Achieving subtemporal resolution in the analysis of two-state single-molecule trajectories

Erik Clarkson and Tobias Ambjörnsson*
Centre for Environmental and Climate Science, Lund University, Lund, Sweden

(Received 25 September 2025; accepted 28 December 2025; published 30 January 2026)

Although spatial resolution in fluorescence microscopy and related fields has advanced to the nanometer scale, time resolution has remained essentially unchanged and is set by the camera system's imaging time. Yet adequate time resolution is crucial for information acquisition about, for instance, dynamical processes in cells. This acquisition typically proceeds by analyzing biomolecular trajectories from single-particle tracking experiments, in terms of a standard discrete-time hidden Markov model. This type of analysis assumes, however, that subsampling time events that happen during imaging time can be neglected, which is rarely the case. To remedy this, we here introduce an algorithm that efficiently calculates the exact contribution of state switches to the likelihood of observed trajectories. This is made possible by our analytic derivation of generalized transition probabilities—which we call *transition-accretion probabilities*—that probabilistically capture unseen switching behavior during data acquisition. We do *in-silico* Bayesian model selection and parameter inference, and demonstrate that our *subsampling time hidden Markov model* approach outperforms the standard variant (applicable to slow kinetics). Our method and associated free-to-use software opens up for precise and reliable parameter estimation across a variety of single-molecule experiments, irrespective of the temporal resolution of the setup.

DOI: 10.1103/ccdh-hhx8

I. INTRODUCTION

Single-molecule experiments and theory enable precise quantitative analysis beyond ensemble averages [1]. For example, single-particle tracking (SPT) reveals individual biomolecular trajectories, allowing high-accuracy estimation from fluctuating quantities [2] while preserving biological functionality [3]. Key estimates include kinetic rates for biochemical reactions and diffusion constants for transport phenomena. These techniques advance research in cell communication [4], molecular motors [3,5], drug discovery [3,6], and synthetic molecular motor construction [7].

The resolution of single-molecule trajectories—both spatial and temporal—determines analysis complexity. High spatial resolution is needed to identify molecules and estimate positions, while high time resolution is essential for capturing dynamics, such as switching and binding. However, these resolutions are interlinked and constrained by photon collection: Greater localization precision requires longer imaging, reducing temporal resolution—*the photon budget dilemma* [8]. As a result, most setups are limited to millisecond sampling times [3], precluding detection of rapid biological events like molecular binding or conformational changes.

Biomolecular interactions often induce state switching in single-molecule experiments. In SPT, this can occur due to, e.g., binding/unbinding events or conformational changes [9]. Analyzing this switching behavior is particularly challenging when transitions occur faster than the imaging time; this is illustrated in Fig. 1. Two additional effects are motion blur [10,11] (due to molecule motion during integration time) and localization errors [10,12] (due to center-of-mass positional estimation), both of which make the analysis more difficult. State-of-the art methods for analysis of diffusion trajectories with hidden state switching use hidden Markov models (HMM), where molecules "emit" observations based on time-dependent states. In the HMM, state switching happens instantaneously in-between sampling times (discrete time); every observed displacement is assumed to correspond to a single state. This assumption is reasonable only for slow switching in relation to the sampling time. The HMM method has been complemented by maximum likelihood estimation, Bayesian inference, and Markov chain Monte Carlo sampling [12,13]. Alternative strategies include variational Bayes [11,14] and Viterbi-based approaches [15]. Recently, Simon *et al.* [16] took a first step toward including subsampling time switching events in the analysis. However, the method in Ref. [16] is approximate and treats the subsampling time states in a brute-force discrete-time manner, which makes the method computationally infeasible for general switching dynamics.

To deal with state-switch effects in SPT, we here introduce *subsampling time HMM (stHMM)* for exact yet feasible integration over hidden states across sampling times (see Ref. [17] for our public Matlab code). The input to our stHMM pipeline is continuous-time data, which has been averaged over a set of sampling-time windows. We marginalize over all possible

*Contact author: tobias.ambjornsson@mgeo.lu.se

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by Bibsam.

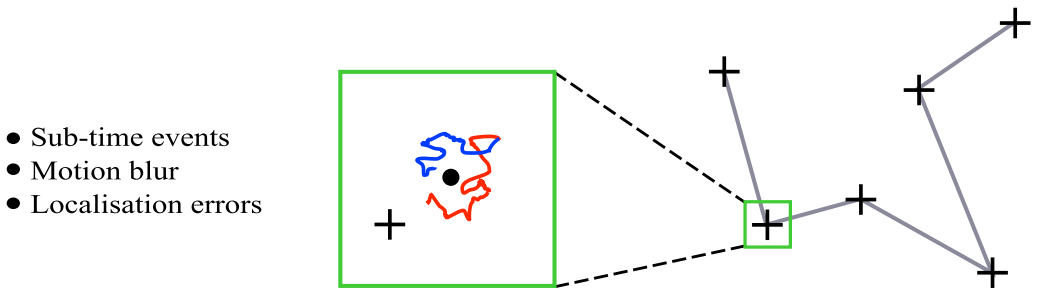


FIG. 1. The SPT inference problem. Observed trajectories (right) are constructed from measured positions, denoted by + signs. The underlying motion (unobserved) is, however, more intricate as can be seen in the enlarged square (left). First, the recorded center-of-mass position of the biomolecule is its average position during imaging time (motion blur [10,11]), seen as the red-blue curve. Second, the measured positions deviate from the true positions—here denoted by a black dot—due to the limited number of photons and errors in detection algorithms (localization noise [10,12]). Third, the biomolecule may interact with its environment and thus switch diffusive state at any time during camera integration. This third SPT challenge is the main focus herein and is illustrated by the coloring of the curve, where each color represents a free/bound state. To analyze this continuous-time data averaged over sampling-time windows (i.e., account for subsampling time switches, motion blur, and localization noise), we use an HMM-like likelihood, built on our transition-accretion probabilities coupled with a single distribution of observed displacements that includes contributions from both diffusive states. Using the likelihood in a Bayesian framework allows us to, for the first time, do model selection and parameter estimation of diffusion constants and kinetic rates, irrespective of the exposure time of the imaging setup.

hidden states, both at the end of every sampling time interval n (parametrized by s_n similar to the HMM) and also inside those intervals (parametrized by w_n—the fraction of time spent in state 1). Our stHMM is based on theoretical results in the form of generalized transition probabilities $P(w_n, s_n|s_{n-1}, \theta)$, henceforth referred to as *transition-accretion probabilities*. Analytic expressions for these are given in Sec. III and derived in Appendix C, using techniques from mathematical physics. The transition-accretion probabilities statistically describe accumulated subsampling time events, thus enabling a continuous-time capacity in an HMM-like framework. A forward algorithm ensures computational feasibility, maintaining linear scaling with trajectory length [18]. Additionally, by using nested sampling we integrate stHMM into a Bayesian framework for parameter inference and model selection. Experimentally, stHMM mitigates the photon budget dilemma, allowing extended imaging times without losing resolution in rapid kinetics.

This article is structured as follows: In Sec. II, we demonstrate that stHMM has a superior inference performance compared to HMM on simulated trajectories. Section III presents our theoretical foundations and describes our procedure for Bayesian inference and model selection within stHMM. Section IV discusses our results, and we also give an outlook and list fields where stHMM may find future applications.

II. RESULTS

In a two-state diffusion process, a biomolecule can switch between two states (labeled by 1 and 2) with rates k_{12}, k_{21}. When in state 1 (2), the molecule performs a random walk characterized by a diffusion constant D_1 (D_2) (see Fig. 1). The challenge at hand is to estimate $\theta = \{D_1, D_2, k_{12}, k_{21}\}$ from a single, or a set of, biomolecular trajectories. To tackle this challenge, we here use synthetic trajectories and compare our parameter estimation method (stHMM) to the current state-of-the-art method (HMM). By using synthetic data, we know the "true model" and its ground truth parameter values, as required for quantitative model selection and parameter estimation. Note that in our analysis pipelines we neglect correlations in observed displacements due to localization errors and motion blur (see Sec. III B 2).

A. Simulation results

Figure 2 demonstrates the underlying motion for two-state diffusion trajectories, where blue codes for the bound state and red for the free state (see Sec. III A for simulation details). We also see observed trajectories, whose positions are marked by black plus-signs and whose displacements are gray connecting lines. The plus-signs are formed from the mean molecule position during imaging time (motion blur) with additive errors (localization errors). The localization error variance σ_E^2 is here assumed to be known from control experiments, e.g., analysis of immobilized fluorophores [12].

Different kinetic switching regimes correspond to "slow," "intermediate," and "fast" switching. The expected number of binding events ν_1 (unbinding events ν_2) during a time interval is

$$\nu_1 = k_{21}\tau, \quad \nu_2 = k_{12}\tau, \tag{1}$$

where τ is the exposure (sampling) time. Figure 2(a) shows a trajectory with slow kinetics, where the biomolecule is in the bound state during the three sampling times. For longer trajectories, the molecule will also be in the free state. We quantify the state-switching dynamics in terms of w_n, the fraction of time spent in state 1 during the nth exposure time (the nth observed displacement in the trajectory). The corresponding histogram in Fig. 2(b) has color-coded w_n values for the molecule state. Evidently, transitions in the slow switching regime ($\nu < 0.1$, where $\nu = \nu_1 = \nu_2$) are so rare that practically all observations come from state 1 or state 2. For intermediate kinetics ($0.1 < \nu < 0.4$), seen in Fig. 2(c), there are observed displacements due to both underlying states. Subsampling time state transitions form a uniform

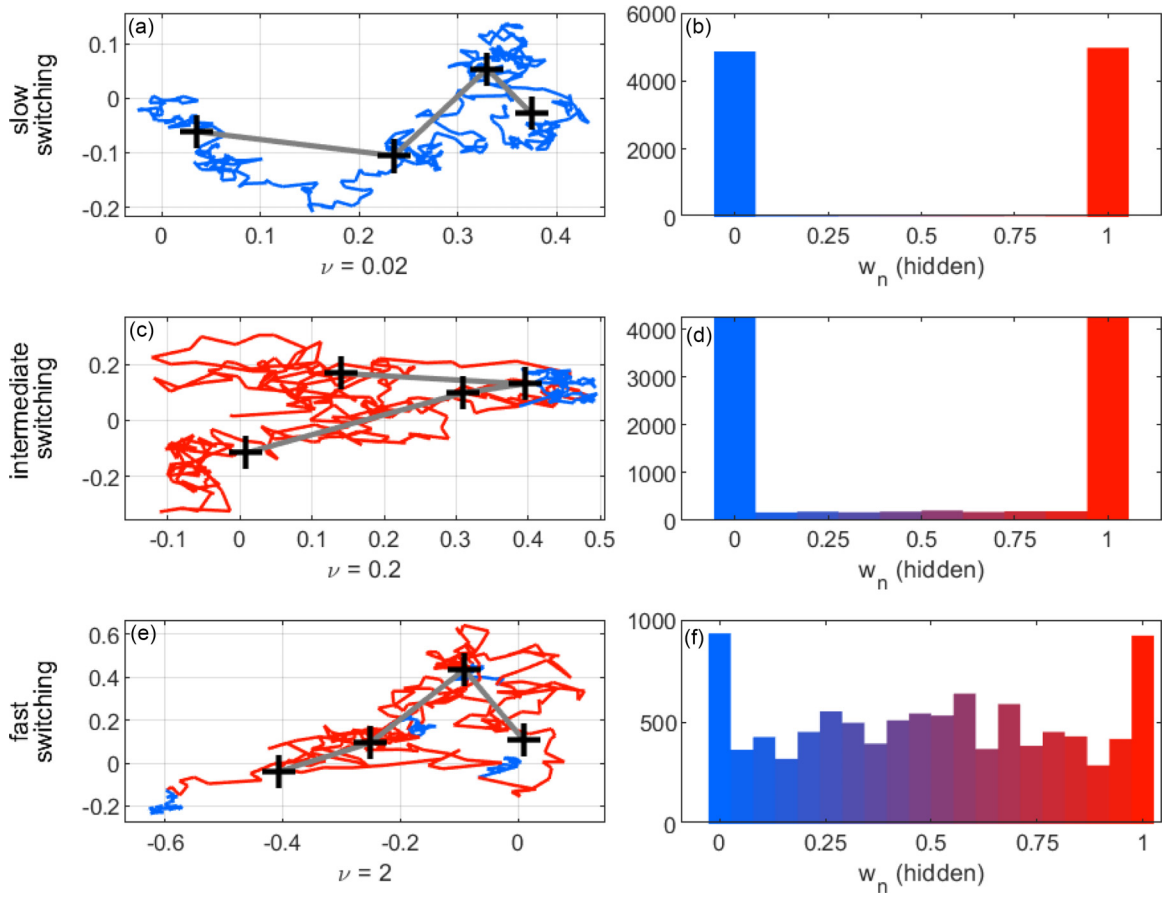


FIG. 2. Synthetic two-state diffusion data. The data for each row of panels are a trajectory of 10^4 sampling times, where the left panels [(a), (c), and (e)] show three such sampling times color-coded according to the free (red) or bound (blue) state at each time, as in Fig. 1. These truncated underlying trajectories are plotted at a time resolution of 100 subtimes per sampling time. Observed trajectories, which include also motion blur and localization errors, are marked with plus signs at estimated positions connected by gray lines. The gray lines show *observed displacements* used as input in the data analysis pipelines. x and y coordinates are in units of μm and mean number of state transitions per time interval is ν. Right [panels (b), (d), and (f)]: corresponding histogram statistics of state 1 weights w_n for all 10^4 sampling times. If state transitions happen during time interval n, the corresponding observed displacement is distributed according to a w_n-dependent probability density function (PDF), see Sec. III B 2. Diffusion constants are $D_1 = 10\ \mu\text{m}^2/\text{s}$ and $D_2 = 1\ \mu\text{m}^2/\text{s}$, and the exposure time is $\tau = 0.005$ s.

distribution of w_n values between 0 and 1 [see Fig. 2(d)], as the molecule is equally likely to switch state at any time during camera exposure. For rapid kinetics ($\nu > 0.4$), the subsampling time state transitions are frequent, as seen in Fig. 2(e). Most observed displacements will therefore contain hidden state switches. Likewise, the histogram in Fig. 2(f) reveals how the distribution of w_n values becomes centered on $k_{21}/(k_{12} + k_{21})$ ($= 0.5$ for our choice of parameters), which is the stationary probability of the biomolecule residing in state 1.

