, v\right) \qquad \text{for} \quad m = 1, 2, ..., M \tag{A.13}$$

where $\mathcal{T}(\cdot)$ is formulated as follows

$$\mathcal{T}(t) = \begin{cases} s_0 & \text{if} \quad t_0 \leq t < t_0 + d_0 \\ s_1 & \text{if} \quad t_0 + d_1 \leq t < t_0 + d_0 + d_1 \\ \vdots & \vdots \\ s_{M-1} & \text{if} \quad t_0 + d_0 + \cdots + d_{M-2} \leq t < t_0 + d_0 + d_1 + \cdots + d_{M-1}. \end{cases} \tag{A.14}$$

with M determined based on the first time

$$t_0 + \sum_{m=0}^{M-1} d_m \geq t_N. \tag{A.15}$$

In this manuscript as in the case of HMM framework, we set $\alpha = A = 1, \text{FWHM} = 0.25$, $H = \text{mean(data)}$ au, and $V = 1$ au.

A.2.1.2 Description of the Computational Scheme

Our MCMC exploits Gibbs sampling scheme (38, 39, 77). Accordingly posterior samples are generated by updating each variables involved sequentially by sampling conditioned on all other variables and the measurements $\mathbf{x}$. Conceptually, the steps involved in the generation of each posterior $\left(\mathcal{T}\left(\cdot\right), \bar{\bar{\pi}}, \bar{\rho}, \bar{\lambda}, \bar{\mu}\right)$ involve

1. Updating the trajectory $\mathcal{T}\left(\cdot\right)$,

2. Updating the transition probability matrix $\bar{\bar{\pi}}$ and $\bar{\rho}$,
3. Updating $\bar{\lambda}, \bar{\mu}$.

Specifically, for step 1, to sample the trajectory $\mathcal{T}(\cdot)$, we need two different strategies. As the state dynamics of the HMM evolve in discrete time, we use a *forward filtering backward sampling* scheme (34, 38–43). For the HMJP, as the state dynamics evolve in continuous time, we perform *uniformization* prior to sampling a new trajectory (64, 65, 68, 99, 100) and then we perform Gibbs sampling to obtain $s_{1:M}$.

For steps 2, 3 we have conjugate priors. Thus, we can do direct sampling.

It is in step 1, where we update the trajectory of the system, where the HMM and HMJP differ most in methodology and computational cost. This is because of the *uniformization* (99) required of the HMJP that sets M as described in Eq. (8).

A.2.2 Hidden Markov Model

A.2.2.1 Summary of Equations

For $K = 2$, the full set of HMM equations is

$$\bar{\rho} \sim \mathbf{Dirichlet}\left(\frac{\alpha}{2}, \frac{\alpha}{2}\right) \tag{A.16}$$

$$\bar{P}_{\sigma_1} \sim \mathbf{Dirichlet}\left(\frac{A}{2}, \frac{A}{2}\right) \tag{A.17}$$

$$\bar{P}_{\sigma_2} \sim \mathbf{Dirichlet}\left(\frac{A}{2}, \frac{A}{2}\right) \tag{A.18}$$

$$\mu_{\sigma_1} \sim \mathbf{Normal}\left(H, V\right) \tag{A.19}$$

$$\mu_{\sigma_2} \sim \mathbf{Normal}\left(H, V\right) \tag{A.20}$$

$$c_0 | \bar{\rho} \sim \mathbf{Categorical}\left(\bar{\rho}\right) \tag{A.21}$$

$$c_n | c_{n-1}, \bar{P}_{\sigma_1}, \bar{P}_{\sigma_2} \sim \mathbf{Categorical}\left(\bar{P}_{c_{n-1}}\right) \quad \text{for} \quad n = 2, ..., N \tag{A.22}$$

$$x_n | c_n, \mu_{\sigma_1}, \mu_{\sigma_2} \sim \mathbf{Normal}\left(\mu_{c_n}, v\right) \quad \text{for} \quad n = 1, ..., N \tag{A.23}$$

A.2.2.2 Description of the Computational Scheme

The joint probability distribution of our framework is $\mathbb{P}\left(\bar{\rho}, \bar{\bar{P}}, \bar{\mu}, \mathcal{T}(\cdot) \left| \mathbf{x} \right.\right)$. As we do not have conjugacy between dynamics and the measurements, it is not possible to have a direct sampling from the posterior distribution. For this reason, we develop a specialized Markov Chain Monte Carlo (MCMC) scheme that can be used to generate pseudorandom samples (77). This scheme is explained in detail below. A working implementation of the resulting scheme in source code and GUI forms provided through the author's website.

