Pitching single-focus confocal data analysis one photon at a time with Bayesian nonparametrics

Meysam Tavakoli
Department of Physics, Indiana University-Purdue University Indianapolis, IN 46202

Sina Jazani, Ioannis Sgouralis
Center for Biological Physics, Department of Physics, Arizona State University, Tempe, AZ 85287

Omer M. Shafraz, Sanjeevi Sivasankar
Department of Biomedical Engineering, University of California, Davis, CA 95616

Bryan Donaphon
Biodesign Institute, Arizona State University, Tempe, AZ 85287

Marcia Levitus
Center for Biological Physics, Department of Physics, Arizona State University, Tempe, AZ 85287
Biodesign Institute, Arizona State University, Tempe, AZ 85287 and
School of Molecular Sciences, Arizona State University, Tempe, AZ 85287

Steve Pressé*
Center for Biological Physics, Department of Physics,
Arizona State University, Tempe, AZ 85287 and
School of Molecular Sciences, Arizona State University, Tempe, AZ 85287

Fluorescence time traces are used to report on dynamical properties of molecules. The basic unit of information in these traces is the arrival time of individual photons, which carry instantaneous information from the molecule, from which they are emitted, to the detector on timescales as fast as microseconds. Thus, it is theoretically possible to monitor molecular dynamics at such timescales from traces containing only a sufficient number of photon arrivals. In practice, however, traces are stochastic and in order to deduce dynamical information through traditional means–such as fluorescence correlation spectroscopy (FCS) and related techniques–they are collected and temporally autocorrelated over several minutes. So far, it has been impossible to analyze dynamical properties of molecules on timescales approaching data acquisition without collecting long traces under the strong assumption of stationarity of the process under observation or assumptions required for the analytic derivation of a correlation function. To avoid these assumptions, we would otherwise need to estimate the instantaneous number of molecules emitting photons and their positions within the confocal volume. As the number of molecules in a typical experiment is unknown, this problem demands that we abandon the conventional analysis paradigm. Here, we exploit Bayesian nonparametrics that allow us to obtain, in a principled fashion, estimates of the same quantities as FCS but from the direct analysis of traces of photon arrivals that are significantly smaller in size, or total duration, than those required by FCS.

Keywords: Single photon detection, Spectroscopy, Confocal microscopy, FCS, Bayesian nonparametrics

I. INTRODUCTION

Methods to capture static molecular structures, such as super-resolution microscopy [36, 51, 67], provide only snapshots of life in time. Yet life is dynamical and obtaining a picture of life in action–one that captures diffraction-limited biomolecules as they move, assemble into and disassemble from larger bimolecular complexes–remains an important challenge [60]. In fact, the creative insights directly leading to fluorescence correlation spectroscopy (FCS) [30, 69]–and related methods such as FCS-FRET [84, 104] and FCCS [92]–have shown that deciphering dynamical information from molecules, often biomolecules, does not demand spatial resolution or spatial localization. Rather, the key is to inhomogeneously illuminate a sample over a small volume.

As fluorescently-labeled molecules diffuse across this inhomogeneously illuminated volume, they emit photons (*i.e.*, they fluoresce) in a way that is proportional to the illumination at their respective locations [57]. Single photon detectors, often photo-multiplier tubes or avalanche photodiodes, are then used to record these photons. In principle, with the appropriate electronics, photons can be recorded within μs-ms. This suggests that information on the molecules' motion could be drawn from the

* Corresponding author. Email address: spresse@asu.edu; Website: http://statphysbio.physics.asu.edu.

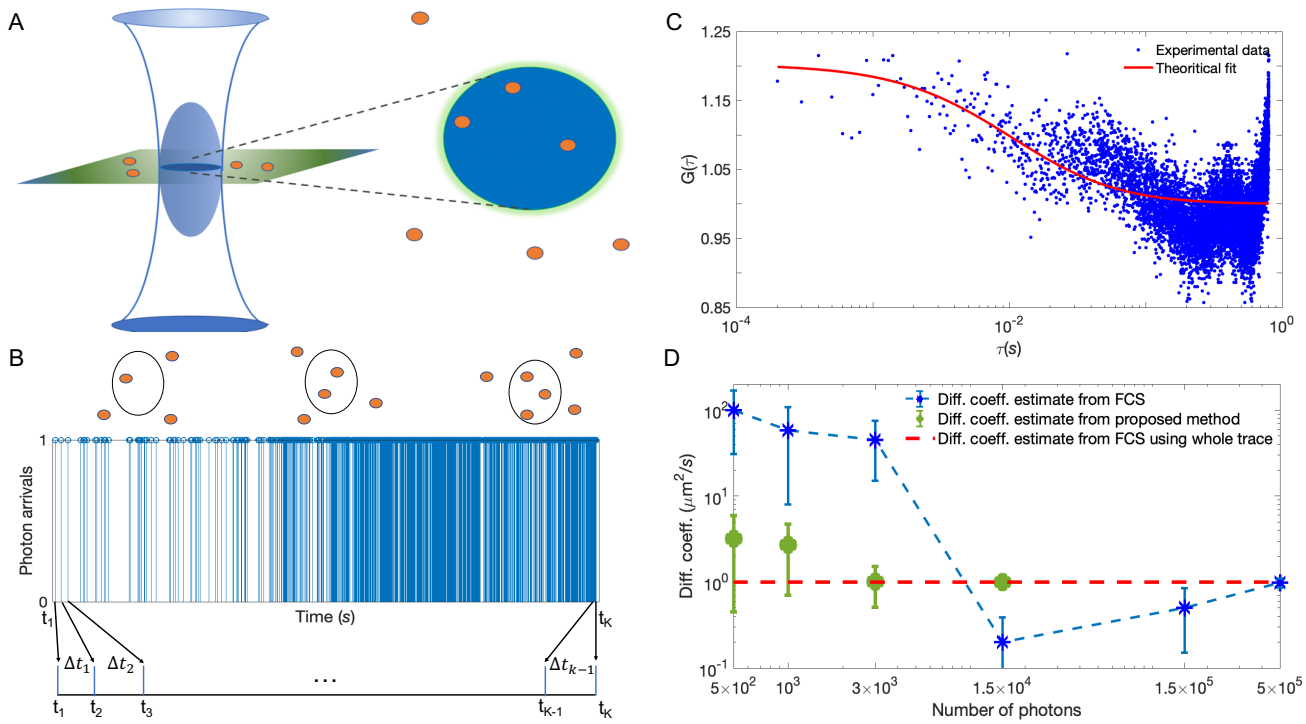

FIG. 1. **Photon arrival times can characterize dynamical properties of molecules on fast, photon-detection, timescales**. (A) Schematic of an illuminated confocal volume (blue) with fluorescent molecules emitting photons based on their location within that volume. (B) Synthetic trace containing ≈ 1500 photon arrivals produced by 4 molecules diffusing at 1 $\mu m^2/s$ for a total time of 30 ms under background and molecule photon emission rates of 10^3 photons/s and 4×10^4 photons/s, respectively. (C) Autocorrelation curve, $G(\tau)$, of the trace in (B), binned at 100 μs. On account of the limited data available in the trace, any reasonable fit is impossible. Normally, in FCS analysis, much longer traces are used to generate smoother $G(\tau)$ that are fitted to determine a diffusion coefficient. In Fig. A1 of the Appendix, we show that the quality of the fit does not improve considerably by fitting to a semi-logarithmic curve. (D) Comparison between diffusion coefficient estimates using our proposed method (detailed later) and FCS as a function of the number of photon arrivals in the analyzed trace. Since by 1.5×10^4 photon arrivals our method has converged, we avoid analyzing larger traces.

data on fast timescales that approach data acquisition, *i.e.* no more than a few μs-ms.

The fundamental quantities measured in a confocal optical setup are individual photon arrival times, from which photon inter-arrival times, *i.e.*, the intervals between adjacent photon arrivals, can be readily obtained [59]. When imaging molecules fixed in space and under homogeneous (uniform) illumination, these inter-arrival times–excluding other experimental and label artifacts such as detector noise, background photons, and label photo-physical kinetics–are independent and identically distributed and so uncorrelated with each other. However, inter-arrival times measured in conventional confocal experiments encode the number of molecules in the vicinity of the confocal volume, their diffusion dynamics, their position with respect to the confocal center in addition to an array of experiment specific artifacts such as detector characteristics and label photokinetics. Consequently, inter-arrival times are correlated with each other and, in principle, these correlations can be exploited to characterize the dynamics of the underlying molecular system.

Thus far, correlations in the inter-arrival times are exploited by collecting photons over long periods [85] and temporally autocorrelating the resulting fluorescence intensity measurements [15, 30, 35, 69]. For sufficiently long intensity traces, the stochasticity in the number of labeled molecules contributing photons, as well as their positions in the illuminated volume and their instantaneous photon emission rates, are averaged out. As such, the mathematical expression for the fluorescence intensity time-autocorrelation function takes a simple form that–under strong assumptions on the illuminated volume's geometry and the molecules' photon emission rate–can be summarized in analytic formulas that are fitted on the acquired measurements.