The observed trajectories in Fig. 2 (plus-sign marked) serve as synthetic data to be analyzed herein. Using our stHMM analysis pipeline, we demonstrate how to do both model selection and parameter estimation, and compare to the ground truth.

B. Model selection

Our first challenge is to do model selection, i.e., tackle the problem "what model does the data come from?" (see Sec. III C 3). The SPT data, as exemplified by the plus-sign-marked gray trajectories in Figs. 2(a), 2(c), and 2(e), serve as input data to our stHMM analysis pipeline. The (unobserved) underlying dynamics includes motion blur and localization errors in $d = 2$ dimensions (Sec. III A) as well as subsampling time state switches (the underlying motion is seen in the colored trajectories in Fig. 2). In our model selection, we compare the model support of one-state diffusion (with a diffusion constant D as its only parameter), HMM and stHMM. Note the slightly liberal terminology here: We name the models by the corresponding analysis pipelines; e.g., the HMM pipeline is simply called "HMM." The model probabilities are proportional to a quantity known as the evidence of that model [Z in Eq. (4)]. It requires the data, a model (a likelihood function), and a prior (Table I) as input.

Analysis of experimental data begins with picking a suitable model. Following Bayesian decision theory [19], the optimal model is the one that minimizes the probability of misclassification error, $1 - P_{\text{chosen}}$, where P_{chosen} is the probability of the chosen model. Earlier hidden-state SPT studies

TABLE I. Prior parameters as used for the analysis of SPT trajectories. All individual priors are lognormal with parameters μ_p and σ_p, which are the logarithm of the location and scale parameter, respectively. Units for D (one-state diffusion constant), D_1, and D_2 are $\mu m^2/s$, while k_{12} and k_{21} have units of inverse time (1/s).

Parameter	μ_p	σ_p
D	$\log((D_1+D_2)/2)$	$\log(1.5)$
D_1	$\log(D_1)$	$\log(1.5)$
D_2	$\log(D_2)$	$\log(1.5)$
k_{12}	$\log(200)$	$\log(3.5)$
k_{21}	$\log(200)$	$\log(3.5)$

failed to provide this number. We here, for the first time in the SPT field, do model selection of hidden two-state and one-state models via direct calculation of the evidence, using nested sampling. Through our method, we have the error probability of our model choice before turning to parameter estimation.

Figure 3 shows the posterior model probabilities inferred from the analyzed SPT data, given our set of candidate models. In Fig. 3(a), the underlying diffusion constants are separated by a factor of 10. The stHMM model selection capabilities here quantitatively confirm our expectation of two-state dynamics by assigning the one-state model a negligible probability. The question that remains is whether the continuous-time nature of the kinetics is significant (stHMM) or negligible (HMM). We see that, overall, stHMM always has a probability >0.6 and a probability >0.8 above $\nu = 0.05$. We conclude that stHMM is the best descriptor of the data. From $\nu = 0.5$ onward, we see a slight decline in the stHMM model probability—a hint that the observations begin to look similar and become more reminiscent of a one-state model. Indeed, the figure inset, which has a logarithmic vertical axis, shows the one-state probability rising rapidly for increasing ν [more clearly seen in Fig. 3(b)].

In Fig. 3(b), the underlying diffusion constant of the free state is made smaller. Compared to Fig. 3(a), the difference between stHMM and HMM probabilities is smaller and stHMM's peak around $\nu = 0.35$ is more pronounced. This is expected, as stHMM's ability to discern the two diffusion constants makes less of a difference when their separation is small. Furthermore, we are shown that it is not as clear that we are dealing with a two-state system. As such, we see a significant probability of the one-state model for the fastest kinetics ($\nu = 4$), where the stHMM probability falls sharply and also the HMM probability falls. The model selection capabilities provided by our stHMM pipeline thus capture the fact that there is less kinetic information in the data whenever the temporal resolution of the camera is low in relation to the studied kinetics. For extremely low temporal resolution, the one-state model should become the most probable one due to its simplicity (Occam's razor [20]).

Note that our model choice criteria favor stHMM over HMM in all regimes. That in turn suggests the importance of analyzing data with stHMM over HMM, especially for intermediate or fast switching dynamics.

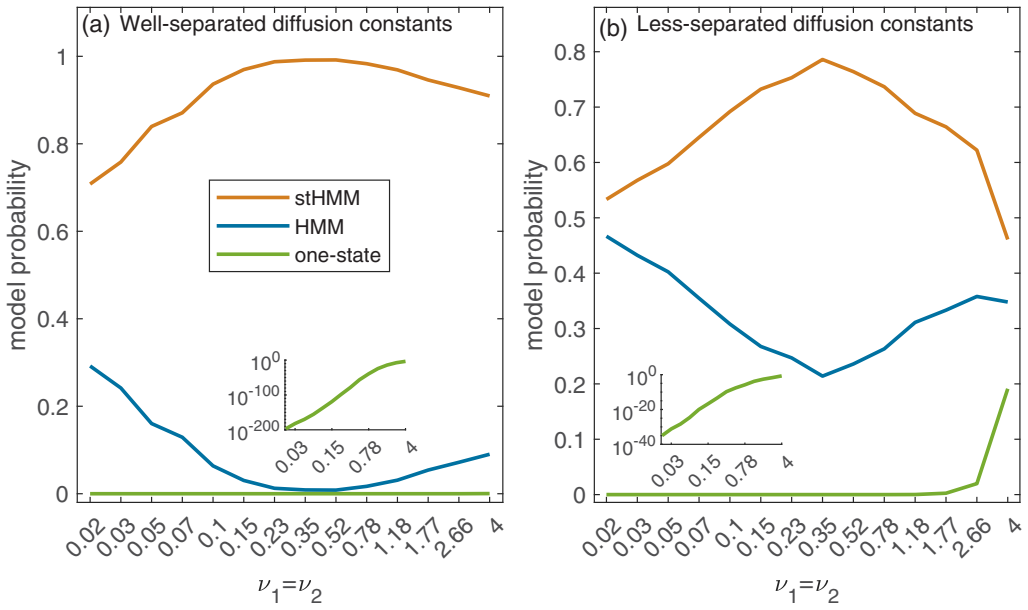


FIG. 3. Model selection of simulated SPT data. (a) Diffusion constants are $D_1 = 10\ \mu m^2/s$ and $D_2 = 1\ \mu m^2/s$. (b) Diffusion constants are $D_1 = 3\ \mu m^2/s$ and $D_2 = 1\ \mu m^2/s$. The models correspond to the stHMM, HMM, and one-state diffusion methods, respectively. Their posterior model probabilities, each with equal prior probability, come from the analysis of heterogenous diffusion trajectories with state switching, motion blur, and localization errors (known to be $\sigma_E = 0.010\ \mu m$). To calculate the model probability, we need the evidence of each model [see Eq. (18)]. These will in turn be based on the respective likelihood functions [i.e., Eq. (2) and Eqs. (B1)–(B3)]. The model probabilities are calculated for seven data sets of distinct mean number of state switches from states 2 to 1 (1–2) ν_1 (ν_2) increasing from left to right. The analyzed data $\mathbf{O}$ consist of 10 independent trajectories with 100 observed displacements each and a sampling time of $\tau = 0.005$ s. Each data point is the mean value of 100 independent runs on data sets that share the same ground truth parameters. Priors used are lognormal with parameters detailed in Table I. We choose the model with the highest probability (P_{chosen}), and the error in this model choice is $1 - P_{\text{chosen}}$.

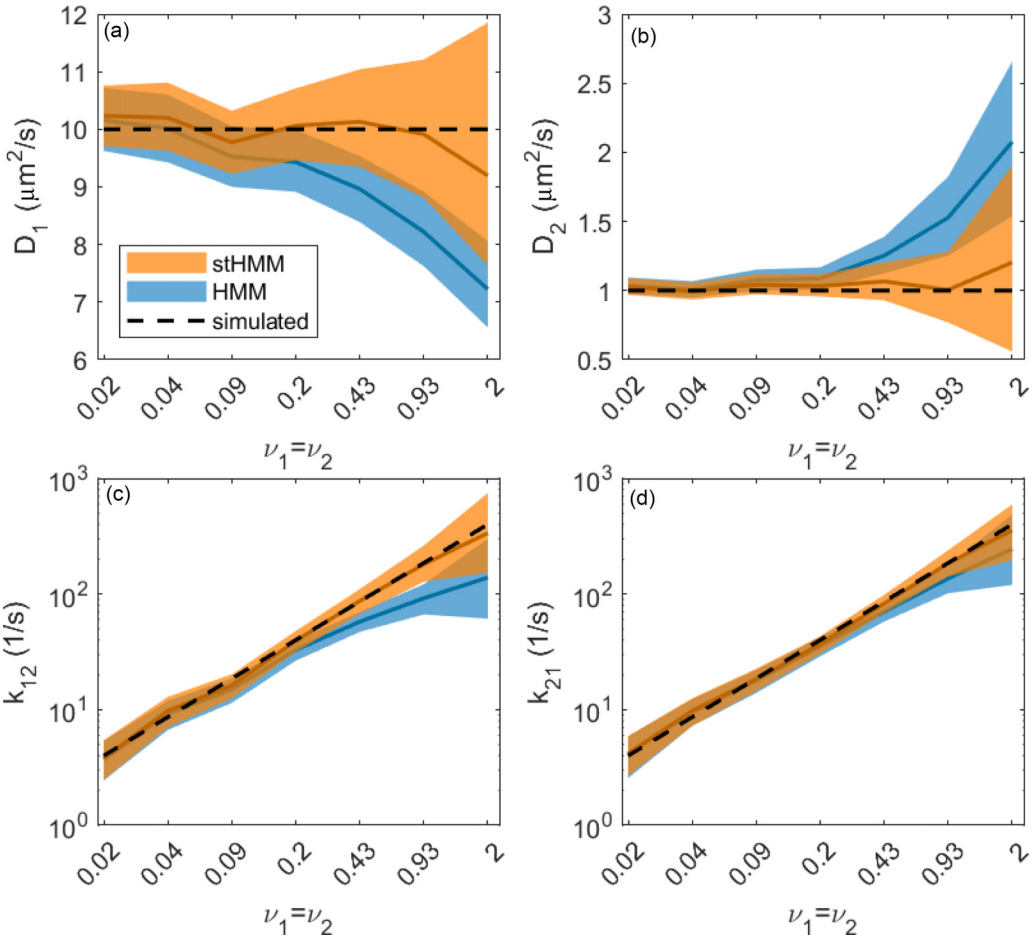


FIG. 4. Parameter inference of simulated SPT data. (a)–(d) Inference results of the unknown parameters, which are diffusion constants D_1, D_2 and kinetic rates k_{12}, k_{21}. They are estimated for seven data sets of distinct mean number of state switches from states 2 to 1 (1–2) ν_1 (ν_2) increasing from left to right in each subplot. Solid lines and shaded regions mark Bayesian estimates and 95% credibility levels for the two color-coded analysis methods employed. Ground truth values are marked by black dotted lines. The analyzed data **O** consist of 10 independent trajectories with 300 observed displacements each and a sampling time of $\tau = 0.005$ s. The localization error standard deviation is $\sigma_E = 0.010$ μm. Priors used are lognormal and detailed in Table I.

C. Parameter inference

Having established that stHMM provides means for exact computation of the evidence for two-state diffusive dynamics (and hence capabilities for doing model selection), we go on to study its performance with respect to parameter estimation. We seek to estimate model parameters from SPT data in $d = 2$ dimensions that include motion blur and localization errors (Sec. III A).