Our MCMC exploits Gibbs sampling scheme (38–41, 43, 77). Accordingly posterior samples are generated by updating each one of the variables involved sequentially by sampling conditioned on all other variables and the measurements $\mathbf{x}$. Conceptually, the steps involved in the generation of each posterior $\left(\bar{\rho}, \bar{\bar{P}}, \bar{\mu}, \vec{c}\right)$ are

1. Update the trajectory $\vec{c}$,
2. Update transition probabilities $\bar{\bar{P}}$,
3. Update transition probabilities $\bar{\rho}$,
4. Update transition probabilities $\bar{\mu}$.

Sampling the trajectory $\mathcal{T}(\cdot) = \vec{c}$ involves forward filtering-backward sampling scheme (38, 39). Updating the transition probability matrix can be carried out for every row separately on account of the independence of the Dirichlet prior on each $\bar{P}_{\sigma_k}$ and $\bar{\rho}$. Because of the conjugacy between the Dirichlet distribution and Categorical distribution, we have direct sampling updates for all $\bar{P}_{\sigma_k}$ and $\bar{\rho}$ (77). Similarly, due to conjugacy of the Normal distribution with itself, we also have direct updates for μ_{σ_k} for all $k = 1, 2$.

We mention that updating $\vec{c}$ is carried out via forward filtering backward sampling (77) approach (38, 39, 43, 77).

In the process of generating pseudorandom numbers from the posterior distribution described above, the first ~ 1000 pseudorandom numbers are discarded to account for MCMC burn-in. The rest of the generated numbers contribute to the posterior probability distribution over the trajectories, transition probability matrix and state levels.

Next, we provide the detailed information regarding these listed updates mentioned above.

A.2.2.3 Overview of the Sampling Updates

1. **Sampling a trajectory** $\mathcal{T}(\cdot)$. We will sample continuous time trajectories using *uniformization* (64, 68, 99) whose key steps 1, 2, 3 are illustrated in Fig. A.10. Uniformization uses ideas from discrete time-discrete state space Markov processes.

   To sample new trajectories, we start from rates, see Eq. (A.1) and old trajectories (see Fig. A.10 panel (a)) generating these trajectories using a Markov jump process. The trajectories are characterized by states "$\vec{S}$" and associated holding times,"$\vec{D}$" namely,

$$\mathcal{T}(\cdot) = \left(\vec{S}, \vec{D}, M\right) \quad \textbf{(A.24)}$$

$$\vec{S} = \{s_0, s_1, ..., s_{M-1}\} \quad \textbf{(A.25)}$$

$$\vec{D} = \{d_0, d_1, d_2, ..., d_{M-1}\}. \quad \textbf{(A.26)}$$

   Following Eq. (13) of the main, we can construct a generator matrix $\bar{\bar{G}}$. As a reminder, the system occupies each state s_m during the interval $[t_m, t_{m+1})$ with $d_m = (t_{m+1} - t_m)$ holding times for all $m = 0, 1, 2, ..., M-1$. As we can see from the cartoon figure in Fig. A.10 panel (b), we must now add virtual jumps. At the mth time interval, we add a number virtual jumps and distribute these *uniformly* in $t_{m+1} - t_m$. We select the number of jumps according to a Poisson distribution with intensity $\nu + \bar{\bar{G}}_{s_m s_m}$. Here $\nu > \max_k \lambda_{\sigma_K}$. To achieve this, we set $\nu = \left(\max_k \lambda_{\sigma_K}\right)\zeta$. In this manuscript, we choose $\zeta = 2$ as suggested in (68).