However, despite the elegance and simplicity of the mathematics involved in the derivation of the time-autocorrelation function [15, 30, 35, 69], a critical limitation of autocorrelative methods, including all those within the FCS framework, remains the stark timescale separation between data collection (*e.g.*, typical time between successive photon arrivals) and the timescale required to deduce a meaningful dynamical interpretation (*e.g.*, typical duration between first and last photon arrivals used); see Fig. (1). A method that takes direct ad-

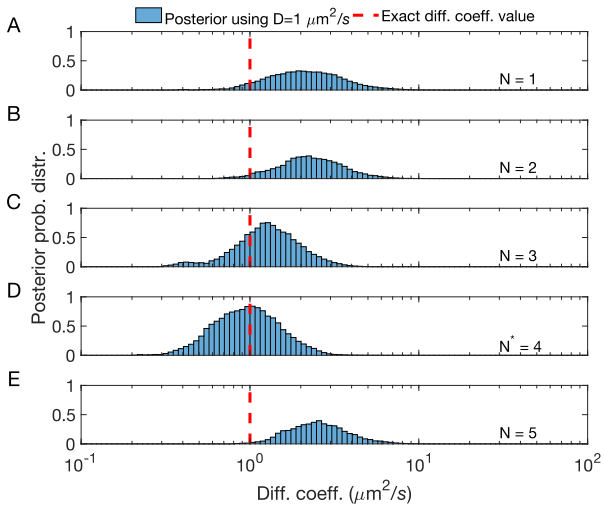

FIG. 2. **Estimates of diffusion coefficients from photon arrival traces strongly depend on the number of molecules assumed to be contributing to the trace.** The trace analyzed contained $\approx$ 1800 photon arrivals produced by 4 molecules diffusing at 1 $\mu m^2/s$ for a total time of 30 ms under background and molecule photon emission rates of 10^3 photons/s and 4×10^4 photons/s, respectively. To estimate D *parametrically,* we assumed a fixed number of molecules, $N = 1$ (A); $N = 2$ (B); $N = 3$ (C); $N = 4$ (D); and $N = 5$ (E). The correct estimate in (D)–and the mismatch in all others–underscores why it is critical to estimate the number of molecules contributing to the trace to deduce quantities such as diffusion coefficients from single photon arrivals.

vantage of single photon arrivals, without using intensity traces (*i.e.,* downsampled photon arrivals), has the potential to reveal dynamical information on timescales several orders of magnitude faster than traditional FCS analysis. As a result, rapid or non-equilibrium processes and, as such, abrupt changes in molecular chemistry, could be studied. Furthermore, provided such a method can utilize substantially shorter traces, the total duration of experiments can be shrunk and the phototoxic damage induced on biological samples can be reduced substancially [67, 70, 83, 103]. This is especially relevant for *in vivo* FCS applications [28, 80, 93, 105].

Previously proposed methods to analyze single photon measurements [2, 4, 40, 41, 45, 75, 76, 82, 108] make assumptions that render them inappropriate for imaging molecules moving through inhomogeneously illuminated volumes [40]. For example, for the analysis of single molecule fluorescence resonance energy transfer (FRET), existing methods assume that the photon inter-arrival times reflect only biomolecular conformational transitions [40–42] but not diffusive motion of the entire biomolecule [26, 27, 39, 41], and so are appropriate only for experiments on immobilized molecules. Along the same lines, existing methods combine FRET with FCS [90] to quantify ns dynamics; however, they do not directly exploit single photon measurements. Rather, they operate on downsampled measurements, achieved through binning, similar to traditional FCS, and therefore inherit the same limitations and drawbacks.

To be able to use single photon arrival times to estimate the diffusion coefficient of labeled molecules in a confocal experiment, as in most biological applications, we must be able to determine the particular number of molecules responsible for the observed photon arrival time trace. Otherwise, naively, many molecules with low diffusion coefficients emitting photons at the periphery of the illuminated confocal volume could be mistaken for fewer molecules with higher diffusion coefficients in the center region which is most illuminated. As we illustrate in Fig. (2), misidentifying the number of molecules, or incorrectly assessing their positions, may give rise to incorrect diffusion coefficient estimates.

More concretely, to obtain quantitative estimates of the diffusion coefficient, we need to formulate a likelihood [9, 37, 107]. In turn, to formulate a likelihood for photon arrival data demands that we know the number of molecules contributing photons as well as their locations across time. As the number of molecules instantaneously located within the confocal volume is unknown, all reasonable possibilities need to be considered and rank-ordered using expensive pre- or post-processing model selection heuristics [60, 100]. This has not been achieved yet, in part, because of the prohibitive computational cost it entails. Analyzing single photon arrivals from a confocal setup to derive dynamical information therefore demands fundamentally new tools.

The conceptually novel framework that we propose in this study can winnow down infinite possibilities (*i.e.*, infinite populations of molecules potentially contributing photons) to a finite, computationally manageable, number in a mathematically exact manner. Such a framework avoids compromising temporal resolution, as it requires no intensity trace to be formed (*i.e.,* no downsampling), and allows us to directly deduce dynamical quantities, such as diffusion coefficients, efficiently from raw single photon arrivals. The underlying theory, Bayesian nonparametrics (BNPs) [34], is a powerful set of tools still under active development and largely unknown to the Physical Sciences [19, 49, 53, 60, 79, 95–98, 100].

Mathematical devices within BNPs, such as the beta-Bernoulli process [1, 16, 77], allow us to place priors not only on parameters themselves, as traditional parametric Bayesian methods, but also on distributions over an infinite number of candidate models to which parameters are associated [50]. Concretely, for the case of our single photon time traces, BNPs and in particular beta-Bernoulli processes can be used to assign posterior probabilities over an array of quantities including all possible number of molecules responsible for producing the data and their associated locations at each photon arrival time. With these devices, as we describe herewith, we turn the otherwise difficult problem of model-selection–that is, determining how many molecules contribute photons–into a parameter estimation problem that remains computationally tractable [1, 16, 77].

II. MATERIALS AND METHODS

Here, we describe the mathematical formulation of our BNPs method for the analysis of confocal single photon data. We begin with the overall input which consists of photon inter-arrival times, $\boldsymbol{\Delta t} = (\Delta t_1, \Delta t_2, \ldots, \Delta t_{K-1})$ where Δt_k represents the time interval between adjacent observations of photons, which occur at times t_k with $k = 1, \ldots, K$. We also use as input the illuminated confocal volume's shape and background photon emission rate which we can determine separately through calibration [13].

To derive estimates for the diffusion coefficient from $\boldsymbol{\Delta t}$, we need to determine intermediate quantities which include: i) photon emission rates of molecular labels; and, most importantly, ii) the unknown number of molecules contributing photons to the trace $\boldsymbol{\Delta t}$, as well as their location with respect to the center of the confocal volume.

A graphical summary of our formulation is shown in Fig. (3). Below, we explain briefly each step involved. More details, and an implementation of the whole method, are available in the Appendix. In addition, source code and a GUI version of our implementation are provided through the Supplementary Materials.

A. Model Formulation

We begin with the distribution according to which the kth observation, Δt_k, is derived

$$\Delta t_k \sim \text{Exponential}\left(\mu_k\right). \tag{1}$$

Accordingly, Δt_k follows an exponential probability distribution [42, 81] with rate μ_k. In fact, the rate μ_k gathers the photon emission rates of all molecules which depend on their respective locations relative to the confocal center (see below) [13]. In addition to the molecule photon emissions rates, μ_k also includes background photons

$$\mu_k = \mu_{back} + \sum_n \mu_k^n, \tag{2}$$

where $\sum_n \mu_k^n$ is the sum over photon emission rates μ_k^n gathered from the individual molecules, that we index with $n = 1, 2, \ldots,$ and μ_{back} is the background photon emission rate. In our formulation, μ_k^n and μ_{back} are the emission rates of photons that reach our detectors which, due to optical and detector limitations, are typically lower than the rates of actual photon emissions [71, 72].

Next, we incorporate the dependency of the emission rate μ_k^n on location [5, 22, 111] with other effects such as camera pinhole shape and size, the laser intensity, laser wavelength, and quantum yield [13] into a characteristic point spread function (PSF). To be more precise, a PSF characterizes the optical response of an imaging system [12, 59, 114]. Although this term is mostly used for wide-field microscopes to describe the emission PSF, here, we follow the FCS literature, and use it to describe the confocal microscope, *i.e.,* both emission and detection PSFs. Consistent with FCS [15, 30, 35, 69], we assume a 3D Gaussian geometry [57]

$$\mu_k^n = \mu_{mol} \exp\left(-2\frac{(x_k^n)^2 + (y_k^n)^2}{\omega_{xy}^2} - 2\frac{(z_k^n)^2}{\omega_z^2}\right), \tag{3}$$

where (x_k^n, y_k^n, z_k^n) is the position of the nth molecule at time t_k and the parameter μ_{mol} indicates the brightness of a single molecule. This is the rate of detected photon emissions achieved when the molecule is at the center of the confocal volume where illumination is highest.