Figure 4 shows inference results for the stHMM and HMM methods, as a function of the mean number of state switches per time interval ν. Beginning with the slow kinetics regime ($\nu < 0.1$), we see that HMM and stHMM yield accurate and similar results. The reason is that very few observations are badly resolved (i.e., include a state transition), making HMM a limiting case of stHMM. In the intermediate regime ($0.1 < \nu < 0.4$), HMM's inference of diffusion constants in Figs. 4(a) and 4(b), begins to deviate noticeably from those of stHMM. In particular, HMM systematically underestimates D_1 and overestimates D_2. This is because it tries to match transition-associated observations to a model with fixed diffusion constants. stHMM successfully separates them, by carefully accounting for all possible subsampling time events weighted by their respective probability. In the rapid switching regime ($\nu > 0.4$), many more observations carry subtime events. HMM's systematic errors in estimating the diffusion constants then accentuate. stHMM's uncertainty estimates for diffusion constants increase with rapid dynamics, since unobserved subtime events increase in frequency. Moreover, stHMM "keeps up" with the kinetic rates, which are clearly too high to handle for HMM [notice the logarithmic scale on the vertical axis in Figs. 4(c) and 4(d)].

Improved error estimates also deserve some attention. We see in Fig. 4 that they become larger for rapid kinetics (especially for stHMM). That is expected since kinetic information is difficult to extract with low temporal resolution. In the extreme limit, all sampling intervals get a w_n value close to the stationary probability of the molecule being in state 1, i.e., the distribution approaches a delta function as its width diminishes (see Appendix E). Then the emission PDF becomes similar for all time intervals, thus resembling the emission PDF of a one-state model with a weighted diffusion constant.

Typical run times of stHMM for a dataset consisting of 10 trajectories of 300 measurements each are around 7 min on a laptop with a 12th-generation Intel i5 processor.

III. METHODS

A. Generating synthetic data

Let us describe the physical system considered and introduce some notation. We assume the state-switching behavior of every biomolecule to be governed by continuous-time Markov dynamics. Hence, they switch from state 1 to 2 with rate k_{12} and back from state 2 to 1 with rate k_{21}. Note that these transitions, in contrast to standard HMM, can happen at any time. When in state i ($i = 1, 2$), a biomolecule undergoes diffusion with a diffusion constant D_i. What we actually measure, Δ_n, are observed displacements. The data are collected during time intervals of duration τ; multiple observations form a trajectory $O = \{\Delta_1, \Delta_2, \ldots, \Delta_N\}$ in time.

To generate synthetic SPT trajectories, we proceed as follows. We divide each time interval into M small subsampling time intervals such that the subtime intervals are chosen to be much shorter than the inverse of any rate constant for the process. Computationally, we initialize a "state matrix" whose rows indicate the state (1 or 2) in sampling time intervals and whose columns indicate the states of subtime intervals. For example, the second row, first column gives the state at the start of the second sampling time interval. The state sequences are generated via Eq. (D10) with time set to $\delta\tau = \tau/M$. For each subtime interval, we draw a zero-mean Gaussian random number with variance $2D_i\delta\tau$. Finally, we cumulatively sum the displacements over all subtime intervals, leaving only position values at the sampling times.

In addition, SPT data suffer from two major error-inducing effects (see Fig. 1). The first is motion blur, which is added by setting the final sampling time positions equal to the average of their subsampling time equivalents (assuming an open shutter mode) [10]. The second are localization errors, modeled as time-independent Gaussian random numbers with zero mean and variance σ_E^2. We incorporate this noise by adding it to the final positions, after which one-dimensional displacements are computed [10]. All steps up to this point are done independently for each spatial dimension. Finally, we compute d-dimensional "observed" displacements (in Sec. II, we set $d = 2$).

B. The stHMM analysis pipeline

1. Analysis overview

Let us now introduce the physical model used in analyzing the data. Since the underlying biomolecule state is hidden to us at any given time (the red-blue trajectories in Figs. 1 and 2), we must condition each observation (the gray trajectories with plus signs in Figs. 1 and 2) on the current state s_n and marginalize over all such possible states $s_1, \ldots, s_N$. Just as with the standard HMM [13,18], we consider start and end states of each time interval. But in contrast to the standard HMM, we let the molecule switch state at *any time during data acquisition*. Such subsampling time events are parametrized by weights w_n for the fraction of time spent in state 1 during time interval n (through w_n, we thus group all subtime contributions by their occupation times in the two states). Subsampling time events lead to contributions from both states to the "emission distribution," which determines what we actually measure (see Fig. 5). In mathematical form, we have

$$\begin{aligned} &P(\Delta_1, \Delta_2, \ldots, \Delta_N|\theta) \\ &= \sum_{s_0=1}^{2}\sum_{s_1=1}^{2}\cdots\sum_{s_N=1}^{2}\int_0^1 \mathrm{d}w_1 \int_0^1 \mathrm{d}w_2 \cdots \int_0^1 \mathrm{d}w_N \\ &\quad P(\Delta_1|w_1,\theta)P(s_0)P(w_1,s_1|s_0,\theta) \\ &\quad P(\Delta_2|w_2,\theta)P(w_2,s_2|s_1,\theta) \\ &\quad \cdots P(\Delta_N|w_N,\theta)P(w_N,s_N|s_{N-1},\theta), \end{aligned} \tag{2}$$

where $P(w_n, s_n|s_{n-1}, \theta)$ is the the transition-accretion probabilities, as before. The quantities $P(\Delta_n|w_n, \theta)$ are the emission probability density functions. We here assume that the joint emission probability density factorizes, i.e., there is no correlation between observed displacements. Given this assumption, Eq. (2) gives the exact likelihood for specific parameter values θ. It is known as the likelihood function when used within a Bayesian setting. If the system at hand has had time to reach a stationary state, the initial probabilities $P(s_0)$ are the stationary probabilities

$$P(s_0 = 1) = \frac{k_{21}}{k_{12} + k_{21}}, \tag{3a}$$

$$P(s_0 = 2) = \frac{k_{12}}{k_{12} + k_{21}}. \tag{3b}$$

Equation (2) contains a central theoretical quantity: the conditional probability $P(w_n, s_n|s_{n-1}, \theta)$. We call these generalized transition probabilities *transition-accretion probabilities*, and now discuss their meaning. As time progresses within a sampling time interval, the fraction of time spent in state 1 "accretes" into w_n. The transition-accretion probabilities contain all possible transitions that may happen below the time resolution limit. Our elaborate analytical derivation of $P(w_n, s_n|s_{n-1}, \theta)$ in Appendix C underpins the entire stHMM analysis pipeline. Final forms are provided in Eqs. (9) and (10).

Given our likelihood function [Eq. (2)], we use Bayesian inference to estimate the unknown parameters θ from the observations $O = \{\Delta_1, \ldots, \Delta_N\}$ (in general, from a set of trajectories $\mathbf{O} = O^{(1)}, \ldots, O^{(M)}$, where $O^{(k)}$ is trajectory number k). For our type of data, the parameters are diffusion constants and kinetic rates: $\theta = \{D_1, D_2, k_{12}, k_{21}\}$. For the parameter estimation, we apply the Bayes formula [21]

$$P(\theta|\mathbf{O}) = \frac{P(\mathbf{O}|\theta)P(\theta)}{Z}, \tag{4}$$

which relates the likelihood of observations $P(\mathbf{O}|\theta)$ to the sought posterior probability of parameter values $P(\theta|\mathbf{O})$. Because our initial guess for parameter values, $P(\theta)$, is a distribution, so is $P(\theta|\mathbf{O})$. The evidence $Z = P(\mathbf{O})$ is just a normalization constant in the context of parameter estimation, but is central for doing model selection (see Sec. III C 3). Sampling the multivariate posterior distribution is nontrivial; we use the nested sampling algorithm outlined in Sec. III C 2, which also outputs Z. The main challenge is to evaluate

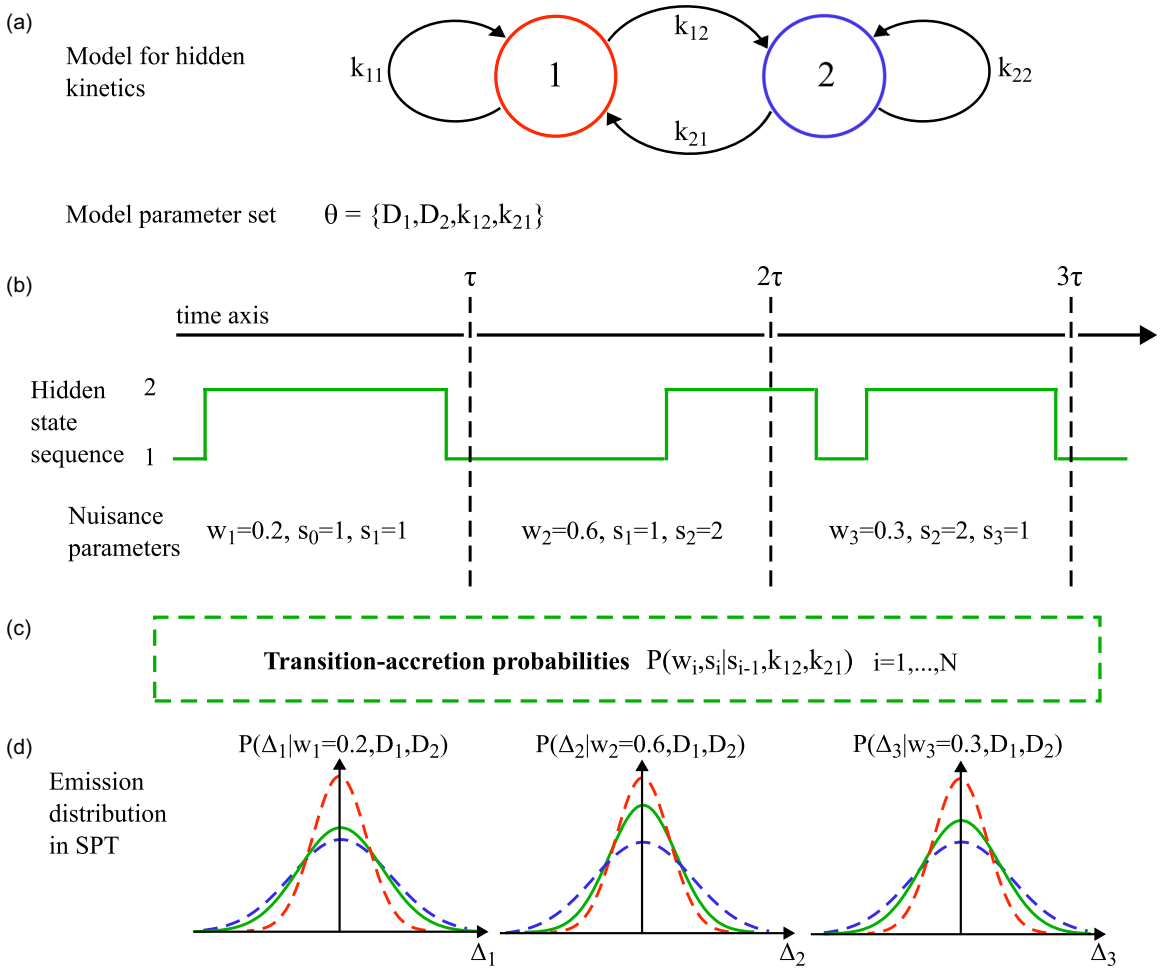


FIG. 5. The stHMM analysis pipeline. (a) The physical system studied exhibits hidden continuous-time Markovian two-state kinetics with state switches possible at any time, parametrized by switching rates k_{12}, k_{21}. The parameter set θ consists of these, together with two more parameters that characterize the emission distribution, which for SPT are diffusion constants D_1, D_2. (b) Observations Δ_i are recorded with imaging time τ (continuous-time data averaged over sampling-time windows), during which the system randomly switches state in a hidden state sequence. Example observations can be seen in Fig. 2. The fraction of time spent in state 1 accretes into w_i for time interval i, and determines the probability distribution of Δ_i for given θ [see Eq. (7) and panel (d)]. (c) The w_i contribution is accurately described by our transition-accretion probabilities, which include dependency also on the hidden end states s_i and the switching rates. Still, w_i and s_i are hidden "nuisance parameters"—which we marginalize over in our inference procedure. This can be seen in the likelihood function [Eq. (2)], where both transition-accretion probabilities and emission probabilities enter. (d) SPT emission distributions for each imaging time (solid green) plotted with those of state 1 (dashed red) and state 2 (dashed blue) for reference. The SPT situation considered herein occurs whenever the variances of the single-state Gaussian PDFs are different but their means are equal (and = 0).

Eq. (2) efficiently enough that sampling of the posterior distribution via Eq. (4) is possible (see Sec. III C 1).

2. *Emission distribution*

We now turn to study the (conditional) emission PDF $P(\Delta_n|w_n,\theta)$ of Eq. (2), which determines the probabilities of observations. The conditioning on w_n means that we include emissions from all states visited during the sampling time. This is depicted in Fig. 5(d).