   As such we initially overestimate the number of jump locations. We subsequently need to prune these down by determining if these virtual jumps coincide with real jumps. In principle we can set ν to as large a value as we like, this only increases the computational cost of pruning.

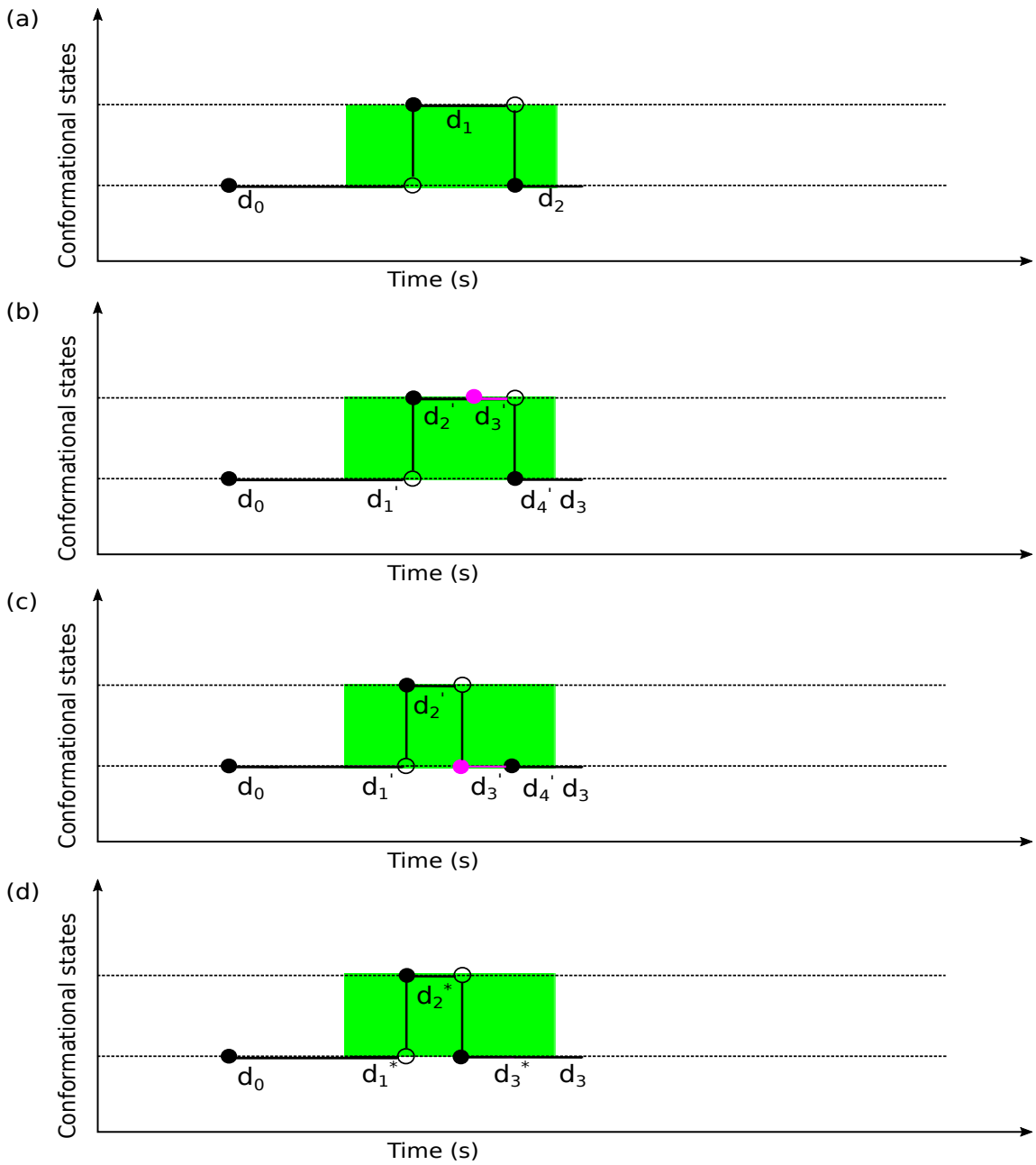

Fig. A.10. ***A conceptual description of uniformization on sampling a Markov jump process***. *Here, we explain how we perform uniformization to obtain a new trajectory estimate using the HMJP in* 4 *panels starting from an old trajectory estimate. In panel (a), we show the old trajectory estimate that we would like to update. This green region is associated with the* n^{th} *integration period that give rise to measurement* x_n*. We note that there is* 1 *jump in this old trajectory estimate with* 2 *holding times that are labeled with* d_1, d_2*. We simply define the states associated with* d_0, d_1, d_2, d_3 *as* s_0, s_1, s_2, s_3*, respectively. In panel (b), we introduce virtual jump events with circles in magenta and then we relabel the trajectory with new holding times and states. Upon relabeling, we now have states* $s_0, s_1', s_2', s_3', s_4', s_3$ *associated with the holding times* $d_0, d_1', ..., d_4', d_3$*. Next, we apply Gibbs sampling to sample each* s_m' *for all* $m = 1, 2, 3, 4$ *conditioned on all other variables. We show the newly sampled states in panel (c). The last step in this uniformization is to remove the self-transitions from the new sampled trajectory so that it remain a Markov jump process. To do so, we remove the self-transitions and in panel (d), we show the new sampled trajectory that is a Markov jump process with states* ${s_1}^*, {s_2}^*, {s_3}^*$ *and associated holding times* ${d_1}^*, {d_2}^*, {d_3}^*$*, respectively.*

To prune, as shown in panel (b) of Fig. A.10, we now must revert to a discrete Markov chain picture. We begin by defining the transition matrix $\bar{\bar{B}}$ that coincides with the Poisson intensity of $\nu + \bar{\bar{G}}_{s_m s_m}$. It is

$$\bar{\bar{B}} = \frac{\bar{\bar{G}}}{\nu} + \bar{\bar{I}} \tag{A.27}$$

Given a transition matrix $\bar{\bar{B}}$, we can now straightforwardly apply the forward filter-backward sampler to sample those states visited to sample the states at each virtual jump point.