Finally, for a molecule diffusing along one direction, the probability distribution $p(x,t)$ of its position x at time t satisfies the diffusion equation [6, 21, 55]

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2}. \tag{4}$$

To solve this equation, we assume that the molecule is located at x_{k-1} at time t_{k-1} and we obtain

$$p(x,t) = \frac{\exp\left(-\frac{(x-x_{k-1})^2}{4(t-t_{k-1})D}\right)}{\sqrt{4\pi(t-t_{k-1})D}}, \tag{5}$$

which is the probability density of a normal random variable with mean x_{k-1} and variance $2(t-t_{k-1})D$. Therefore, at time $t = t_k$, we write

$$x_k \sim \text{Normal}\left(x_{k-1}, 2D\Delta t_{k-1}\right), \tag{6}$$

where $\Delta t_{k-1} = t_k - t_{k-1}$ and D is the molecule's diffusion coefficient. Similarly, solving the diffusion equation for molecules following isotropic diffusion in free space along all three Cartesian directions, we obtain

$$x_k^n \sim \text{Normal}(x_{k-1}^n, 2D\Delta t_{k-1}) \tag{7}$$
$$y_k^n \sim \text{Normal}(y_{k-1}^n, 2D\Delta t_{k-1}) \tag{8}$$
$$z_k^n \sim \text{Normal}(z_{k-1}^n, 2D\Delta t_{k-1}). \tag{9}$$

B. Model Inference

All quantities which we need to infer–such as the diffusion coefficient, D, locations of molecules through time, (x_k^n, y_k^n, z_k^n) and the molecule photon emission rate μ_{mol}– are formulated as model variables. We estimate these variables within the Bayesian paradigm [37, 60, 100]. The model parameters such as D and μ_{mol} require priors. Additionally, we have to consider priors on the initial molecule locations, *i.e.*, at the time of the very first photon arrival, (x_1^n, y_1^n, z_1^n). Options for these priors are straightforward and, for computational convenience, we adopt the distributions described in the Appendix.

Meanwhile, before we proceed any further with our BNPs formulation, we need to revise eq. (3) as follows

$$\mu_k^n = b^n \mu_{mol} \exp\left(-2\frac{(x_k^n)^2 + (y_k^n)^2}{\omega_{xy}^2} - 2\frac{(z_k^n)^2}{\omega_z^2}\right). \tag{10}$$

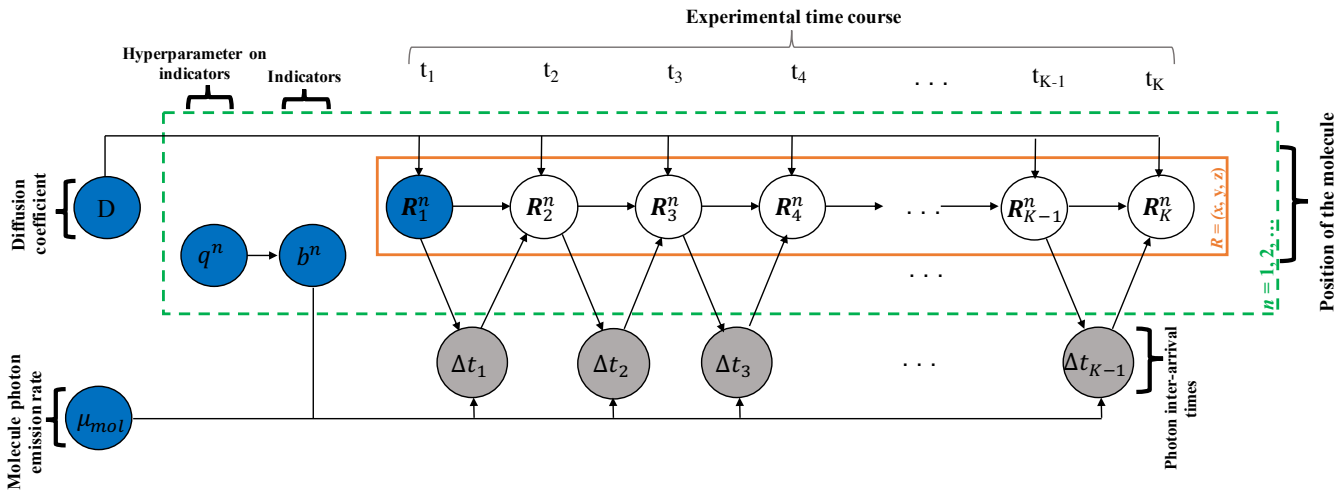

FIG. 3. **BNP formulation used for the analysis of photon arrival traces**. Molecules, indexed $n = 1, 2, \ldots$, evolve over the experimental time course which is indexed by $k = 1, 2, \ldots, K$. Here, $R_k^n = (x_k^n, y_k^n, z_k^n)$ indicates the location of molecule n at time t_k. During the experiment, only a single observation (inter-arrival time) Δt_k is recorded, thereby combining photon emissions from every molecule and the background. The diffusion coefficient D determines the evolution of the molecular positions which influence the photon emission rates and eventually the recorded Δt_k. The indicator variables b^n are introduced to infer the unknown molecule population size. In the graphical model, the measured data are highlighted by grey shaded circles and the model variables, which require priors, are designated by blue circles.

The variables b^n, defined for each model molecule, take only values 1 or 0. Specifically, we have $b^n = 0$ when the nth model molecule *does not* contribute photons to the measurements as in this case the molecule is *decoupled* from the overall photon emission rate μ_k. This indicator variable allows us to operate on an arbitrarily large population of model molecules; technically, an infinite population. The ability to recruit, from a potentially infinite pool of model molecules, the precise number that contributes to the measured trace $\mathbf{\Delta t}$ is the chief reason we abandon the parametric Bayesian paradigm and adopt BNPs. After introducing the indicators b^n, we can estimate the number of molecules that contribute photons, *i.e.*, those molecules where $b^n = 1$, simultaneously with the remaining of the parameters simply by having each b^n as a separate parameter and estimating its value.

To estimate b^n, we consider a Bernoulli prior with a beta hyper-prior

$$b^n | q^n \sim \text{Bernoulli}(q^n), \tag{11}$$

$$q^n \sim \text{Beta}(A_q, B_q), \tag{12}$$

where A_q and B_q are (hyper-hyper-)parameters specifically chosen to allow for $n \to \infty$. In this limit, eqs. (11) and (12) can be combined resulting in a beta-Bernoulli process [1, 16, 77]; see Appendix for more details.

With the specified priors, we can now form a joint posterior probability $p(D, \mu_{mol}, (x_k^n, y_k^n, z_k^n)_k^n, (b^n, q^n)^n | \mathbf{\Delta} t)$ including all unknown variables which we seek to determine. Nevertheless, the nonlinear dependence of the PSF on the molecules' positions $(x_k^n, y_k^n, z_k^n)_k^n$ and the non-parametric prior on the indicators $(b^n)^n$ exclude a closed form for our posterior. For this reason, we develop a Markov Chain Monte Carlo scheme [37, 58, 86] that exploits results from the theory of Computational Statistics and Non-linear filtering to generate pseudo-random samples from this posterior that we use in obtaining our estimates [37, 86]. A technical description of this scheme can be found in the Appendix and a ready-to-use implementation is available through the Supplementary Materials.

C. Data Acquisition

1. *Acquisition of Synthetic Data for Figs. (4)-(7)*

We acquire the synthetic data shown in the Results section by computer simulations [8, 33, 44, 47, 52] that represent Brownian motion of point molecules moving through a typical illuminated confocal volume. We provide finer details and complete parameter choices in the Appendix.

2. *Acquisition of Experiment data for Figs. (8)-(12)*

For these experiments we used Cy3 fluoresent dyes. Solutions were made by suspending Cy3 dye in glycerol/buffer (pH 7.5, 10 mM Tris-HCl, 100 mM NaCl and 10 mM KCl, 2.5 mM $CaCl_2$) at various v/v, to a final concentration of either 100 pM or 1 nM. The solution was placed in a glass- bottomed fluid-cell, assembled on a custom designed confocal microscope [63] and a 532 nm laser beam was focused to a diffraction-limited spot on the glass coverslip of the fluid-cell using a 60x, 1.42 N.A., oil-immersion objective (Olympus). In our setup, the