Assuming normal diffusion in d dimensions and ignoring measurement noise (see the below paragraphs), the emission PDF for given weight w_n is

$$P(\Delta_n|w_n,\theta) = \frac{1}{(2\pi\sigma^2(w_n))^{d/2}} e^{-\frac{\Delta_n^2}{2\sigma^2(w_n)}}, \tag{5}$$

which follows from considering the PDF as a sum of random state-dependent subsampling time emissions. Here, Δ_n^2 is the sum of squared displacements in the d dimensions. The variance is a linear combination of the individual variances of each state:

$$\sigma^2(w_n) = w_n\sigma_1^2 + (1-w_n)\sigma_2^2 \tag{6}$$

$$= 2\tau(w_nD_1 + (1-w_n)D_2). \tag{7}$$

In order to make Eq. (5) more useful for SPT, we incorporate two sources of "noise": motion blur and localization error, as done previously [11,12]. Briefly, motion blur makes fluorescently labeled molecules ("dots") appear smeared-out due to a biomolecule's motion during camera exposure time. Localization errors, on the other hand, arise in dot detection, making every recorded position differ slightly from the true

position. For motion blur, we assume an open shutter mode so that the motion blur coefficient is $R = 1/6$, and use the correction formula $\sigma^2(w_n) \to (1-2R)\sigma^2(w_n)$ in similarity with Refs. [10,11]. For localization errors, we follow Slator *et al.* [12] and assume that the variance is small relative to the Brownian motion: $\sigma_E^2 \ll \min(\sigma_1^2, \sigma_2^2)$, such that we may use their modification $\sigma^2(w_n) \to \sigma^2(w_n) + 2\sigma_E^2$. In actuality, motion blur and localization errors both induce correlations among nearby observed displacements [11], which we neglect throughout this paper. In summary, the noise-corrected emission PDF is given by Eq. (5) with the substitution

$$\sigma^2(w_n) \to (1-2R)\sigma^2(w_n) + 2\sigma_E^2. \tag{8}$$

σ_E^2 will in a real experiment depend on the photobleaching rate, point spread function, pixel size, algorithm for dot detection and localization, exposure time, etc.

3. Transition-accretion probabilities

Having seen how the emissions are affected by subsampling time events, the next step is to describe the probabilities of those events. The transition-accretion probability densities appearing in Eq. (2) are derived in Appendix C using techniques from mathematical physics. In our derivation, we consider a single sampling time interval n with fixed end states, and start by calculating the moment generating function of time spent in state 1 with these conditions. By transforming also with respect to the sampling time duration τ, we reach a closed-form expression for this double transform, which we then inverse-transform back to the original variables. Upon changing the time variable to time fraction w_n, we find

$$P(w_n, s_n = 1|s_{n-1} = 1, \theta) = Q_\delta^{(1)} + Q_s^{(1)}, \tag{9a}$$

$$P(w_n, s_n = 2|s_{n-1} = 2, \theta) = Q_\delta^{(2)} + Q_s^{(2)}, \tag{9b}$$

$$P(w_n, s_n = 2|s_{n-1} = 1, \theta) = Q_o^{(1)}, \tag{9c}$$

$$P(w_n, s_n = 1|s_{n-1} = 2, \theta) = Q_o^{(2)}. \tag{9d}$$

Here, we introduced the quantities

$$Q_\delta^{(1)} = e^{-\tau k_{12}}\delta(1-w_n), \tag{10a}$$

$$Q_\delta^{(2)} = e^{-\tau k_{21}}\delta(w_n), \tag{10b}$$

$$Q_o^{(1)} = \tau k_{12} e^{-\tau(k_{12}w_n + k_{21}(1-w_n))} I_0(2\tau\sqrt{k_{12}k_{21}w_n(1-w_n)}), \tag{10c}$$

$$Q_o^{(2)} = \tau k_{21} e^{-\tau(k_{12}w_n + k_{21}(1-w_n))} I_0(2\tau\sqrt{k_{12}k_{21}w_n(1-w_n)}), \tag{10d}$$

$$Q_s^{(1)} = \tau e^{-\tau(k_{12}w_n + k_{21}(1-w_n))}\sqrt{\frac{k_{12}k_{21}w_n}{1-w_n}} I_1(2\tau\sqrt{k_{12}k_{21}w_n(1-w_n)}), \tag{10e}$$

$$Q_s^{(2)} = \tau e^{-\tau(k_{12}w_n + k_{21}(1-w_n))}\sqrt{\frac{k_{12}k_{21}(1-w_n)}{w_n}} I_1(2\tau\sqrt{k_{12}k_{21}w_n(1-w_n)}), \tag{10f}$$

where $\delta(w)$ is a delta function and I_0, I_1 are modified Bessel functions of the first kind [22] of order 0 and 1, respectively.

Let us briefly interpret these terms. $Q_\delta^{(i)}$ is the probability to stay in state i throughout time interval n, given start state i ($i = 1, 2$). Thus, $Q_\delta^{(1)}$ ($Q_\delta^{(2)}$) picks up w_n value 1 (0). $Q_s^{(i)}$ also corresponds to starting and ending in state i, but includes an even number of state switches in-between. It has a w_n dependence, which allows for values $w_n \in (0, 1)$. Finally, $Q_o^{(i)}$ is the PDF that, given start state i, end up in state $j \neq i$ at the end of the time interval. Since it ends up in the opposite state, it includes an odd number of state switches and is defined for $w_n \in (0, 1)$.

Two limits are of interest here. In the slow switching limit, $Q_s^{(i)}$ becomes negligible, and $Q_o^{(1)}$ ($Q_o^{(2)}$) simplifies to τk_{12} (τk_{21}) as shown in Appendix F. In the fast switching limit, $Q_\delta^{(i)}$ instead becomes negligible, while $Q_s^{(i)}$ and $Q_o^{(i)}$ approach Gaussians with w_n centered at the stationary probability of Eq. (3a) (see Appendix E).

C. Parameter inference

Having an expression for the likelihood in Eq. (2), the final challenge is to numerically sum over the hidden state sequence $(s_0, s_1, \ldots, s_N)$ of start/end states in each time interval. We detail our method for evaluating these sums in Sec. III C 1 and for sampling the posterior distribution of parameter values in Secs. III C 2–III C 4.

1. Forward algorithm

We need an efficient way of computing Eq. (2) for parameter estimation purposes. That trajectory likelihood includes 2^N terms—an intractable number even for small N. To manage this, we update the standard HMM forward algorithm to incorporate the weights w_n [see Eq. (13)]. The likelihood is written as

$$P(O|\theta) = \sum_{s_N=1}^{2} \alpha_N(s_N), \tag{11}$$

where α_N is calculated recursively from

$$\alpha_n(s_n) = \sum_{s_{n-1}=1}^{2} \alpha_{n-1}(s_{n-1}) P(\Delta_n, s_n|s_{n-1}, \theta), \tag{12}$$

with $s_n = 1, 2$ fixed, $n = 2, 3, \ldots, N$ and where $\alpha_n(s_n)$ is the forward variable, defined as the probability of all observed displacements up to step n and ending in state s_n, given specific values of the model parameters θ. Also, we defined the *weighted joint probabilities* as

$$P(\Delta_n, s_n|s_{n-1}, \theta) = \int_0^1 P(\Delta_n|w_n, \theta) P(w_n, s_n|s_{n-1}, \theta)\, \mathrm{d}w_n. \tag{13}$$

Explicit expressions for these are given by Eq. (A1). Note that the weighted joint probabilities involve both emission and transition-accretion probabilities, and that the hidden parameter w_n is integrated out. We employ Matlab's formula-based

ALGORITHM 1. stHMM forward algorithm for calculating the loglikelihood of a trajectory of N measurements. The weighted joint probabilities are obtained by numerical integration, see Eq. (13).

$\theta \leftarrow \{D_1, D_2, k_{12}, k_{21}\}$
$\text{logsum}(x, y) = \begin{cases} x + \log(1 + e^{y-x}) & \text{if } x \leqslant y \\ y + \log(1 + e^{x-y}) & \text{otherwise} \end{cases}$
$\log[\alpha_1(s_1 = i)] \leftarrow \text{logsum}[\log(P(s_0 = i)) + \log(P(\Delta_1, s_1 = i|s_0 = i, \theta)), \log(P(s_0 = j)) + \log(P(\Delta_1, s_1 = i|s_0 = j, \theta))];\ i, j = 1, 2, i \neq j$
for $n = 2$ to $n = N$ **do**
  $\log[\alpha_n(s_n = i)] \leftarrow \text{logsum}[\log(\alpha_{n-1}(s_{n-1} = i)) + \log(P(\Delta_n, s_n = i|s_{n-1} = i, \theta)), \log(\alpha_{n-1}(s_{n-1} = j)) + \log(P(\Delta_n, s_n = i|s_{n-1} = j, \theta))];\ i, j = 1, 2, i \neq j$
end for
$\log[P(O|\theta)] \leftarrow \text{logsum}(\log[\alpha_N(s_N = 1)], \log[\alpha_N(s_N = 2)])$

integral method for all integral evaluations. For $n = 1$,

$$\alpha_1(s_1) = \sum_{s_0=1}^{2} P(s_0)P(\Delta_1, s_1|s_0, \theta), \tag{14}$$

where initial probabilities $P(s_0)$ are set according to the system at hand. For instance, if the system has had time to equilibrate they are given by Eq. (3), and if the system is prepared to be in, e.g., state $s_0 = 1$ at the initial time we set $P(s_0 = 1) = 1, P(s_0 = 2) = 0$.

The full stHMM forward algorithm is presented in Algorithm 1. We use logarithms to avoid numerical underflow.

The total likelihood of an ensemble of M independent trajectories is obtained simply by multiplying together the likelihood of each one:

$$P(\mathbf{O}|\theta) = \prod_{k=1}^{M} P(O^{(k)}|\theta). \tag{15}$$

This product thus becomes a sum when considering log-likelihoods.

2. Bayesian inference with the nested sampling algorithm

In order to evaluate the posterior probability given in Eq. (4) with a likelihood given in Eq. (2) [evaluated using Eqs. (11)–(14)], we need an algorithm to compute the posterior and evidence. To that end, we use nested sampling [23], which is a Monte Carlo algorithm with focus on computing the evidence Z. Tackling this multivariate integral directly involves heavy computation due to poor dimensional scaling, known as "the curse of dimensionality" [24]. Yet, nested sampling overcomes this difficulty by grouping sums by likelihood [23]:

$$Z \simeq \sum_i L_i \omega_i, \tag{16}$$

where i is an iteration index, L_i is the likelihood sampled at iteration i, and ω_i is a corresponding weight. Moreover, the posterior average of an arbitrary function $f(\theta)$ is [25]

$$\langle f(\theta) \rangle \simeq \sum_i \frac{L_i \omega_i}{Z} f(\theta), \tag{17}$$

where the posterior weights $L_i\omega_i/Z$ are calculable once the evidence has been estimated. In particular, Eq. (17) lets us compute Bayesian parameter estimates (posterior means) by setting $f(\theta) = \theta$.

To find the upper limit of the credibility interval of a parameter, we sum up the marginalized posterior weights from the smallest value and stop when reaching the 97.5% level (i.e., we prune away the top 2.5%). The lower limit is found analogically by working from the largest sampled value and proceeding to smaller ones (i.e., pruning away the bottom 2.5%).

As stopping criterion, we set the fractional increase in estimated evidence to 10^{-4}. The number of live points we set to 200, which makes for a smooth posterior distribution without being very computationally intensive [26]. Using nested sampling in this context was done in Ref. [27], and can be regarded as an improvement to earlier employed Markov chain Monte Carlo methods [12,13] which come with a number of tunable hyper-parameters, and which do not provide an estimate of the evidence (Sec. III C 3).

When drawing random numbers from a likelihood-constrained prior we use a bounding ellipsoid [26] in a probability-conserving transformed space. In this space, all live points' positions are uniformly distributed, such that a simple method [28] of drawing uniform spherical random numbers and mapping them to the ellipsoid can be used. The ellipsoid is constructed based on the covariance matrix among the live points' positions and scaled by a factor of 1.06^2 [26] to encompass the true likelihood contour. Thanks to the small dimensionality of only four parameters, the ellipsoidal method achieves a high acceptance ratio when using at least 100 live points [24].

3. Model selection

In reality, we may not know precisely what model to use, e.g., what type of diffusion is being studied or whether to include various noise corrections or not. This problem of model selection is within the Bayesian framework dealt with by comparing the posterior probabilities $P(M_i|\mathbf{O})$ of the competing models M_i [21]. Assuming that these have equal prior probabilities, the posterior probability of M_k given observations $\mathbf{O}$ is

$$P(M_k|\mathbf{O}) = \frac{Z(M_k)}{\sum_i Z(M_i)} \tag{18}$$

due to Bayes formula (4) when written in terms of models rather than parameters. We here made the conditioning on the model explicit; i.e., above $Z(M_k)$ is the normalizing constant Z obtained from Eq. (4) for model M_k. One then picks that model, k^*, which has the highest evidence (support) in the analyzed data (model selection). The model probability is $P_{\text{chosen}} = P(M_{k^*}|\mathbf{O})$, and the probability that we chose the wrong model is estimated as $1 - P_{\text{chosen}}$. The nested sampling algorithm employed here directly outputs the evidence, making model selection easy.

4. Selecting priors

Because our analysis is fully Bayesian, we need to specify the prior distribution $P(\theta)$ in Eq. (4). We use independent

logarithmic Gaussian (lognormal) priors for all four parameters, with parameter values listed in Table I. The prior reflects our complete knowledge before seeing the data; we here assume only an order-of-magnitude idea of ground truth values.