Below we provide the details of sampling a part of a trajectory that is associated with observation x_n with Gibbs sampling scheme.

First, we write down the target distribution and then we factorize it as follows

$$\mathbb{P}\left(s'_{1:4}\middle|s_0, s_3, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}, x_n\right) \tag{A.28}$$
$$= \mathbb{P}\left(s'_4\middle|s_0, s_3, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}, x_n\right) \mathbb{P}\left(s'_3\middle|s'_4, s_0, s_3, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}, x_n\right) \tag{A.29}$$
$$\times \mathbb{P}\left(s'_2\middle|s'_3, s_0, s_3, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}, x_n\right) \mathbb{P}\left(s'_1\middle|s'_2, s_0, s_3, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}, x_n\right). \tag{A.30}$$

Next, we have to sample each part of this target distribution given above according to the following

(a)

$$\mathbb{P}\left(s'_1\middle|s'_2, s_0, s_3, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}, x_n\right) = \mathbb{P}\left(s'_1\middle|s'_2, s_0, x_n, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}\right) \tag{A.31}$$
$$\propto \mathbb{P}\left(s'_2\middle|s'_1\right)\left(s'_1\middle|s_0\right)\mathbb{P}\left(s'_1\middle|x_n, \dots\right) \tag{A.32}$$
$$\mathbb{P}\left(s'_1|x_n, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}\right) = \sum_{s_0}\mathbb{P}\left(s'_1, s_0|x_n, \dots\right) \propto \mathbb{P}\left(x_n|s_1, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}\right)\sum_{s_0}\mathbb{P}\left(s'_1|s_0\right) \tag{A.33}$$

(b)

$$\mathbb{P}\left(s'_2\middle|s'_3, s_0, s_3, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}, x_n\right) = \mathbb{P}\left(s'_2\middle|s'_3, x_n, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}\right) \propto \mathbb{P}\left(s'_3\middle|s'_2\right)\mathbb{P}\left(s'_2\middle|x_n, \dots\right) \tag{A.34}$$
$$\mathbb{P}\left(s'_2|x_n, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}\right) = \sum_{s'_1}\mathbb{P}\left(s'_2, s'_1|x_n, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}\right) \propto \mathbb{P}\left(x_n|s'_1, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}\right) \tag{A.35}$$
$$\times \sum_{s'_1}\mathbb{P}\left(s'_2|s'_1\right)$$