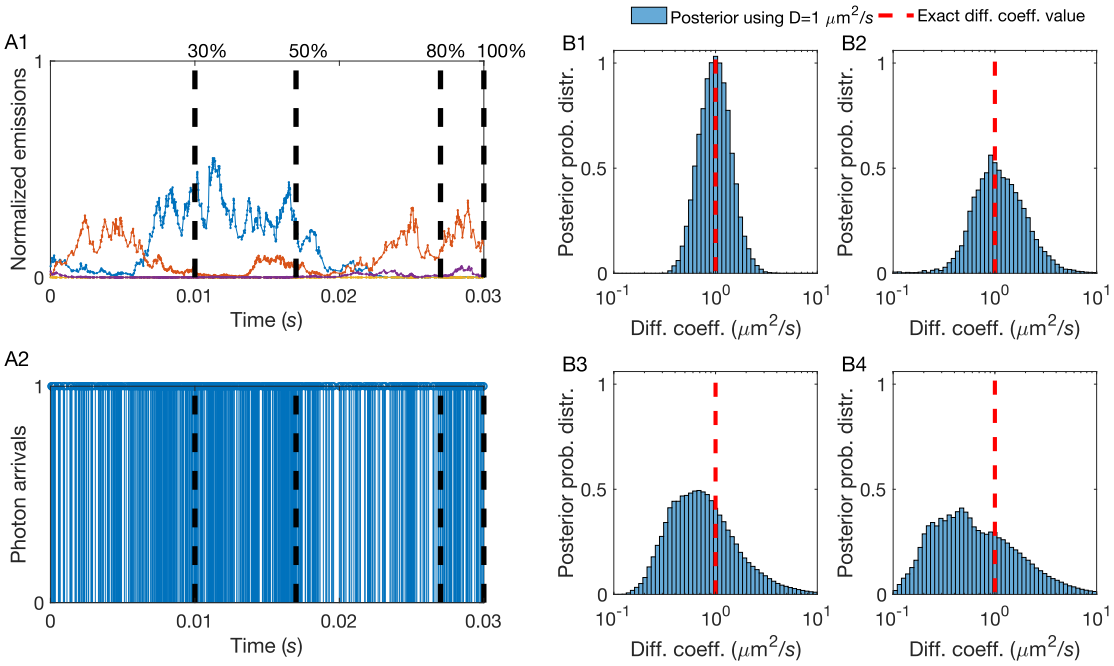

FIG. 4. **A higher number of total photon arrivals provide more photons per unit time and sharper diffusion coefficient estimates**. (A1) Instantaneous molecule photon emission rates μ_k^n, normalized by μ_{mol}. (A2) Photon arrival trace resulting from combining photon emissions from every molecule and the background. This synthetic trace contains ≈ 2000 photon arrivals produced by 4 molecules diffusing at 1 $\mu m^2/s$ for a total time of 30 ms under background and molecule photon emission rates of 10^3 photons/s and 4×10^4 photons/s, respectively. The dashed lines show the initial 30%, 50%, 80%, and 100% portions of the original trace containing ≈ 600, ≈ 1000, ≈ 1600, ≈ 2000 photon arrivals, respectively. (B1-B4) Posterior probability distributions drawn from traces with differing length (shown in (A2)). As expected, for the longer traces, the peak of the posterior matches with the exact value of D (dashed line). Gradually, as we decrease the total number of photon arrivals analyzed, the estimation becomes less reliable.

laser beam is focused at the glass-water/glycerol interface and the beam is refocused by visual inspection at the beginning of every measurement. Emitted fluorescence was collected from the same objective and focused onto a Single Photon Avalanche Diode (SPAD, Micro Photon Devices) with a maximum count rate of 11.8 Mc/s. A bandpass filter placed in front of the detector blocked all back-scattered excitation light and relayed only fluorescence from Cy3. Individual photon arrivals on the detector triggered TTL pulses and were both timestamped and registered at 80 MHz. This was achieved using a field programmable gate array (FPGA, NI Instruments) and custom LabVIEW software [89].

3. *Acquisition of Experimental Data for Fig. (13)*

For these experiments we used 5-TAMRA fluorescent dyes. The excitation source was a supercontinuum fiber laser Fianium WhiteLase SC480 (NKT Photonics, Birkerod, Denmark) operating at a repetition rate of 40 MHz. The excitation wavelength (550 nm) was selected by an acousto-optic tunable filter (AOTF), and the exiting beam was collimated and expanded by approximately a factor of three to slightly overfill the back aperture of the objective lens. The light was reflected into the objective lens (Zeiss EC Plan-Neofluar 100x oil, 1.3 NA pol M27, Thornwood, NY, USA) by a dichroic mirror (Chroma 89016bs). The same objective was used to collect the fluorescence from the sample, and passed through a band pass filter (Chroma ET575/50m) before being focused into a position motorized pinhole wheel set at 25 μm. The output of the pinhole was focused on a multimode hybrid fiber optic patch cable (M18L01, Thorlabs, NJ, USA) which was coupled to a single-photon avalanche diode (SPCM AQRH-14, Excelitas Technologies, Quebec, Canada). The detected photons were recorded by a TimeHarp 200 time-correlated single photon counting board (PicoQuant, Berlin, Germany) operating in T3 mode. The sample ($\approx$50 μL) was contained in a perfusion chamber gasket (CoverWell) adhered on a glass coverslip. The sample was 20 nM 5-Carboxytetramethylrhodamine (5-TAMRA, purchased from Sigma-Aldrich, USA) dissolved in doubly distilled water at room temperature.

III. RESULTS

Our goal is to characterize quantities that describe molecular dynamics, especially dynamics encountered in biological samples, such as diffusion coefficients, at the

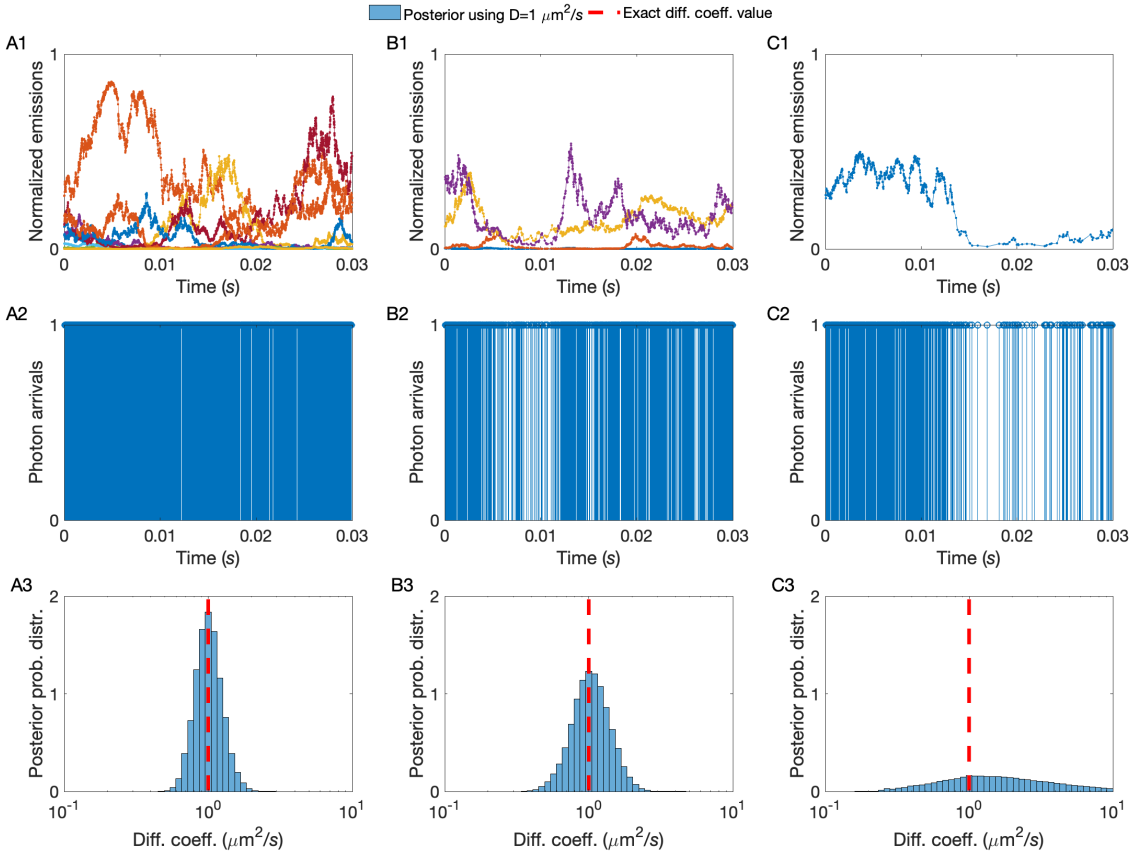

FIG. 5. **A higher molecular concentration provides more photons per unit time and sharper diffusion coefficient estimates**. (A1, B1, C1) Instantaneous molecule photon emission rates μ_k^n, normalized by μ_{mol}. (A2, B2, C2) Photon arrival traces resulting from combining photon emissions from every molecule and the background. These are produced by 10 molecules containing $\approx$ 3000 photon arrivals (A2), 4 molecules containing $\approx$ 2000 photon arrivals (B2), and 1 molecules containing $\approx$ 1000 photon arrivals (C2), diffusing at 1 $\mu m^2/s$ for a total time of 30 ms under background and molecule photon emission rates of 10^3 photons/s and 4×10^4 photons/s, respectively. (A3, B3, C3) Posterior probability distributions drawn from traces with differing number of molecules (shown in (A2, B2, C2)). As expected, for the traces with higher number of molecules, the peak of the posterior matches with the exact value of D (dashed line). Gradually, as we decrease the total number of molecules the estimation becomes less reliable.

data-acquisition timescales of conventional single-focus confocal setups. Our input consists of: i) the measured photon inter-arrival times $\boldsymbol{\Delta t} = (\Delta t_1, \Delta t_2, ..., \Delta t_{K-1})$; ii) the background photon emission rate; and iii) the geometry of the illuminated volume specified through a characteristic PSF.