IV. DISCUSSION

In this paper, we have studied the parameter inference and model selection problems for two-state diffusion where state switches, for instance, due to binding events, can occur at any time. The problems share a natural use of the HMM framework, which has become widespread in related fields. Yet, it is only through this study that the HMM has been adapted to the problem at hand: allowing for state transitions to happen at any time during data acquisition (continuous-time data averaged over sampling-time windows). By doing so, our stHMM approach retains the advantages of the battle-tested HMM while extending its validity and inference precision. The power of the stHMM approach originates from the concept of transition-accretion probabilities, paired with an updated forward algorithm for efficient likelihood evaluation.

Perhaps the most striking conclusion is how our stHMM pipeline provides accurate parameter inference across all kinetic switching regimes, including its reliable error estimates. The advantages of this are twofold. First, an experimentalist need not worry whether the inference is trustworthy or not. This is helpful as, in practice, the relevant regimes are seldom known *a priori*. Second, stHMM provides superior inference for rapid kinetics. The results validate not only stHMM's ability to handle subsampling time state switches but also its ability to handle motion blur and localization errors through the approximation in Eq. (8).

We began our analysis by comparing the evidence of three candidate models. The basic one-state diffusion model we suggest always be included, since data of very poor temporal resolution cannot be readily analyzed with a two-state model. In such a case, the one-state model is favored due to it being less complex, i.e., having only one parameter. In contrast to other methods [14,29] in the SPT field, we do model selection via direct evidence calculation and leverage the built-in model complexity penalization of the Bayesian framework (the Occam's razor [20] effect). Thereby, we have for the first time exactly compared the evidence Z computed for the models behind stHMM, HMM, and one-state diffusion.

From the experimental point of view, our stHMM analysis pipeline provides a solution to the experimentalist's photon budget dilemma. By using our method and software, the experimentalist can choose a sufficiently large imaging time to be able to collect enough photons to detect the biomolecules of interest spatially, while relying on our method to still achieve accurate estimation of parameters related to rate processes occurring on timescales faster than the imaging time.

From a broader perspective, there is an interest in HMMs in fields that deal with continuous-time data averaged over sampling-time windows. A continuous-time HMM, however, assumes the observations to be dependent only on the instantaneous state at the time of emission (the "snapshot property") [30]. That "deficiency" requires an extra assumption of slow state switching, similar to the usual HMM. In light of this, the development of exact methods is described as a promising area of research by Glennie *et al.* [30]. We argue that stHMM marks a major step in that direction.

For the future, it would be natural to generalize our method to deal with systems of more than two states. One would then need to derive the transition-accretion probability densities of a K-state Markov chain, analogous to Appendix C. Marginalization would be over the resulting $K-1$ unknown weight factors being the fraction of time spent in states $1,\ldots,K-1$. Such an analysis can be combined with evidence computation, so as to estimate K through model selection should it be unknown. Due to the extra numerical integrations, such an analysis scheme would be more computationally demanding than our two-state version.

Another future study would be to estimate state sequences with the stHMM. For each time interval, both an end state s_n and a weight w_n with associated credibility intervals would be estimated. In other words, one could construct a forward-backward algorithm [18], just as we in this paper have constructed a forward algorithm.

We expect stHMM to be of use, in its current or modified form, in other fields that employ HMMs for continuous-time data averaged over a set of sampling-time windows, e.g., in bioacoustic sound analysis [31,32], animal movement analysis [30], financial market analysis [33], and earth quake prediction [34].

To summarize, our stHMM framework is the first that provide reliable means for computationally fast estimation of transition rate constants and diffusion constants for any SPT-like data. The stHMM methodology and public software provided in this study widen the applicability of the HMM framework to intermediate and rapid kinetics. Thereby, temporal resolution no longer puts a hard constraint on the feasibility of data analysis.

ACKNOWLEDGMENTS

E.C. is grateful for discussions with Michael Lomholt. T.A. acknowledges research Grant No. 2022-03475 from the Swedish Research Council.

DATA AVAILABILITY

The data that support the findings of this article are openly available [17].

APPENDIX A: FORMULAE FOR COMPILED QUANTITIES USED IN THE stHMM FORWARD ALGORITHM

Substituting Eqs. (5), (9), and (10) into Eq. (13) gives us the form of the four *weighted joint probabilities* $P(\Delta_n, s_n|s_{n-1}, \theta)$, used in our forward algorithm. They are

$$P(\Delta_n, s_n = 1|s_{n-1} = 1, \theta) = e^{-\tau k_{12}} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{\Delta_n^2}{2\sigma_1^2}} + \tau \int_0^1 \frac{1}{\sqrt{2\pi\sigma(w_n)^2}} e^{-\frac{\Delta_n^2}{2\sigma(w_n)^2}} e^{-\tau(k_{12}w_n + k_{21}(1-w_n))} \sqrt{\frac{k_{12}k_{21}w_n}{1-w_n}} I_1(2\tau\sqrt{k_{12}k_{21}w_n(1-w_n)})\,\mathrm{d}w_n, \tag{A1a}$$

$$P(\Delta_n, s_n = 2|s_{n-1} = 2, \theta) = e^{-\tau k_{21}} \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{\Delta_n^2}{2\sigma_2^2}} + \tau \int_0^1 \frac{1}{\sqrt{2\pi\sigma(w_n)^2}} e^{-\frac{\Delta_n^2}{2\sigma(w_n)^2}} e^{-\tau(k_{12}w_n + k_{21}(1-w_n))} \sqrt{\frac{k_{12}k_{21}(1-w_n)}{w_n}} I_1(2\tau\sqrt{k_{12}k_{21}w_n(1-w_n)})\,\mathrm{d}w_n, \tag{A1b}$$

$$P(\Delta_n, s_n = 2|s_{n-1} = 1, \theta) = \tau k_{12} \int_0^1 \frac{1}{\sqrt{2\pi\sigma(w_n)^2}} e^{-\frac{\Delta_n^2}{2\sigma(w_n)^2}} e^{-\tau(k_{12}w_n + k_{21}(1-w_n))} I_0(2\tau\sqrt{k_{12}k_{21}w_n(1-w_n)})\,\mathrm{d}w_n, \tag{A1c}$$

$$P(\Delta_n, s_n = 1|s_{n-1} = 2, \theta) = \tau k_{21} \int_0^1 \frac{1}{\sqrt{2\pi\sigma(w_n)^2}} e^{-\frac{\Delta_n^2}{2\sigma(w_n)^2}} e^{-\tau(k_{12}w_n + k_{21}(1-w_n))} I_0(2\tau\sqrt{k_{12}k_{21}w_n(1-w_n)})\,\mathrm{d}w_n. \tag{A1d}$$

A number of computationally heavy integrals need to be evaluated, though Eqs. (A1c) and (A1d) are equal up to a constant prefactor and need thus only be evaluated once. Note that the fundamental kinetic rates k_{12}, k_{21}, whose inverses give the mean dwell times, above appear as model parameters. In HMM, these parameters can only be included in an approximate way (see Appendix D).

Equations (A1a) and (A1b) display integrable singularities for $w_n = 1$ and $w_n = 0$, respectively. For accurate numerical integration, we rewrite $I_1(z)$ using recurrence relation

$$I_\alpha(z) = \frac{z(I_{\alpha-1}(z) - I_{\alpha+1}(z))}{2\alpha}, \tag{A2}$$

with $\alpha = 1$, so that in Eq. (A1a)

$$\sqrt{\frac{k_{12}k_{21}w_n}{1-w_n}} I_1(2\tau\sqrt{k_{12}k_{21}w_n(1-w_n)}) = \tau k_{12}k_{21}w_n[I_0(2\tau\sqrt{k_{12}k_{21}w_n(1-w_n)}) - I_2(2\tau\sqrt{k_{12}k_{21}w_n(1-w_n)})] \tag{A3}$$

and in Eq. (A1b)

$$\sqrt{\frac{k_{12}k_{21}(1-w_n)}{w_n}} I_1(2\tau\sqrt{k_{12}k_{21}w_n(1-w_n)}) = \tau k_{12}k_{21}(1-w_n)[I_0(2\tau\sqrt{k_{12}k_{21}w_n(1-w_n)}) - I_2(2\tau\sqrt{k_{12}k_{21}w_n(1-w_n)})]. \tag{A4}$$

Both of these equations are free from apparent divergences, thus being straightforward to integrate numerically.

APPENDIX B: HMM AND ONE-STATE DIFFUSION ANALYSES

1. HMM

Here, we give a brief account of how we implement the standard HMM method. The HMM likelihood given in Ref. [13], with modifications for motion blur and localization errors as in Refs. [11,12], is evaluated via the forward algorithm [13,18]. We use Eq. (D10) to estimate kinetic rates from the discrete-time transition probabilities. The likelihood function of a trajectory is

$$P(\Delta_1, \Delta_2, \ldots, \Delta_N|\theta) = \sum_{s_1=1}^{2}\sum_{s_2=1}^{2}\cdots\sum_{s_N=1}^{2} P(\Delta_1|s_1,\theta)P(s_1|\theta)P(\Delta_2|s_2,\theta)P(s_2|s_1,\theta) \cdots P(\Delta_N|s_N,\theta)P(s_N|s_{N-1},\theta), \tag{B1}$$

with state-dependent emission PDF

$$P(\Delta_n|s_n,\theta) = \frac{\exp\left\{-\frac{\Delta_n^2}{2((1-2R)\sigma^2(s_n)+2\sigma_E^2)}\right\}}{\left[2\pi\left((1-2R)\sigma^2(s_n)+2\sigma_E^2\right)\right]^{d/2}}, \tag{B2}$$

with variance $\sigma^2(s_n) = 2\tau D(s_n)$, i.e., $D(s_n = 1) = D_1$ and $D(s_n = 2) = D_2$.

2. One-state diffusion

The one-state diffusion model is simply a product of Gaussian emission PDFs:

$$P(\Delta_1, \Delta_2, \ldots, \Delta_N|D) = \prod_{n=1}^{N} P(\Delta_n|D), \tag{B3}$$

with each PDF similar to Eq. (B2) but with $\sigma^2(s_n) = \sigma^2 = 2\tau D$. The only parameter here is the diffusion constant D.

APPENDIX C: DERIVATION OF TRANSITION-ACCRETION PROBABILITIES

In this Appendix, we derive the occupation time distribution in a continuous-time two-state Markov chain with fixed end states—the transition-accretion probabilities. We start with a discrete Markov model, which we later turn into a continuous-time form by letting the step size $\delta\tau$ approach zero. Our strategy is to compute the moment generating function (MGF) as a device for later computing the PDF. In the last part of the derivation we adapt a calculation similar to the one by Pedler [35].

1. Transition matrix

The transition matrix in a two-state setting is

$$\mathbf{P}(\tau) = \begin{pmatrix} p_{1,1}(\tau) & p_{1,2}(\tau) \\ p_{2,1}(\tau) & p_{2,2}(\tau) \end{pmatrix} = \begin{pmatrix} 1 - p_{1,2}(\tau) & p_{1,2}(\tau) \\ p_{2,1}(\tau) & 1 - p_{2,1}(\tau) \end{pmatrix}, \tag{C1}$$

where $p_{u,v}(\tau) = \mathrm{Prob}(v|u)$ $(u, v = 1, 2)$, i.e., $p_{u,v}$ is the probability that, given that we start in state u, find the system in state v after a time τ. For short times $\delta\tau$, we can write

$$\mathbf{P}(\delta\tau) = \mathbf{I} + \delta\tau\mathbf{R} + \mathcal{O}((\delta\tau)^2), \tag{C2}$$

where $\mathcal{O}(\ldots)$ denotes the characteristic scaling of remainder terms, and

$$\mathbf{R} = \begin{pmatrix} -k_{12} & k_{12} \\ k_{21} & -k_{21} \end{pmatrix} \tag{C3}$$

is the rate matrix.

Suppose now that we divide time into M subintervals:

$$M\delta\tau = \tau, \tag{C4}$$

and let $M \longrightarrow \infty$, $\delta\tau \longrightarrow 0$ such that τ remains constant. Moreover, in a discrete calculation, time is advanced by multiplying by transition (transfer) matrices. Thereby,

$$\begin{aligned} \mathbf{P}(\tau) &= [\mathbf{P}(\delta\tau)]^M \\ &= [\mathbf{I} + \delta\tau\mathbf{R} + \mathcal{O}(\delta\tau^2)]^M \\ &= \left[\mathbf{I} + \frac{\tau\mathbf{R}}{M} + \mathcal{O}(\delta\tau^2)\right]^M \\ &\overset{M\longrightarrow\infty}{\longrightarrow} \exp(\tau\mathbf{R}), \end{aligned} \tag{C5}$$

which is the continuous-time transition matrix. This expression will be of use in the next subsection.