(c)

$$\mathbb{P}\left(s'_3\middle|s'_4, s_0, s_3, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}, x_n\right) = \mathbb{P}\left(s'_3\middle|s'_4, x_n, \dots\right) \propto \mathbb{P}\left(s'_4\middle|s'_3\right)\mathbb{P}\left(s'_3\middle|x_n, \dots\right) \tag{A.36}$$
$$\mathbb{P}\left(s'_3|x_n, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}\right) = \sum_{s'_2}\mathbb{P}\left(s'_3, s'_2|x_n, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}\right) \propto \mathbb{P}\left(x_n|s'_2, \dots\right)\sum_{s'_2}\mathbb{P}\left(s'_3|s'_2\right) \tag{A.37}$$

(d)

$$\mathbb{P}\left(s'_4\middle|s_0, s_3, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}, x_n\right) = \mathbb{P}\left(s'_4\middle|s_3, x_n, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}\right) \propto \mathbb{P}\left(s_3\middle|s'_4\right) \tag{A.38}$$
$$\times \mathbb{P}\left(s'_4\middle|x_n, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}\right) \tag{A.39}$$
$$\mathbb{P}\left(s'_4|x_n, \dots\right) = \sum_{s'_3}\mathbb{P}\left(s'_4, s'_3|x_n\right) \propto \mathbb{P}\left(x_n|s'_4, d'_{1:4}, \mu_{\sigma_1}, \mu_{\sigma_2}\right) \tag{A.40}$$
$$\times \sum_{s'_3}\mathbb{P}\left(s'_4|s'_3\right).$$

We repeat this process of sampling trajectory for all observations x_n where $n = 1, 2, ..., N$. We should note that $\bar{\bar{B}}_{s'_k s'_{k'}} = \mathbb{P}\left(s'_k | s'_{k'}\right)$ with $k, k' = 1, 2, 3, 4$.

Once we obtain the new trajectory (see Fig. A.10 panel (c)) we observe that this new trajectory includes self transition events at the jump times.

However, occupying the same steps after the jump times is not allowed for continuous time processes, therefore we can not have self transition events in Markov jump processes. To obtain a Markov jump process from this sampled new trajectory, we drop the self transition events (see Fig. A.10 panel (d)) thereby obtaining a new trajectory which is a Markov jump process.

Next we explain how to update the transition probability matrix $\bar{\bar{\pi}}$.

2. **Sampling transition probabilities $\bar{\bar{\pi}}$ and $\bar{\rho}$.** We placed conjugate Dirichlet distributions to the Categorical distribution for the rows of $\bar{\bar{\pi}}$. Therefore, we update each row of $\bar{\bar{\pi}}$ via direct sampling as we do for the case of the HMM

$$\mathbb{P}\left(\bar{\pi}_k \middle| \mathbf{x}, s_{0:M}, d_{1:M}, \bar{q}, \bar{\mu}\right) = \left(\prod_{m=1}^{M} \prod_{s_{n-1}=\sigma_k} \mathbb{P}\left(s_m | s_{m-1} = \sigma_k, \bar{\pi}_k\right)\right) \mathbb{P}\left(\bar{\pi}_k\right). \tag{A.41}$$

Similar direct sampling formulation applies to $\bar{\rho}$ as follows

$$\mathbb{P}\left(\bar{\rho} \middle| \mathbf{x}, s_{0:M}, d_{1:M}, \bar{q}, \bar{\mu}\right) \propto \mathbb{P}\left(s_0 \middle| \bar{\rho}\right) \mathbb{P}\left(\bar{\rho}\right). \tag{A.42}$$

Next, we provide the details of sampling $\bar{\lambda}_{\sigma_k}$ for all $k = 1, 2$.