As we explain in the Methods section, in order to estimate the molecules' diffusion coefficient, D, we also estimate intermediate quantities (namely, molecule photon emission rates, molecule positions over time and the molecule numbers in the first place). These intermediate quantities demand that we use BNPs to determine quantities that *a priori* may be arbitrarily large such as the number of molecules contributing photons to our datasets $\boldsymbol{\Delta t}$.

Within the Bayesian paradigm [60, 107], our estimates take the form of posterior probability distributions over the unknown quantities. These distributions combine parameter values, probabilistic relations among different parameters, as well as the associated uncertainties. According to the common statistical interpretation [37, 107], the sharper the posterior, the more conclusive (and certain) the estimate. To quantify the uncertainty, we compute a posterior variance and use the square root of this variance to construct error-bars (*i.e.*, credible intervals) [37, 107]. In Table I in the Appendix, we summarize the mean values and error bars of our analyses.

Below, we validate first our method on synthetic data where the ground truth is available. For these, we use a confocal volume of typical size $\omega_{xy} = 0.3$ μm and $\omega_z = 1.5$ μm [59]. We then test our method on experimental data collected in two labs utilizing different FCS setups. For the latter cases, we demonstrate the advantages of our method by comparing our results to the results obtained from autocorrelative methods used in FCS analysis.

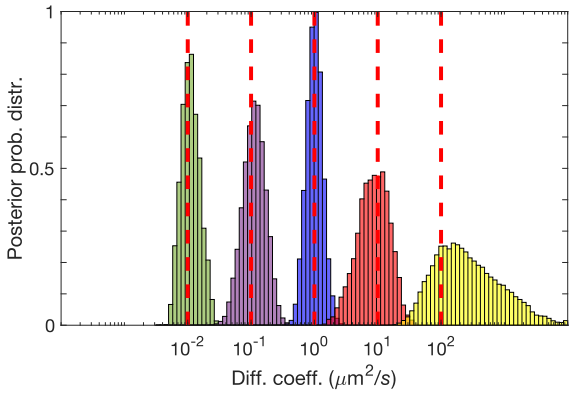

FIG. 6. **A lower diffusion coefficient provides more photons per unit time and sharper diffusion coefficient estimates**. Posterior probability distributions drawn from traces containing ≈ 2000 photon arrivals produced by 4 molecules diffusing at $D = 0.01, 0.1, 1, 10\ \mu m^2/s$ for a total time of 30 ms under background and molecule photon emission rates of 10^3 photons/s and 4×10^4 photons/s, respectively. For molecules diffusing at $D = 100\ \mu m^2/s$, under similar conditions, we used a trace containing ≈ 3000 photons for a total time of 50 ms, since we needed a longer trace to gather sufficient information for drawing a posterior.

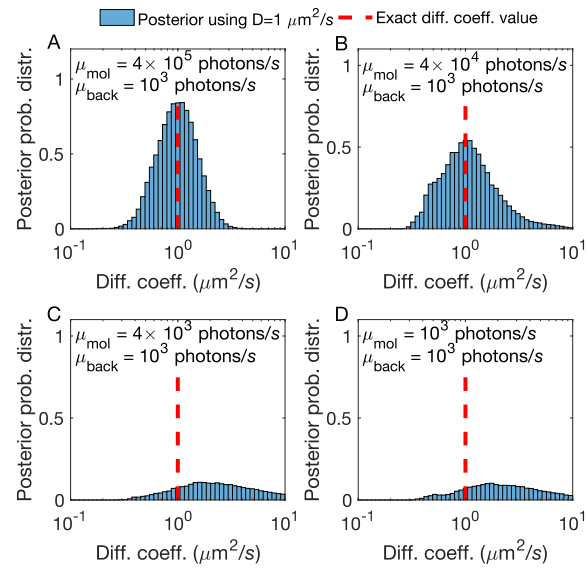

FIG. 7. **A higher molecule photon emission rate provides more photons per unit time and sharper diffusion coefficient estimates**. (A, B, C, D) Posterior probability distributions drawn from traces produced by 4 molecules diffusing at 1 $\mu m^2/s$ for a total time of 30 ms under background photon emission rate of 10^3 photons/s and molecule photon emission rates 4×10^5, 4×10^4, 4×10^3, 10^3 photons/s, respectively. As expected, under higher molecule photon emission rates, the peak of the posterior matches sharply with the exact value of D (dashed line). Gradually, as we decrease the molecule photon emission rate, the estimation becomes less reliable.

A. Method Validation using Simulated Data

To demonstrate the robustness of our approach, we simulate raw single photon arrival traces under a broad range of: i) total photon arrivals, Fig. (4); ii) concentrations of labeled molecules, Fig. (5); iii) diffusion coefficients, Fig. (6); and iv) molecule photon emission rates, Fig. (7). The parameters not varied are held fixed at the following baseline values: diffusion coefficient of 1 $\mu m^2/s$ which is typical of slower *in vivo* conditions [7, 68, 88, 105], molecule photon emission rates of 4×10^4 photons/s [62, 82], and 4 as the number of labeled molecules contributing photons. We chose 4, a small number of molecules (as opposed to a larger number of molecules), because this scenario presents the greatest analysis challenge as very few photons, and thus little data, are gathered to aid the analysis.

As illustrated in Fig. (1), a critical and recurring point throughout this section is that the traces we analyze are shorter than those that could be meaningfully analyzed using FCS. While we focus on the diffusion coefficient estimation here, we note that our framework supports more detailed parameter estimation which we provide in the Appendix.

1. *Total Photon Arrivals*

We evaluate the robustness of our method with respect to the length of the trace (*i.e.*, the total number of photon arrivals recorded) at a fixed number of molecules, diffusion coefficient, and molecule photon emission rates. The first important finding is that, for the values of parameters selected, we need 2 orders of magnitude less data than FCS; see Fig. (1D). For instance, to obtain an estimate of the diffusion coefficient within 10% of the ground truth value, we require $\approx 10^3$ photons (directly emitted from the labeled molecule), while FCS requires $\approx 10^5$ photons. Under our simulated scenario, these correspond to traces of total duration $30 - 50$ ms and 50 s, respectively. To determine our error, we chose the mean value of the diffusion coefficient's marginal posterior, $p(D|\mathbf{\Delta} t)$, and measure the percentage difference of this mean value to the ground truth.

In general, the precise photon numbers demanded by our method and traditional FCS depend on a broad range of experimental parameter settings. This is the reason, we explore different settings in subsequent subsections as well as the Appendix.

An important overarching concept is the concept of a photon arrival as a unit of information. The more photon arrivals we have in the analyzed trace, the sharper our diffusion coefficient estimates become. This is valid, as we see in Fig. (1D) and Fig. (4), for increasing total photon arrivals. Similarly, as we see in subsequent subsections, we also collect more photons as we increase the concentration of labeled molecules (and thus the number of molecules contributing photons to the trace), increase the molecule photon emission rates of molecular labels,

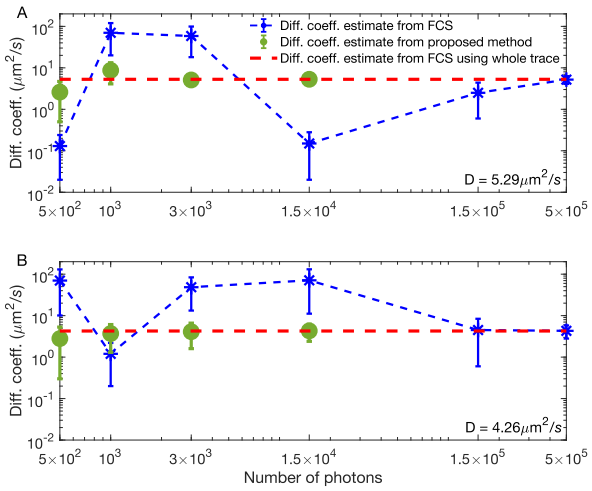

FIG. 8. **Higher molecular concentrations in experimental traces provide more photons per unit time resulting in sharper diffusion coefficient estimates**. Estimates shown are drawn from experimental traces with a low (100 pM) (A) and high (1 nM) (B) concentration of Cy3 dye molecules and 75% glycerol at a fixed laser power of 100 μW. Similarly to Fig. (1), we compare our method's diffusion coefficient estimate (circle green dots) to FCS (blue asterisk) as a function of the number of photons used in the analysis. Since by 1.5×10^4 photon arrivals our method has converged, we avoid analyzing larger traces. The red dash line is the FCS estimate produced from the entire 5 min trace containing $\approx 3 \times 10^6$ photon arrivals.

or decrease diffusion coefficients of molecules. In the latter case, a slower diffusion coefficients provides more time for each molecule to traverse the illuminated region, in turn, resulting in more photon arrivals.

2. *Molecule Concentration*

To test the robustness of our method under different concentrations of labeled molecules at fixed diffusion coefficient, and molecule photon emission rates, we simulate molecules diffusing at 1 μm^2/s for a total time 30 ms with: i) average concentrations of 10 molecules/μm^3, Fig. (5A1, A2); ii) 4 molecules/μm^3, Fig. (5B1, B2); and iii) 1 molecule/μm^3, Fig. (5C1, C2). The molecule and background photon emission rates are taken to be 4×10^4 photons/s and 10^3 photons/s respectively, which are typical of confocal imaging [82].