2. Moment generating function of time spent in state 1

Consider a single displacement, say number n, in a longer trajectory of displacements. We fix the biomolecule's end states s_{n-1} and s_n during the corresponding sampling time interval. In discrete time, we thus have a sequence $(s_{n-1}, \eta_1^{(n)}, \eta_2^{(n)}, \ldots, \eta_M^{(n)}, s_n)$ of subsampling time states. We want to sum out the M state variables $\eta_1^{(n)}, \ldots, \eta_M^{(n)}$; for ease of notation, we drop the superscripts below. By its Markov property, the joint conditional probability is

$$\begin{aligned} &P(\eta_1, \ldots, \eta_M, s_n|s_{n-1}) \\ &\quad = p_{s_{n-1},\eta_1} p_{\eta_1,\eta_2} p_{\eta_2,\eta_3} \cdots p_{\eta_{M-1},\eta_M} p_{\eta_M,s_n}, \end{aligned} \tag{C6}$$

where every state is either 1 or 2, i.e., $s_i, \eta_i = 1, 2$. We now seek the probability, P_k, of the biomolecule being k times $(0 \leqslant k \leqslant M)$ in state 1 and ending in state s_n, given that it started in state s_{n-1}. Following Ref. [36], we define "occupation vectors" $\mathbf{e}(u)$, which have all elements set to 0 except for element u which is 1. That is,

$$\mathbf{e}(1) = (1, 0)^{\mathrm{T}}, \tag{C7}$$

$$\mathbf{e}(2) = (0, 1)^{\mathrm{T}}. \tag{C8}$$

A given (subsampling time) state sequence can then be represented in terms of these occupation vectors as

$$\{\eta_1, \eta_2, \ldots, \eta_M\} \equiv \{\mathbf{e}(\eta_1), \mathbf{e}(\eta_2), \ldots, \mathbf{e}(\eta_M)\}. \tag{C9}$$

Moreover, we can collect all times spent in state 1—a random variable—by scalar multiplication, i.e.,

$$\begin{aligned} n^{(1)} &= \text{Number of times we are in state 1} \\ &= \sum_{i=1}^{M} \mathbf{e}(\eta_i) \cdot \mathbf{e}(1). \end{aligned} \tag{C10}$$

We find the distribution of $n^{(1)}$, with fixed end states, by using Kronecker (or Dirac) delta functions:

$$\begin{aligned} &P(k, s_n|s_{n-1}, M) \\ &= \langle \delta_{k,n^{(1)}} \rangle \\ &= \sum_{\eta_1=1}^{2} \cdots \sum_{\eta_M=1}^{2} \delta_{k,n^{(1)}} \\ &\quad p_{s_{n-1},\eta_1} p_{\eta_1,\eta_2} \cdots p_{\eta_{M-1},\eta_M} p_{\eta_M,s_n}, \end{aligned} \tag{C11}$$

where $\langle \cdots \rangle$ denotes an expectation value. Summing over k and s_n gives unity, as required by normalization.

In our search for the continuum limit $(M \longrightarrow \infty)$ of the above expression, we will calculate the MGF with respect to the number of state-1 occupations k (the discrete analog of occupation time), take the continuum limit, and later inverse transform. Thus, the discrete transform of Eq. (C11) is

$$\begin{aligned} &\psi^{\mathrm{disc}}(\tilde{\zeta}; s_n, s_{n-1}, M) \\ &\quad = \sum_{k=0}^{M} \exp(-\tilde{\zeta}k) P(k, s_n|s_{n-1}, M), \end{aligned} \tag{C12}$$

where $\tilde{\zeta}$ is the transformed variable corresponding to k. We write out the full expression by plugging in Eq. (C11):

$$\begin{aligned} &\psi^{\mathrm{disc}}(\tilde{\zeta}; s_n, s_{n-1}, M) \\ &= \sum_{k=0}^{M} \exp(-\tilde{\zeta}k) \Bigg(\sum_{\eta_1=1}^{2} \cdots \sum_{\eta_M=1}^{2} \\ &\quad \times \delta_{k,n^{(1)}} p_{s_{n-1},\eta_1} p_{\eta_1,\eta_2} \cdots p_{\eta_{M-1},\eta_M} p_{\eta_M,s_n} \Bigg). \end{aligned} \tag{C13}$$

Using Eq. (C10) and the Dirac-delta property, we now get

$$\psi^{\rm disc}(\tilde{\zeta}; s_n, s_{n-1}, M) = \sum_{\eta_1=1}^{2} \cdots \sum_{\eta_M=1}^{2} p_{s_{n-1},\eta_1} \exp[-\tilde{\zeta}\mathbf{e}(\eta_1)\cdot\mathbf{e}(1)] \cdots p_{\eta_1,\eta_2} \exp[-\tilde{\zeta}\mathbf{e}(\eta_2)\cdot\mathbf{e}(1)] \cdots p_{\eta_{M-1},\eta_M} \exp[-\tilde{\zeta}\mathbf{e}(\eta_M)\cdot\mathbf{e}(1)] p_{\eta_M,s_n}, \quad \text{(C14)}$$

which can be conveniently formulated as M matrix multiplications. Therefore, we define a matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \text{(C15)}$$

which accounts only for time spent in state 1. Equation (C14) can now be written as

$$\psi^{\rm disc}(\tilde{\zeta}; s_n, s_{n-1}, M) = \mathbf{e}^{\rm T}(s_{n-1})[\mathbf{T}(\tilde{\zeta})]^M \mathbf{P}\mathbf{e}(s_n), \quad \text{(C16)}$$

where

$$\mathbf{T}(\tilde{\zeta}) = \mathbf{P}\exp(-\tilde{\zeta}\mathbf{A}), \quad \text{(C17)}$$

where $\mathbf{P}$ is the transition matrix. The matrix multiplications in Eq. (C16) successively sum state-1 occupations in the transformed space, one step at a time along the Markov chain.

In preparation for taking the continuum limit, we consider again Eq. (C12). By defining a Laplace variable (with dimension inverse time)

$$\zeta = \frac{\tilde{\zeta}}{\delta\tau}, \quad \text{(C18)}$$

we find from Eq. (C12)

$$\begin{aligned}\psi^{\rm disc}(\tilde{\zeta}; s_n, s_{n-1}, M) &= \sum_{k=0}^{M} \exp(-\tilde{\zeta}k) P(k, s_n|s_{n-1}, M) \\ &= \sum_{k\delta\tau=0}^{M\delta\tau} \exp\left(-\frac{\tilde{\zeta}}{\delta\tau}\delta\tau k\right) P(k\delta\tau, s_n|s_{n-1}, M\delta\tau) \\ &= \sum_{t=0}^{\tau} \exp(-\zeta t) P(t, s_n|s_{n-1}, \tau) \\ &= \psi^{\rm disc}(\zeta; s_n, s_{n-1}, \tau), \end{aligned} \quad \text{(C19)}$$

where we made use of the fact that the probability to be in state 1 k times is the same as spending time

$$t = k\delta\tau, \quad \text{(C20)}$$

in state 1 if every interval lasts $\delta\tau$ time units. Thus, $t \leqslant \tau$ as it should, but note that time is so far still a discrete variable (albeit with noninteger values in general). With these variables, Eq. (C16) reads

$$\psi^{\rm disc}(\zeta; s_n, s_{n-1}, \tau) = \mathbf{e}^{\rm T}(s_{n-1})[\mathbf{T}(\zeta)]^M \mathbf{P}(\delta\tau)\mathbf{e}(s_n), \quad \text{(C21)}$$

where [updating Eq. (C17)]

$$\mathbf{T}(\zeta) = \mathbf{P}(\delta\tau)\exp(-\zeta\delta\tau\mathbf{A}), \quad \text{(C22)}$$

with transition matrix $\mathbf{P}(\delta\tau)$ now given by Eq. (C2). The matrix multiplications in Eq. (C21) have a continuous-time limit:

$$\begin{aligned}[\mathbf{T}(\zeta)]^M &= [\mathbf{P}(\delta\tau)\exp(-\zeta\delta\tau\mathbf{A})]^M \\ &= [(\mathbf{I} + \delta\tau\mathbf{R} + \cdots)\exp(-\zeta\delta\tau\mathbf{A})]^M \\ &= [\exp(\delta\tau\mathbf{R})\exp(-\zeta\delta\tau\mathbf{A})]^M \\ &= \left[\exp\left(\frac{\tau\mathbf{R}}{M}\right)\exp\left(-\frac{\zeta\tau\mathbf{A}}{M}\right)\right]^M \\ &\xrightarrow{M\to\infty} \exp(\tau(\mathbf{R} - \zeta\mathbf{A})), \end{aligned} \quad \text{(C23)}$$

where the second line follows from Eq. (C2), the third from Eq. (C5), the fourth from Eq. (C4), and the last from the Lie product formula

$$\lim_{M\to\infty}\left[\exp\left(\frac{\mathbf{B}}{M}\right)\exp\left(\frac{\mathbf{C}}{M}\right)\right]^M = \exp(\mathbf{B} + \mathbf{C}) \quad \text{(C24)}$$

valid for arbitrary (noncommuting) matrices [37]. The continuous-time version of Eq. (C19) now becomes

$$\begin{aligned}&\psi(\zeta; s_n, s_{n-1}, \tau) \\ &= \int_0^{\tau} \exp(-\zeta t) P(t, s_n|s_{n-1}, \tau)\, {\rm d}t \\ &= \int_0^{\infty} \exp(-\zeta t) P(t, s_n|s_{n-1}, \tau)\Theta(\tau - t)\, {\rm d}t, \end{aligned} \quad \text{(C25)}$$

with $\Theta(t)$ being the Heaviside step function that lets us regard this as a Laplace transform. The final expression for this Laplace transform is [see Eqs. (C21) and (C23)]

$$\psi(\zeta; s_n, s_{n-1}, \tau) = \mathbf{e}^{\rm T}(s_{n-1}) e^{\tau(\mathbf{R}-\zeta\mathbf{A})}\mathbf{e}(s_n), \quad \text{(C26)}$$

where we used the fact that the final transition probability matrix in Eq. (C21) yields no contribution: $\mathbf{P}(\delta\tau) \longrightarrow \mathbf{I}$ as $\delta\tau \longrightarrow 0$, as seen from Eq. (C2). Equation (C26) is the continuous-time MGF up until sampling time τ and is a generalization of the results in Ref. [38] for fixed initial and final states over a sampling time interval. The inverse-transform of this MGF is $P(t, s_n|s_{n-1}, \tau)$, which is normalized [cf. Eq. (C11)], i.e., we have $\sum_{s_n}\int P(t, s_n|s_{n-1}, \tau)\, {\rm d}t = 1$.

3. Transition-accretion probability densities

Next, we seek to Laplace transform with respect to the sampling time interval, τ, to easier inverse transform back to the original variable space (i.e., time t spent in state 1 during sampling time τ). Proceeding as in Ref. [35], albeit for general left and right state vectors, we have from Eq. (C26)

$$\Psi(\zeta, \phi; s_n, s_{n-1}) = \mathbf{e}^{\rm T}(s_{n-1})\mathcal{L}\{e^{\tau(\mathbf{R}-\zeta\mathbf{A})}\}(\zeta, \phi)\mathbf{e}(s_n), \quad \text{(C27)}$$

where ϕ is the Laplace variable associated with τ. The above is a standard transform (albeit with matrix argument) and yields

$$\Psi(\zeta, \phi; s_n, s_{n-1}) = \mathbf{e}^{\rm T}(s_{n-1})(-\mathbf{R} + \zeta\mathbf{A} + \phi\mathbf{I})^{-1}\mathbf{e}(s_n), \quad \text{(C28)}$$

with $\mathbf{I}$ being the identity matrix. Matrix inversion of the 2×2 matrix is elementary. To proceed, we need to pick end states s_{n-1}, s_n (four possible cases).

Choosing $s_{n-1} = s_n = 1$, we get

$$\Psi(\zeta, \phi; s_n = 1, s_{n-1} = 1) = \frac{k_{21} + \phi}{(k_{12} + \zeta + \phi)(k_{21} + \phi) - k_{12}k_{21}}. \tag{C29}$$

Let us first consider the simpler transformation

$$\tilde{\Psi}(q, r; s_n = 1, s_{n-1} = 1) = \frac{r}{qr - k_{12}k_{21}} = \frac{1}{q} + \frac{k_{12}k_{21}}{q(qr - k_{12}k_{21})}, \tag{C30}$$

where the variables are independent so that the Laplace inverse will factorize later. The first term is just the double transform of a delta function $\delta(y)$, where (x, y) are the normal variables that correspond to the transformed (q, r). Let us find the inverse transform of the second term. Using the geometric series

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}, \tag{C31}$$

valid for $|x| < 1$, we rewrite the second term (without the $k_{12}k_{21}$ factor) in Eq. (C30) as a power series:

$$\frac{1}{q(qr - k_{12}k_{21})} = \frac{1}{q^2 r}\left(\frac{1}{1 - k_{12}k_{21}/(qr)}\right) = \frac{1}{q^2 r}\sum_{k=0}^{\infty}\left(\frac{k_{12}k_{21}}{qr}\right)^k = \sum_{k=0}^{\infty}\frac{(k_{12}k_{21})^k}{q^{k+2}r^{k+1}}. \tag{C32}$$

We see that Eq. (C32) is the double Laplace transform

$$\mathcal{L}^2\left\{\sum_{k=0}^{\infty}(k_{12}k_{21})^k \frac{x^{k+1}}{(k+1)!}\frac{y^k}{k!}\right\}(q, r) \tag{C33}$$

$$= \sum_{k=0}^{\infty}\mathcal{L}^2\left\{(k_{12}k_{21})^k \frac{x^{k+1}}{(k+1)!}\frac{y^k}{k!}\right\}(q, r), \tag{C34}$$

since

$$\mathcal{L}^2\left\{\frac{x^{k+1}}{(k+1)!}\frac{y^k}{k!}\right\}(q, r) = \int_0^{\infty} \mathrm{d}x\, e^{-xq}\frac{x^{k+1}}{(k+1)!}\int_0^{\infty}\mathrm{d}y\, e^{-yr}\frac{y^k}{k!} = \mathcal{L}\left\{\frac{x^{k+1}}{(k+1)!}\right\}(q)\mathcal{L}\left\{\frac{y^k}{k!}\right\}(r) = \frac{1}{q^{k+2}}\frac{1}{r^{k+1}}. \tag{C35}$$