3. **Sampling $\bar{\lambda}$.** We placed conjugate Gamma prior distributions to the Exponential distribution for $\bar{\lambda}_{\sigma_k}$ where $k = 1, 2$. Therefore updating all $\bar{\lambda}_{\sigma_k}$ with $k = 1, 2$ is carried out with direct sampling based on the following formulation

$$\begin{aligned}
\mathbb{P}\left(\bar{\lambda}_{\sigma_k} \middle| \bar{\bar{\pi}}, \mathbf{x}, s_{0:M}, d_{1:M}, \bar{\mu}\right) &\propto \mathbb{P}\left(\bar{\lambda}_{\sigma_k} \middle| s_{0:M}, d_{1:M}\right) = \prod_{n=1}^{M} \prod_{s_{n-1}=\sigma_k} \mathbb{P}\left(d_n | s_{n-1} = \sigma_k\right) \mathbb{P}\left(\bar{\lambda}_{\sigma_k}\right) && \textbf{(A.43)} \\
&= \prod_{n=1}^{M} \prod_{s_{n-1}=\sigma_k} \mathbf{Exponential}\left(d_n; \bar{\lambda}_{\sigma_k}\right) \mathbf{Gamma}\left(\bar{\lambda}_m; \eta, \frac{b}{\eta}\right) \\
&= \bar{\lambda}_{\sigma_k}^{\left(\sum_{n=1}^{M} \mathbb{I}\left(s_{n-1}=\sigma_k\right) + \frac{\eta}{2} - 1\right)} \exp\left(-\bar{\lambda}_{\sigma_k}\left(\sum_{n=1}^{M} \sum_{s_{n-1}=\sigma_k} d_n + \frac{\eta}{b}\right)\right) && \textbf{(A.44)} \\
&= \mathbf{Gamma}\left(\bar{\lambda}_{\sigma_k}; \sum_{n=1}^{M} \mathbb{I}\left(s_{n-1} = \sigma_k\right) + \eta, \frac{1}{\sum_{n=1}^{M} \sum_{s_{n-1}=\sigma_k} d_n + \frac{\eta}{b}}\right)
\end{aligned}$$

for $k = 1, 2$. In this formulation, $\mathbb{I}$ denotes the indicator function and for example $\mathbb{I}\left(s_n = 2\right)$ is 1 if $s_n = 2$ for some $n = 1, 2, ..., M$ and 0 otherwise.

4. **Sampling $\bar{\mu}$.** We perform the updates of μ_{σ_k} for all $k = 1, 2$ as a part of the Gibbs sampling scheme. Therefore, we need the full conditional distributions of μ_{σ_k} for all $k = 1, 2$ these are $\mathbb{P}\left(\mu_{\sigma_1} | \mu_{\sigma_2}, s_{0:M}, d_{1:M}, \mathbf{x}, \bar{\bar{\pi}}, \bar{\lambda}\right)$, $\mathbb{P}\left(\mu_{\sigma_2} | \mu_{\sigma_1}, s_{0:M}, d_{1:M}, \mathbf{x}, \bar{\bar{\pi}}, \bar{\lambda}\right)$. Below, we provide the explicit update formula for μ_{σ_1} and the similar for-

mulation holds for μ_{σ_2}.

$$\mathbb{P}\left(\mu_{\sigma_1}|\mu_{\sigma_2}, s_{0:M}, d_{0:M}, \mathbf{x}\right) \propto \left(\prod_{n=1}^{N} \mathbb{P}\left(x_n|\mu_{\sigma_1}, \mu_{\sigma_2}, s_{0:M}, d_{0:M}\right)\right) \mathbb{P}\left(\mu_{\sigma_1}\right) \tag{A.45}$$

$$\propto \quad \exp\left(\mu_{\sigma_1}\left(H\frac{1}{V} + \sum_{n=1}^{N}\beta\left(x_n - \sum_{k_n=1}^{n*}\mathbb{I}\left(s_{k_n} \neq 1\right)\mu_{s_{k_n}} d_{k_n}\right)\right)\right) \tag{A.46}$$