Figure (5) summarizes our results and suggests that posteriors over diffusion coefficients are broader–and thus the accuracy with which we can pinpoint the diffusion coefficient drops–when the concentration of labeled molecules is lower. Intuitively, we expect this result as fewer molecules within the confocal volume provide fewer photons arrivals.

3. *Diffusion Coefficients*

We repeat the simulations of the previous subsection to demonstrate, using synthetic data, the robustness of our method with respect to the diffusion coefficient magnitude at fixed number of molecules, and molecule photon emission rates; see Fig. (6). Intuitively, and again on the basis of the fact that photon arrivals carry information, we expect that faster moving molecules give rise to broader posterior distributions as these emit fewer photons, and thus provide less information, while they traverse the confocal volume.

4. *Molecule Photon Emission Rates*

Figure (7) illustrates the robustness of our method with respect to the molecule photon emission rates (*i.e.*, set by the laser power used in the experimental setting and the choice of fluorescent label) by fixing the number of molecules, diffusion coefficient (1 $\mu m^2/s$), and background emission (10^3 photons/s). To accomplish this, we simulate increasingly dimmer molecules until the molecule signature is effectively lost in the background. As expected, dimmer molecules lead to broader posterior estimates over diffusion coefficients as these traces are associated with higher uncertainty.

B. Estimation of Physical Parameters from Experimental Data

To evaluate our BNPs method on real data, we used experimental single photon traces collected under a broad range of conditions. That is, we used measurements from two different experimental setups and different fluorescent dyes, that are commonly used in labeling biological samples, as well as diffusing labeled proteins. Additional differences between the setups include different numerical apertures (NA), laser powers, and overall detection instrumentation as detailed in the Methods section.

Figures (8)-(11) were collected using the Cy3 dye and these results were used to benchmark the robustness of our method on dye concentration, diffusion coefficients, and laser power. Moreover, to evaluate the proposed approach beyond free dyes, in Fig. (12), we used labeled proteins, namely freely diffusing streptavidin labeled with Cy3. For Fig. (13), photon arrivals were collected using 5-TAMRA dye in order to test the robustness of our method on a different fluorophore.

1. *Benchmarking on Experimental Data using Cy3*

We begin by verifying our method on mixtures of water and glycerol. While we only use short segments in our analysis, the collected traces are long enough (≈5 min

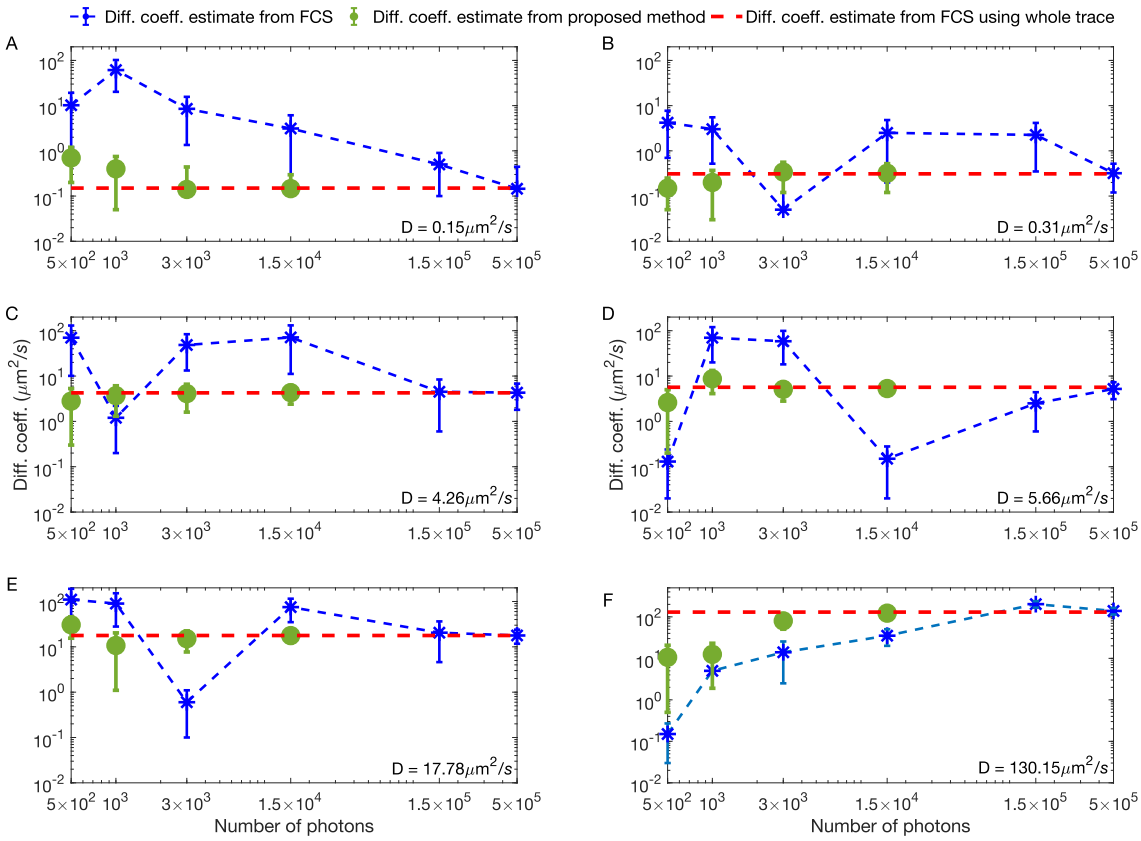

FIG. 9. **Lower diffusion coefficients in experimental traces provide more photons per unit time and sharper diffusion coefficient estimates**. Estimates shown are drawn from experimental traces with 99% glycerol (A), 94% glycerol (B), 75% glycerol (C), 67% glycerol (D), 50% glycerol (E), and 0% glycerol (F) with fixed concentration 1 nM of Cy3 dye molecules and laser power of 100 μW. Similarly to Fig. (1), we compare our method's diffusion coefficient estimate (circle green dots) to FCS (blue asterisk) as a function of the number of photons used in the analysis. Since by 1.5×10^4 photon arrivals our method has converged, we avoid analyzing larger traces. The red dash line is the FCS estimate produced from the entire 5 min trace containing $\approx 3 \times 10^6$ photon arrivals.

each) to be meaningfully analyzed by traditional autocorrelative analysis used in FCS for sake of comparison. The result of the analysis of the full trace by FCS yields a diffusion coefficient that we treat as an effective ground truth. We then ask how long of a trace our method requires, as compared to FCS, in order for our diffusion coefficient estimate to converge to this ground truth.

Our strategy addresses the following complication: we anticipate that the PSF may be distorted from the idealized shape assumed especially with increasing amounts of glycerol [38]. However, the same (possibly incorrect) PSF is used in both FCS and our method in order to compare both methods head-to-head. Thus, concretely, we are asking: how many photon arrivals do we need to converge to the *same* result as FCS (irrespective of whether the FCS result is affected by PSF distortion artifacts)?

Our single photon traces are obtained under a range of conditions, namely different: i) dye concentrations, Fig. (8); ii) diffusion coefficients, Fig. (9); and iii) laser powers, Fig. (10). As before, longer traces, higher concentrations, lower diffusion coefficients, and higher laser powers result, on average, in sharper estimates with the results still converging with at least 2 orders of magnitude fewer photon arrivals than FCS for equal accuracy in Figs. (8), (9), and Fig. (10), respectively. We mention "on average" as individual traces are stochastic. Thus, some traces under higher concentrations of fluorescent molecules may happen to have fewer molecules contribute photons to the traces than experiments with lower concentrations.

Figures (8) recapitulates our expectations derived from the synthetic data shown earlier (Fig. (5)), where dye concentrations are low yielding a wider posterior for our diffusion coefficient and correspondingly sharper posteriors for the higher concentration. Here, similarly to Fig. (1), we compare our method's diffusion coefficient estimate to FCS as a function of the number of photon arrivals used in the analysis, Fig. (8A) and Fig. (8B), both in good agreement with FCS estimates, produced by the entire traces which is $\approx 10^3$ times longer.

Similar to the analysis of synthetic data, by comparing different diffusion coefficients, the slower a diffusing

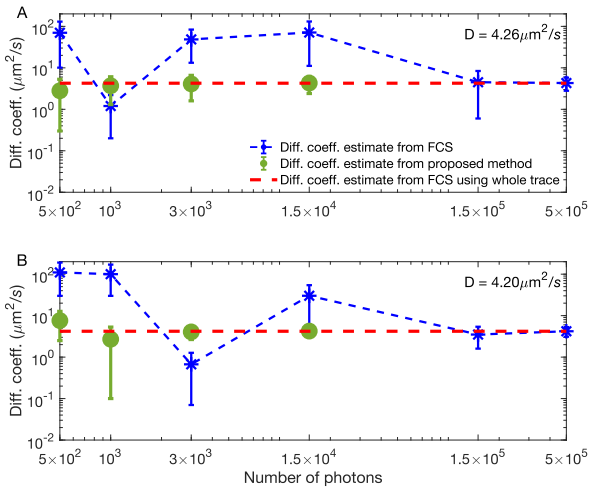

FIG. 10. **Higher laser powers in experimental traces provide more photons per unit time and sharper diffusion coefficient estimates**. Estimates shown are drawn from experimental traces with high (100 μW) (A) and low (25 μW) (B) laser power with fixed concentration 1 nM of Cy3 dye molecules and 75% glycerol. Similarly to Fig. (1), we compare our method's diffusion coefficient estimate (circle green dots) to FCS (blue asterisk) as a function of the number of photons used in the analysis. Since by 1.5×10^4 photon arrivals our method has converged, we avoid analyzing larger traces. The red dash line is the FCS estimate produced from the entire 5 min trace containing $\approx 3 \times 10^6$ photon arrivals.

molecule is, the more time it spends within the confocal volume, the more photons are collected providing us with a sharper posterior estimate of its diffusion coefficient (see Fig. (9)).