The series in Eq. (C33) we identify as a modified Bessel function of order 1:

$$\sum_{k=0}^{\infty}(k_{12}k_{21})^k \frac{x^{k+1}}{(k+1)!}\frac{y^k}{k!} = \sqrt{\frac{x}{k_{12}k_{21}y}}\sum_{k=0}^{\infty}\frac{(k_{12}k_{21}xy)^{k+1/2}}{k!(k+1)!} = \sqrt{\frac{x}{k_{12}k_{21}y}}I_1(2\sqrt{k_{12}k_{21}xy}), \tag{C36}$$

from its power series representation (order m) [23]

$$I_m(z) = \left(\frac{z}{2}\right)^m \sum_{k=0}^{\infty}\frac{1}{k!(k+m)!}\left(\frac{z}{2}\right)^{2k}. \tag{C37}$$

Therefore, we see that

$$\frac{1}{q(qr - k_{12}k_{21})} = \mathcal{L}^2\left\{\sqrt{\frac{x}{k_{12}k_{21}y}}I_1(2\sqrt{k_{12}k_{21}xy})\right\}(q, r). \tag{C38}$$

We now seek the transform $(t, \tau) \longrightarrow (\zeta, \phi)$ in terms of the simpler $(x, y) \longrightarrow (q, r)$ transform above. They are related in Laplace space through

$$q = \zeta + \phi + k_{12}, \tag{C39a}$$

$$r = \phi + k_{21}. \tag{C39b}$$

Laplace shifting by a constant, such as k_{12}, k_{21}, follows the relation

$$\mathcal{L}^2\{e^{-k_{12}\zeta - k_{21}\phi}f(t, \tau)\}(\zeta, \phi) = \mathcal{L}^2\{f(t, \tau)\}(\zeta + k_{12}, \phi + k_{21}). \tag{C40}$$

Moreover, we need to consider the Laplace shift of an independent variable:

$$\begin{aligned}\mathcal{L}^2\{f(t, \tau - t)\Theta(\tau - t)\}(\zeta, \phi) &= \int_0^{\infty}\mathrm{d}t\, e^{-t\zeta}\int_0^{\infty}\mathrm{d}\tau\, e^{-\tau\phi}f(t, \tau - t)\Theta(\tau - t)\\ &= \int_0^{\infty}\mathrm{d}t\, e^{-t\zeta}\int_{-t}^{\infty}\mathrm{d}\tau\, e^{-(u+t)\phi}f(t, u)\Theta(u)\\ &= \int_0^{\infty}\mathrm{d}t\, e^{-t(\zeta+\phi)}\int_0^{\infty}\mathrm{d}\tau\, e^{-t\phi}f(t, u)\\ &= \mathcal{L}^2\{f(t, \tau)\}(\zeta + \phi, \phi),\end{aligned} \tag{C41}$$

where $\Theta(u)$ is the Heaviside step function. Combining Eqs. (C38), (C40), and (C41), we find the inverse of Eq. (C29):

$$P(t, s_n = 1|s_{n-1} = 1, \tau) = e^{-(k_{12}t + k_{21}(\tau - t))}\left[\delta(\tau - t) + \sqrt{\frac{k_{12}k_{21}t}{\tau - t}}I_1(2\sqrt{k_{12}k_{21}t(\tau - t)})\right], \tag{C42}$$

which is the probability to spend time t in state 1 during sampling time τ and end in state 1, given the starting state 1. Here, we have omitted $\Theta(\tau - t)$ since it is clear from the physical situation that $t \leqslant \tau$. With a change of variables from $t \longrightarrow w$, where

$$t = w\tau, \tag{C43}$$

the Jacobian is $\mathrm{d}t/\mathrm{d}w = \tau$ and we retrieve our transition-accretion probability of Eqs. (9a), (10a), and (10c).

By symmetry (or by repeating the calculation), we find the corresponding probability with end states

equal to 2:

$$P(t, s_n = 2|s_{n-1} = 2, \tau) = e^{-(k_{12}t + k_{21}(\tau - t))}\left[\delta(t) + \sqrt{\frac{k_{12}k_{21}(\tau - t)}{t}} \times I_1(2\sqrt{k_{12}k_{21}t(\tau - t)})\right]. \quad \text{(C44)}$$

Let us finally find the last two transition-accretion probabilities. Going back to Eq. (C28) and picking $s_{n-1} = 1,\ s_n = 2$,

$$\Psi(\zeta, \phi; s_n = 2, s_{n-1} = 1) = \frac{k_{12}}{(k_{12} + \zeta + \phi)(k_{21} + \phi) - k_{12}k_{21}}. \quad \text{(C45)}$$

Using again the (q, r) variables Eq. (C39), we find

$$\tilde{\Psi}(q, r; s_n = 2, s_{n-1} = 1) = \frac{k_{12}}{qr - k_{12}k_{21}}. \quad \text{(C46)}$$

With the Laplace transform

$$\frac{1}{qr - k_{12}k_{21}} = \mathcal{L}^2\{I_0(2\sqrt{k_{12}k_{21}xy})\}(q, r), \quad \text{(C47)}$$

we get

$$P(x, s_n = 2|s_{n-1} = 1, y) = k_{12}I_0(2\sqrt{k_{12}k_{21}xy}). \quad \text{(C48)}$$

Using the shift properties of Eqs. (C40) and (C41), we find the inverse transform

$$P(t, s_n = 2|s_{n-1} = 1, \tau) = k_{12}e^{-(k_{12}t + k_{21}(\tau - t))}I_0(2\sqrt{k_{12}k_{21}t(\tau - t)}), \quad \text{(C49)}$$

which after the variable change of Eq. (C43) is identical to Eqs. (9c) and (10c). Similarly, we find

$$P(t, s_n = 1|s_{n-1} = 2, \tau) = k_{21}e^{-(k_{12}t + k_{21}(\tau - t))}I_0(2\sqrt{k_{12}k_{21}t(\tau - t)}). \quad \text{(C50)}$$

APPENDIX D: RELATION TO DISCRETE TRANSITION PROBABILITIES

To connect with the approach by Das *et al.* [13], we here provide an alternative derivation of Eq. (D10). It can be viewed as moving from a continuous-time description to a discrete-time one (then assuming slow switching). This is to say we seek the probability of switching (or staying put) in one state from time 0 to another state at time τ, irrespective of any switches happening during that time interval.

To that end, we want to marginalize the time spent in state 1, t, in Eq. (C26). This can be done indirectly by noting that the defining relation of a Laplace transform

$$\mathcal{L}\{f(t)\}(s) = \int_0^\infty e^{-st} f(t)\, \mathrm{d}t \quad \text{(D1)}$$

implies that the zero-frequency component is

$$\mathcal{L}\{f(t)\}(0) = \int_0^\infty f(t)\, \mathrm{d}t, \quad \text{(D2)}$$

which is precisely the integral we wish to evaluate. As such, we set $\zeta = 0$ in Eq. (C26):

$$\psi(0; s_n, s_{n-1}, \tau) = \mathbf{e}^{\mathrm{T}}(s_{n-1})e^{\tau\mathbf{R}}\mathbf{e}(s_n). \quad \text{(D3)}$$

The end vectors $\mathbf{e}$ are not relevant for the remainder of this section (they serve to "pick out" the end states of interest). Already here, we see agreement with Eq. (C5), meaning that we are about to compute the transition matrix $\mathbf{P}(\tau)$.

Now the challenge is to calculate the matrix exponential, where the rate matrix $\mathbf{R}$ is given by Eq. (C3). The first step is to diagonalize

$$\mathbf{R} = \mathbf{T}\mathbf{D}\mathbf{T}^{-1}, \quad \text{(D4)}$$

where $\mathbf{T}$ contains the eigenvectors of $\mathbf{R}$ and $\mathbf{D}$ its corresponding eigenvalues. Then, by the definition of the exponential function as a power series,

$$e^{\tau\mathbf{R}} = \mathbf{T}e^{\tau\mathbf{D}}\mathbf{T}^{-1}. \quad \text{(D5)}$$

We find

$$\mathbf{D} = \begin{pmatrix} 0 & 0 \\ 0 & -(k_{12} + k_{21}) \end{pmatrix}, \quad \text{(D6)}$$

$$\mathbf{T} = \begin{pmatrix} 1 & -k_{12}/k_{21} \\ 1 & 1, \end{pmatrix} \quad \text{(D7)}$$

and

$$\mathbf{T}^{-1} = \frac{1}{1 + k_{12}/k_{21}}\begin{pmatrix} 1 & k_{12}/k_{21} \\ -1 & 1 \end{pmatrix}. \quad \text{(D8)}$$

Thereby,

$$e^{\tau\mathbf{R}} = \frac{1}{k_{12} + k_{21}}\begin{pmatrix} k_{21} + k_{12}e^{-(k_{12}+k_{21})\tau} & k_{12}(1 - e^{-(k_{12}+k_{21})\tau}) \\ k_{21}(1 - e^{-(k_{12}+k_{21})\tau}) & k_{12} + k_{21}e^{-(k_{12}+k_{21})\tau} \end{pmatrix}. \quad \text{(D9)}$$

Comparing with Eqs. (C1) and (C5), we identify the discrete transition probabilities

$$p_{1,2}(\tau) = \frac{k_{12}}{k_{12} + k_{21}}(1 - e^{-(k_{12}+k_{21})\tau}), \quad \text{(D10a)}$$

$$p_{2,1}(\tau) = \frac{k_{21}}{k_{12} + k_{21}}(1 - e^{-(k_{12}+k_{21})\tau}), \quad \text{(D10b)}$$

as given in the supplementary of Das *et al.* [13]. Hence, $p_{1,2}(\tau)$ is the probability to start in state 1 at time $t = 0$ and end in state 2 at time $t = \tau$, with any number of transitions in-between. While these formulas are exact, they do not change the fact that the HMM pipeline ignores subsampling time state transitions (it assumes a single state for each sampling time interval). To yield correct estimates of kinetic rates and diffusion constants, the hidden states should rather be marginalized together with the emissions, as we do in our stHMM framework [see Eq. (13)].

APPENDIX E: FAST-SWITCHING ANALYSIS OF stHMM

We here seek to analyze the transition-accretion probabilities, given in Eqs. (9) and (10), in the limit where

$$k_{12}\tau \longrightarrow \infty, \quad \text{(E1a)}$$

$$k_{21}\tau \longrightarrow \infty. \quad \text{(E1b)}$$

This limit would arise for fixed kinetic rates and an infinitely long imaging time. As a first observation, we see that

$$Q_\delta^{(i)} \longrightarrow 0, \quad i = 1, 2, \quad \text{(E2)}$$

which means that the probability to stay in the starting state throughout the time interval is zero in this limit. Moving on to the remaining probability densities, we note that they all have the factor

$$U(w_n) = \tau e^{-\tau(k_{12}w_n + k_{21}(1-w_n))} I_\alpha(2\tau\sqrt{k_{12}k_{21}w_n(1-w_n)}) \tag{E3}$$

in common (where $\alpha = 0, 1$). We know the asymptotic result [23]

$$I_\alpha(z) \sim \frac{1}{\sqrt{2\pi z}} e^z, \tag{E4}$$

for $|z| \longrightarrow \infty$ which is valid for $|\arg z| < \pi/2$ irrespective of α. Inserting Eq. (E4) into Eq. (E3),

$$U(w_n) \sim \left(\frac{\tau}{4\pi\sqrt{k_{12}k_{21}w_n(1-w_n)}}\right)^{1/2} e^{-\tau(\sqrt{k_{12}w_n}+\sqrt{k_{21}(1-w_n)})^2}. \tag{E5}$$

The maximum occurs where the argument w_n^* of the exponential vanishes:

$$\sqrt{k_{12}w_n^*} = \sqrt{k_{21}(1-w_n^*)}, \tag{E6}$$

equivalent to

$$w_n^* = \frac{k_{21}}{k_{12}+k_{21}}. \tag{E7}$$

We recognize this from Eq. (3a)—it is the stationary probability of residing in state 1.