$$\times \quad \exp\left(-\mu_{\sigma_1}^2\left(\frac{\frac{1}{V}}{2} + \sum_{n=1}^{N}\frac{\beta}{2}\left(\sum_{k_n=1}^{n*}\mathbb{I}\left(s_{k_n} = 1\right) d_{k_n}\right)^2\right)\right) \tag{A.47}$$

$$\propto \quad \mathbf{Normal}\left(\mu_{\sigma_1}; \frac{\left(H\frac{1}{V} + \sum_{n=1}^{N}\beta\left(x_n - \sum_{k_n=1}^{n*}\mathbb{I}\left(s_{k_n} \neq 1\right)\mu_{s_{k_n}} d_{k_n}\right)\right)}{\left(\frac{1}{V} + \sum_{n=1}^{N}\beta\left(\sum_{k_n=1}^{n*}\mathbb{I}\left(s_{k_n} = 1\right) d_{k_n}\right)^2\right)}, \frac{1}{\left(\frac{1}{V} + \sum_{n=1}^{N}\beta\left(\sum_{k_n=1}^{n*}\mathbb{I}\left(s_{k_n} = 1\right) d_{k_n}\right)^2\right)}\right)$$

where $\mathbf{x} = (x_{1:N})$.

A.3 Notation

Table A.1. *Notation conventions*

Description	Variable	Units
n^{th} measurement	x_n	au
Total number of measurements	N	–
Total number of states	K	–
Total number of jumps	M	–
System trajectory	$\mathcal{T}(\cdot)$	–
HMJP trajectory	$(\vec{S}, \vec{D}, M)$	–
HMM trajectory	$\vec{c}$	–
State of the system in HMM at time t_n	c_n	–
State of the system in HMJP at n^{th} jump	s_n	–
Holding time of the system in HMJP at n^{th} jump	d_{n+1}	s
Initial transition probability matrix	$\bar{\rho}$	–
Transition probability matrix in HMM	$\bar{\bar{P}}$	–
Probability of transitioning from state σ_k to $\sigma_{k'}$ in HMM	$P_{\sigma_k \to \sigma_{k'}}$	–
Transition probability for the embedded discrete Markov chain in HMJP	$\bar{\bar{\pi}}$	–
Escape rate for the kth state of the system	$\lambda_{\sigma_K} = -\lambda_{\sigma_k \to \sigma_k}$	1/s
Escape rate vector	$\bar{\lambda}$	–
State levels	$\bar{\mu}$	au
Generator matrix in HMJP	$\bar{\bar{G}}$	1/s
Rate of transitioning form state σ_k to $\sigma_{k'}$ in HMJP	$\lambda_{\sigma_k \to \sigma_{k'}}$	1/s
Initial time of the experiment	t_0	s
End time of the experiment	t_N	s
Integration period	τ	s
Data acquisition period	Δt	s
Variance of measurements	v	au
Representative of nth data acquisition period	$[t_{n-1}, t_n]$	–
Representative of nth integration period	$[t_n - \tau, t_n]$	–
Concentration parameter for $\bar{\rho}$	α	–
Concentration parameter for $\bar{\bar{\pi}}$ and $\bar{\bar{P}}$	A	–
Shape parameter	η	au
Scale parameter	b	s^{-1}
State level mean	H	au
State level variance	V	au

Table A.2. *Parameter choices and units*

Variable	H	V	η	b	A	α
Units	au	au	au	s^{-1}	-	-
Fig. 2	mean of measurements	1	2	300	1	1
Fig. 3	mean of measurements	1	2	300	1	1
Fig. 4	mean of measurements	1	2	300	1	1
Fig. 5	mean of measurements	1	2	300	1	1
Fig. A.1	mean of measurements	1	2	300	1	1
Fig. A.2	mean of measurements	1	2	300	1	1
Fig. A.3	mean of measurements	1	2	300	1	1
Fig. A.4	mean of measurements	1	2	300	1	1
Fig. A.6	mean of measurements	1	2	300	1	1
Fig. A.7	mean of measurements	1	2	300	1	1
Fig. A.8	mean of measurements	1	2	300	1	1
Fig. A.5	mean of measurements	1	2	300	1	1