Similarly to the synthetic data shown earlier (Fig. (7)), Fig. (10) illustrates the robustness of our method to lower laser power which, as expected, yields a wider posterior for our diffusion coefficient and correspondingly sharper posteriors for higher laser power. Here, we compare our method's diffusion coefficient estimate to FCS as a function of the number of photon arrivals used in the analysis, Fig. (10A) and Fig. (10B), both in good agreement with FCS estimates, produced from the entire trace.

As further controls, Fig. (11) demonstrates a set of analysis where the background photon emission rate is artificially added to real data. In these cases, we test the limits of our method on more challenging imaging conditons. Furthermore, we repeat our analysis on single photon traces produced by a labeled biomolecule. Specifically, in Fig. (12), we use streptavidin proteins labeled with Cy3.

2. *Benchmarking on Experimental Data using 5-TAMRA*

Finally, we switch to a different dye, different setup and acquisition electronics as detailed in the Methods section. Our sample contained 20 nM of 5-TAMRA dissolved in water. As previously, we successfully benchmark our estimates of the diffusion coefficient versus the value obtained from FCS on much longer (≈10 min) traces, see Fig. (13).

IV. DISCUSSION

A single photon arriving at a detector mounted to a confocal microsocope encodes information that reports on the fastest timescale achievable for spectroscopic and imaging applications [59, 73]. Directly exploiting this information can help uncover the dynamics of physical or biological systems at fast timescales with accuracy superior to that obtained from *derived* quantities such as down-sampled intensity traces.

Our method takes a Bayesian nonparametrics (BNPs) approach to tackling single photon arrival data to characterize dynamical quantities from as few as hundreds to thousands of datapoints from confocal imaging. This is by contrast to conventional autocorrelative methods used in FCS [30, 69, 84, 104] that require dramatically more data, *i.e.*, datasets several orders of magnitude larger in either total duration or total number of photon arrivals, to characterize dynamical quantities with similar accuracy.

There have been partial solutions to the challenge of interpreting single molecule data at the single photon level often outside FCS applications. Indeed, existing methods make assumptions that render them inapplicable to diffusion through inhomogenesouly illuminated volumes. For example, they assume uniform illumination [40, 82], apply downsampling or binning and thereby reduce temporal resolution to exploit existing mathematical frameworks such as the hidden Markov model [2, 17, 41, 76, 106], or focus on immobile molecules [26, 27, 39, 41]. More recently, fluorescence-based nanosecond FCS approaches, in which the data are still correlated under the assumption that the time trace reports on processes at equilibrium, have been used to obtain information on rapid fluctuations in proteins [90]. As such, correlative methods largely continue to dominate confocal data analysis almost half a century beyond their inception [30, 57, 69].

To take full advantage of single photon data, new Mathematics are required. These must treat the inherent non-stationarity between photon arrivals arising due to molecular diffusion in an inhomogeneously illuminated volume and the stochastic number of molecules contributing photons. In particular, analyzing data derived from mobile molecules within an illuminated confocal region breaks down the perennial parametric Bayesian paradigm that has been the workhorse of data analysis [17, 46, 60, 65, 76, 100, 105]. We argue here that BNPs–which provide principled extensions of the Bayesian methodology [34, 102]–show promise in Physics [48, 54, 60, 95, 96, 98, 100] and give us a working solution to fundamental parametric challenges.

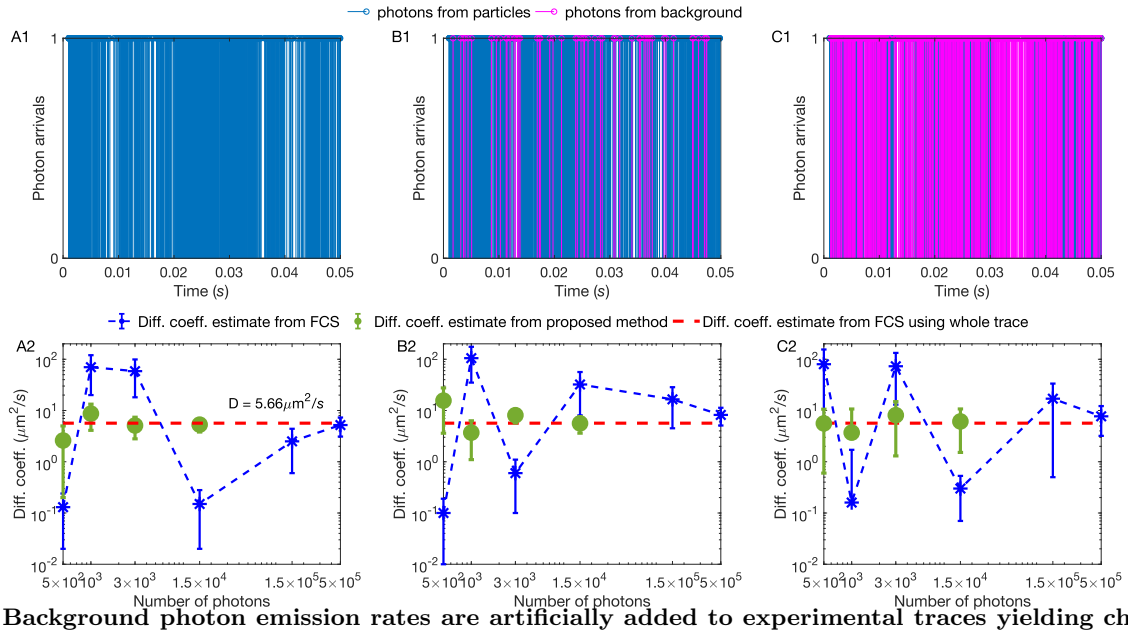

FIG. 11. **Background photon emission rates are artificially added to experimental traces yielding challenging imaging conditions and broader diffusion coefficient estimates**. Experimental traces with fixed concentration 1 nM of Cy3 dye molecules and 67% glycerol and fixed laser power 100 μW. The same total number of photons analyzed under differing (artificially increased) background photon emission rates (0 (A1), 500 (B1), 1000 (C1) photons/s). (A2, B2, C2) Similarly to Fig. (1), we compare our method's diffusion coefficient estimate (green dots) to FCS (blue asterisk) as a function of the number of photons used in the analysis. Since by 1.5×10^4 photon arrivals our method has converged, we avoid analyzing larger traces. The red dash line is the FCS estimate obtained from the entire, 5 min, trace containing $\approx 3 \times 10^6$ photon arrivals.

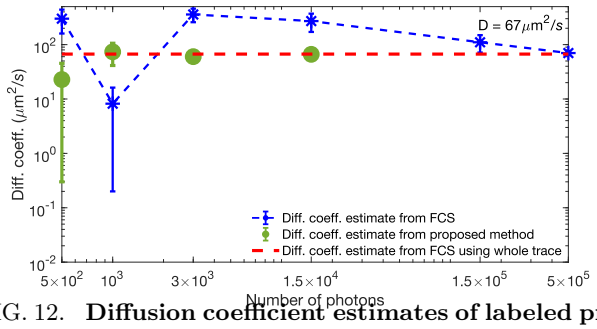

FIG. 12. **Diffusion coefficient estimates of labeled protein.** Estimates shown are drawn from experimental traces with fixed concentration 1 nM of Cy3-labeled streptavidin molecules and laser power 100 μW. Similarly to Fig. (1), we compare our method's diffusion coefficient estimate (green dots) to FCS (blue asterisk) as a function of the number of photons used in the analysis. Since by 1.5×10^4 photon arrivals our method has converged, we avoid analyzing larger traces. The red dash line is the FCS estimate obtained from the entire, 5 min, trace containing $\approx 3 \times 10^6$ photon arrivals.

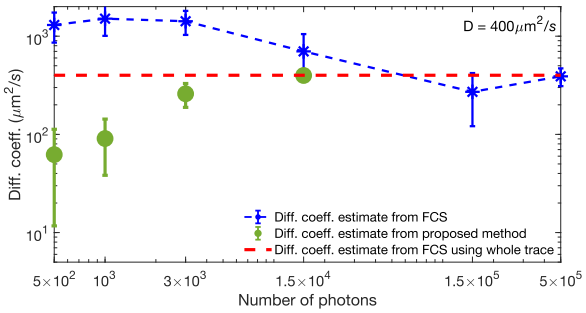

FIG. 13. **Diffusion coefficient estimates of 5-TAMRA dye**. Estimates shown are drawn from experimental traces with fixed concentration 20 nM of 5-TAMRA dye molecules. Similarly to Fig. (1), we compare our method's diffusion coefficient estimate (green dots) to FCS (blue asterisk) as a function of the number of photons used in the analysis. Since by 1.5×10^4 photon arrivals our method has converged, we avoid analyzing larger traces. The red dash line is the FCS estimate obtained from the entire, 10 min, trace containing $\approx 6 \times 10^6$ photon arrivals.