From here we take one step further and look for a Gaussian approximation of Eq. (E5), which can be seen as part of large deviation theory (see, e.g., Ref. [39]). To that end, we write

$$w_n = w_n^* + \Delta w_n, \tag{E8}$$

where $\Delta w_n = w_n - w_n^*$. With this, we investigate the argument of the exponential, denoted by I, near $w_n = w_n^*$:

$$I = \sqrt{k_{12}w_n} - \sqrt{k_{21}(1-w_n)} \tag{E9}$$

$$= \sqrt{k_{12}w_n^*}\sqrt{1+\frac{\Delta w_n}{w_n^*}} + \sqrt{k_{21}(1-w_n^*)}\sqrt{1-\frac{\Delta w_n}{1-w_n^*}}, \tag{E10}$$

and by employing the small-argument Taylor expansion

$$\sqrt{1\pm x} \simeq 1 \pm \tfrac{1}{2}x \tag{E11}$$

for $x = \Delta w_n/w_n^*$ and $x = \Delta w_n/(1-w_n^*)$, we get

$$I = \frac{\Delta w_n}{2}\left(\sqrt{\frac{k_{12}}{w_n^*}} + \sqrt{\frac{k_{21}}{1-w_n^*}}\right) = \frac{\Delta w_n}{2}\left(\sqrt{\frac{k_{12}+k_{21}}{w_n^*(1-w_n^*)}}\right), \tag{E12}$$

where we made use of Eq. (E6). Now, we are ready to find the Gaussian approximation. Inserting Eqs. (E7) and (E12) into Eq. (E5) and setting $w_n = w_n^*$ in the prefactor,

$$U(w_n) \sim \left(\frac{\tau}{4\pi k_{12}k_{21}w_n^*(1-w_n^*)}\right)^{1/2} e^{-\frac{(\Delta w_n)^2(k_{12}+k_{21})\tau}{4w_n^*(1-w_n^*)}} \tag{E13}$$

and if we define

$$\sigma^2 = \frac{2w_n^*(1-w_n^*)}{(k_{12}+k_{21})\tau}, \tag{E14}$$

we have

$$U(w_n) \sim \frac{1}{k_{12}+k_{21}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(w_n-w_n^*)^2}{2\sigma^2}}, \tag{E15}$$

where we used also Eq. (E8). The two transition-accretion probability densities for starting and ending in the same state, have extra prefactors where we set $w_n = w_n^*$ and use Eq. (E7) to get

$$\sqrt{\frac{k_{12}k_{21}w_n^*}{1-w_n^*}} = k_{21}, \tag{E16}$$

$$\sqrt{\frac{k_{12}k_{21}(1-w_n^*)}{w_n^*}} = k_{12}. \tag{E17}$$

Given these and Eqs. (E3), (E7), and (E15), we find the following asymptotic Gaussian approximations of the transition-accretion PDFs:

$$P(w_n, s_n = 1|s_{n-1} = 1, \theta) \sim w_n^* \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(w_n-w_n^*)^2}{2\sigma^2}}, \tag{E18a}$$

$$P(w_n, s_n = 2|s_{n-1} = 2, \theta) \sim (1-w_n^*) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(w_n-w_n^*)^2}{2\sigma^2}}, \tag{E18b}$$

$$P(w_n, s_n = 2|s_{n-1} = 1, \theta) \sim (1-w_n^*) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(w_n-w_n^*)^2}{2\sigma^2}}, \tag{E18c}$$

$$P(w_n, s_n = 1|s_{n-1} = 2, \theta) \sim w_n^* \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(w_n-w_n^*)^2}{2\sigma^2}}. \tag{E18d}$$

Just as one would expect, it does not matter in which state the biomolecule starts, but only in which state it ends. From Eq. (E14), we see that all transition-accretion PDFs become sharply peaked at w_n^* and approach δ-functions as $\tau \longrightarrow \infty$. Normalization is then asymptotically fulfilled: Adding Eqs. (E18a) and (E18c) and integrating over w_n gives probability 1 in the limit [and likewise for Eqs. (E18b) and (E18d)].

APPENDIX F: SLOW-SWITCHING ANALYSIS OF stHMM

In this Appendix, we simplify Eq. (A1) in the limit

$$\nu_i \ll 1, \quad i = 1, 2, \tag{F1}$$

with ν_i defined in Eq. (1). The simplified form is then compared to HMM, providing insight into the approximations made with HMM.

Only for relatively few time intervals will a transition event (or more) occur. To simplify the modified Bessel functions in this regime, we employ the Taylor series expansion in Eq. (C37) so that

$$I_0(z) = 1 + \mathcal{O}(z^2) \tag{F2}$$

and

$$I_1(z) = \frac{z}{2} + \mathcal{O}(z^3). \tag{F3}$$

Furthermore, we use the Taylor expansion of an exponential function

$$\exp(-z) = 1 - z + \mathcal{O}(z^2). \tag{F4}$$

With this at hand, we simplify Eq. (A1) to first order in ν_i:

$$P(\Delta_n, s_n = 1 | s_{n-1} = 1, \theta) \simeq (1 - \tau k_{12}) \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{\Delta_n^2}{2\sigma_1^2}}, \tag{F5a}$$

$$P(\Delta_n, s_n = 2 | s_{n-1} = 2, \theta) \simeq (1 - \tau k_{21}) \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{\Delta_n^2}{2\sigma_2^2}}, \tag{F5b}$$

$$P(\Delta_n, s_n = 2 | s_{n-1} = 1, \theta) \simeq \tau k_{12} \int_0^1 \frac{1}{\sqrt{2\pi\sigma^2(w_n)}} e^{-\frac{\Delta_n^2}{2\sigma^2(w_n)}} \mathrm{d}w_n, \tag{F5c}$$

$$P(\Delta_n, s_n = 1 | s_{n-1} = 2, \theta) \simeq \tau k_{21} \int_0^1 \frac{1}{\sqrt{2\pi\sigma^2(w_n)}} e^{-\frac{\Delta_n^2}{2\sigma^2(w_n)}} \mathrm{d}w_n. \tag{F5d}$$

Of these, Eqs. (F5a) and (F5b) include no state transition and Eqs. (F5c) and (F5d) each include a single transition. We identify the approximate transition probabilities $p_{1,2} = \tau k_{12}$, $p_{2,1} = \tau k_{21}$, and the others follow from normalization. The fact that I_1 terms are of order $\geqslant 2$ in ν_i and the I_0 terms of order $\geqslant 1$ can be explained from physical grounds. Because of the slow kinetics, it is much more probable to switch once than twice. It is even more probable to stay put.

The HMM is closely related to Eq. (F5), but it does not use the correct state-conditional emission distributions in Eqs. (F5c) and (F5d); i.e., the emissions are treated as unaffected by the state switching. To be correct, we see how these emission PDFs should be integrated over all possible weights w_n, reflecting how that single transition is equally likely to happen at any time during data acquisition [see Fig. 2(d)].

[1] J. Elf and I. Barkefors, Single-molecule kinetics in living cells, Annu. Rev. Biochem. **88**, 635 (2019).

[2] P. K. Koo and S. G. J. Mochrie, Systems-level approach to uncovering diffusive states and their transitions from single-particle trajectories, Phys. Rev. E **94**, 052412 (2016).

[3] H. Milller, Z. Zhou, J. Shepherd, A. J. M. Wollman, and M. C. Leake, Single-molecule techniques in biophysics: A review of the progress in methods and applications, Rep. Prog. Phys. **81**, 024601 (2018).

[4] T. Dam, M. Chouliara, V. Junghans, and P. Jönsson, Supported lipid bilayers and the study of two-dimensional binding kinetics, Front. Mol. Biosci. **9**, 833123 (2022).

[5] C. Veigel and C. F. Schmidt, Moving into the cell: Single-molecule studies of molecular motors in complex environments, Nat. Rev. Mol. Cell Biol. **12**, 163 (2011).

[6] J. Renaud, C. Chung, U. Danielson, U. Egner, M. Hennig, R. Hubbard, and H. Nar, Biophysics in drug discovery: Impact, challenges and opportunities, Nat. Rev. Drug Discovery **15**, 679 (2016).

[7] R. Lino, K. Kinbara, and Z. Bryant, Introduction: Molecular motors, Chem. Rev. **120**, 1 (2020).

[8] Y. Miao, Y. Cheng, Y. Xia, Y. Hei, W. Wang, Q. Dai, J. Suo, and C. Chen, Supervised multi-frame dual-channel denoising enables long-term single-molecule FRET under extremely low photon budget, Nat. Commun. **16**, 74 (2025).

[9] M. Fazel and M. J. Wester, Analysis of super-resolution single molecule localization microscopy data: A tutorial, AIP Adv. **12**, 010701 (2022).

[10] A. J. Berglund, Statistics of camera-based single-particle tracking, Phys. Rev. E **82**, 011917 (2010).

[11] M. Lindén, V. Ćurić, E. Amselem, and J. Elf, Pointwise error estimates in localization microscopy, Nat. Commun. **8**, 15115 (2017).

[12] P. J. Slator, C. W. Cairo, and N. J. Burroughs, Detection of diffusion heterogeneity in single particle tracking trajectories using a hidden Markov model with measurement noise propagation, PLoS One **10**, e0140759 (2015).

[13] R. Das, C. W. Cairo, and D. Coombs, A hidden Markov model for single particle tracks quantifies dynamic interactions between LFA-1 and the actin cytoskeleton, PLoS Comput. Biol. **5**, e1000556 (2009).

[14] F. Persson, M. Lindén, C. Unoson, and J. Elf, Extracting intracellular diffusive states and transition rates from single-molecule tracking data, Nat Methods **10**, 265 (2013).

[15] A. Federbush, A. Moscovich, and Y. Bar-Sinai, Hidden Markov modeling of single-particle diffusion with stochastic tethering, Phys. Rev. E **109**, 034129 (2024).

[16] F. Simon, J.-Y. Tinevez, and S. van Teeffelen, Extrack characterizes transition kinetics and diffusion in noisy single-particle tracks, J. Cell Biol. **222**, e202208059 (2023).

[17] E. Clarkson and T. Ambjörnsson, Software for subsampling time hidden Markov model (stHMM) analysis of single molecule trajectories, GitHub, 2025, https://github.com/ewclarkson/stHMM-single-molecule/.

[18] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, Cambridge, 2007), 3rd ed.

[19] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis* (CRC Press, New York, 2013), 3rd ed.

[20] D. S. Sivia and J. Skilling, *Data Analysis A Bayesian Tutorial*, 2nd ed. (Oxford University Press, 2006).

[21] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, Cambridge, 2005).

[22] M. Abramowitz and I. A. Segun, *Handbook of Mathematical Functions* (Dover Publications, New York, 1964), 1st ed.

[23] J. Skilling, Nested sampling for general Bayesian computation, Bayesian Anal. **833**, 1 (2006).

[24] J. Buchner, Nested sampling methods, Stat. Surv. **17**, 169 (2023).

[25] J. Krog, L. H. Jacobsen, F. W. Lund, D. Wüstner, and M. A. Lomholt, Bayesian model selection with fractional Brownian motion, J. Stat. Mech. (2018) 093501.

[26] R. Allison and J. Dunkley, Comparison of sampling techniques for Bayesian parameter estimation, Mon. Not. R. Astron. Soc. **437**, 3918 (2014).

[27] E. Clarkson and T. Ambjörnsson, Bayesian and frequentist analyses of two-state single-molecule diffusion trajectories, J. Phys. A: Math. Theor. **58**, 195004 (2025).

[28] F. Feroz, M. P. Hobson, and M. Bridges, MULTINEST: An efficient and robust Bayesian inference tool for cosmology and particle physics, Mon. Not. R. Astron. Soc. **398**, 1601 (2009).

[29] F. Simon, G. Ramadier, I. Fonquernie, J. Zsok, S. Patskovsky, M. Meunier, C. Boudoux, E. Dultz, and L. E. Weiss, Detecting directed motion and confinement in single-particle trajectories using hidden variables, eLife **13**, RP99347 (2024).

[30] R. Glennie, T. Adam, V. Leos-Barajas, T. Michelot, T. Photopoulou, and B. T. McClintock, Hidden Markov models: Pitfalls and opportunities in ecology, Methods Ecol. Evol. **14**, 43 (2023).

[31] R. L. Putland, L. Ranjard, R. Constantine, and C. A. Radford, A hidden Markov model approach to indicate Bryde's whale acoustics, Ecol. Indic. **84**, 479 (2018).

[32] Y. Ren, M. T. Johnson, P. J. Clemins, M. Darre, S. S. Glaeser, T. S. Osiejuk, and E. Out-Nyarko, A framework for bioacoustic vocalization analysis using hidden Markov models, Algorithms **2**, 1410 (2009).

[33] B. Manjunatha, R. Karthik, N. R. Kiran, A. P. Naik, G. N. Damodhara, H. Gunashekhar, S. M. Prem, and K. R. Mahendra, Theoretical foundations and application of hidden Markov models, J. Sci. Res. Rep. **30**, 837 (2024).

[34] C. F. Yip, W. L. Ng, and C. Y. Yau, A hidden Markov model for earthquake prediction, Stoch. Environ. Res. Risk Assess. **32**, 1415 (2018).

[35] P. J. Pedler, Occupation times for two state Markov chains, J. Appl. Probab. **8**, 381 (1971).

[36] R. J. Elliott, C. C. Liew, and T. K. Siu, Characteristic functions and option valuation in a Markov chain market, Comput. Math. Appl. **62**, 65 (2011).

[37] G. Herzog, A proof of Lie's product formula, Am. Math. Mon. **121**, 254 (2014).

[38] J. N. Darroch and K. W. Morris, Passage-time generating functions for continuous-time finite Markov chains, J. Appl. Probab. **5**, 414 (1968).

[39] L. Peliti and S. Pigolotti, *Stochastic Thermodynamics: An Introduction* (Princeton University Press, Princeton, 2021), 1st ed.