Our new tools open up the possibility to explore at the single photon level non-equilibrium processes resolved on fast timescales [3, 74], reaching ms or even below, that have been the focus of recent attention [25]. Moreover, and of immediate relevance for biophysical applications, if a single molecule photobleaches after emitting just a few hundred photons, then our novel method can still provide a diffusion coefficient estimate. Additionally, by analyzing single photon data pointwise, as we do in this study, we obtain a better handle on error bars than analyzing post-processed, such as correlated, data where the error bars can become difficult to compute or interpret [56, 87]. As such, a sharp diffusion coefficient posterior may not only suggest a good estimate of the diffusion coefficient but also suggest that the underlying model, such as normal diffusion, is appropriate and *vice*

versa a broad posterior may suggest a poor estimate or an inappropriate motion model.

Furthermore, armed with a transformative framework, founded upon rigorous Statistics, it is now possible to extend the proof-of-principle study to treat effects that lie beyond the current scope of this work. In particular, we can extend our framework to treat multiple color imaging [24], triplet effect and complex molecule photophysics [43] (such as molecular blinking [97, 113] and photobleaching [61, 99]), more complex molecule motion models [101, 112] other than free diffusion [53], distorted or abberated PSF models [32], or even incorporate chemical reactions among the molecules [10, 109]. As our BNP framework explicitly represents the instantaneous position of each involved molecule throughout the experiment's time course, these are extensions that require modest modifications.

ACKNOWLEDGEMENTS

S.P. acknowledges support from the NIH NIGMS (R01GM130745) for supporting early efforts in nonparametrics and NIH NIGMS (R01GM134426) for supporting single photon efforts. S.S. acknowledges support from the NIH NIGMS (R01GM121885).

AUTHOR CONTRIBUTIONS

MT analyzed data and developed analysis software; MT, SJ, IS developed computational tools; OS, SS, BD, ML contributed experimental data; MT, SJ, IS, SP conceived research; SP oversaw all aspects of the projects.

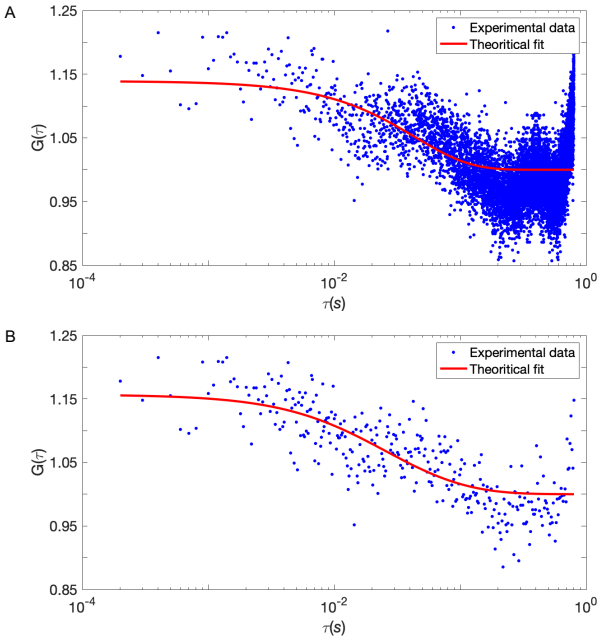

FIG. A1. **FCS curves resulting from exceedingly short traces (same synthetic data as Fig. 1) with linear (A) and semi-logarithmic (B) binning.** Due to the limited data, the quality of the fitted autocorrelation curve, $G(\tau)$, does not improve considerably for (B) as compared to (A).

APPENDIX

Additional Analysis Results

In Fig. (A2) we illustrate the weakness of FCS analysis when applied on limited datasets such as those in the scope of our method. Additionally, using synthetic data, in Fig. (A3) we estimate diffusion coefficients faster than those in the Results section and in Fig. (A4) we estimate photon emission rates. Finally, using experimental data of Cy3 and 5-TAMRA dyes obtained as described in the Methods section, in Figs. (A5) and (A6), respectively, we benchmark the same estimates on real data.

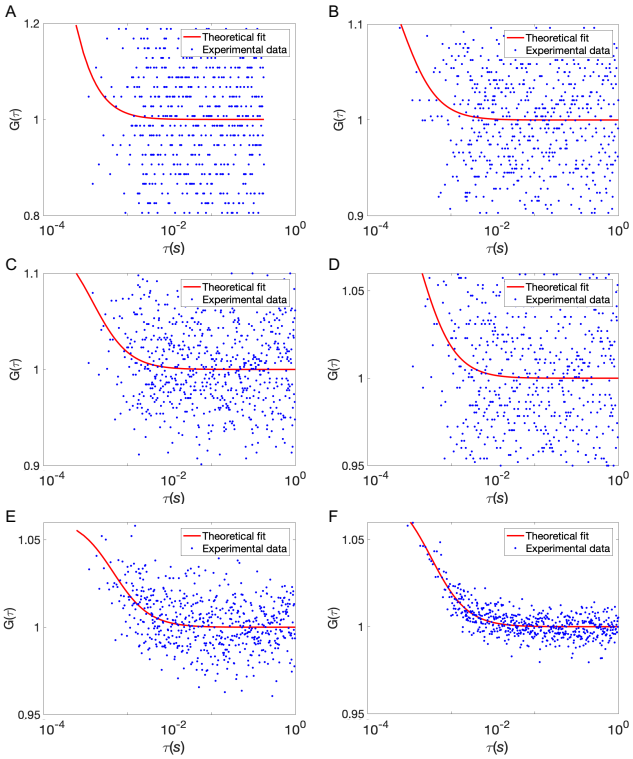

FIG. A2. **FCS curves resulting from exceedingly short traces.** Shown are autocorrelation curves, $G(\tau)$, of 5-TAMRA experimental traces, binned at 10 μs, for 100 ms and $\approx$ 500 photon arrivals (A); 200 ms and $\approx$ 1000 photon arrivals (B); 300 ms and $\approx$ 3000 photon arrivals (C); 2 s and $\approx$ 15000 photon arrivals (D); 30 s and $\approx 15 \times 10^5$ photon arrivals (E); 100 s and $\approx 15 \times 10^6$ photon arrivals (F). Even a visual inspection illustrates how poorly FCS applies on traces as sort as those analyzed by our BNP method.

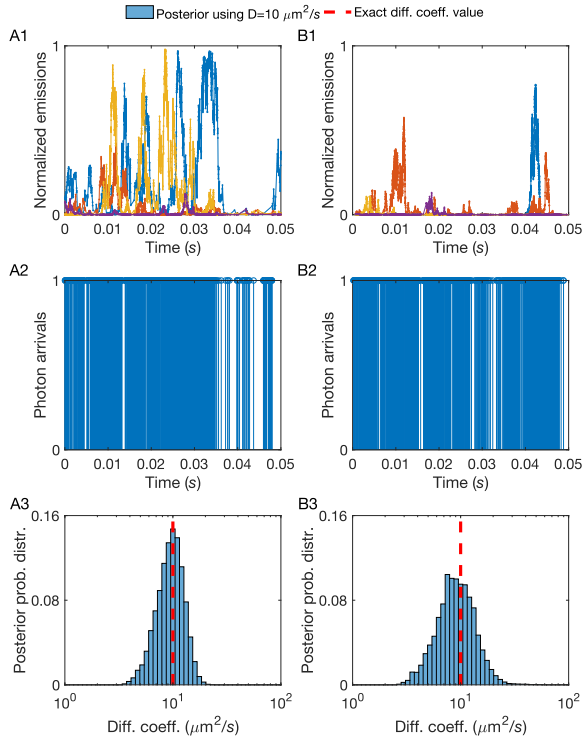

FIG. A3. **A larger molecule photon emission rate provides more photons per unit time and sharper diffusion coefficient estimates**. (A1, B1) Instantaneous molecule photon emission rates μ_k^n, normalized by μ_{mol}. (A2, B2) Photon arrival trace resulting from combining photon emissions from every molecule and the background. These traces are produced by 10 molecules diffusing at 10 $\mu m^2/s$ for a total time of 50 ms under background photon emission rate of 10^3 photons/s and molecule photon emission rate 4×10^5 photons/s containing ≈ 3000 photon arrivals (A2), and molecule photon emission rate 4×10^4 photons/s containing ≈ 2000 photon arrivals (B2). (A3, B3) Posterior probability distributions drawn from traces with differing molecule photon emission rates (shown in (A2, B2)). As expected, for the traces with higher molecule photon emission rate, the peak of the posterior sharply matches with the exact value of D (dashed line). Gradually, as we decrease the molecule photon emission rate, the estimation becomes less reliable.

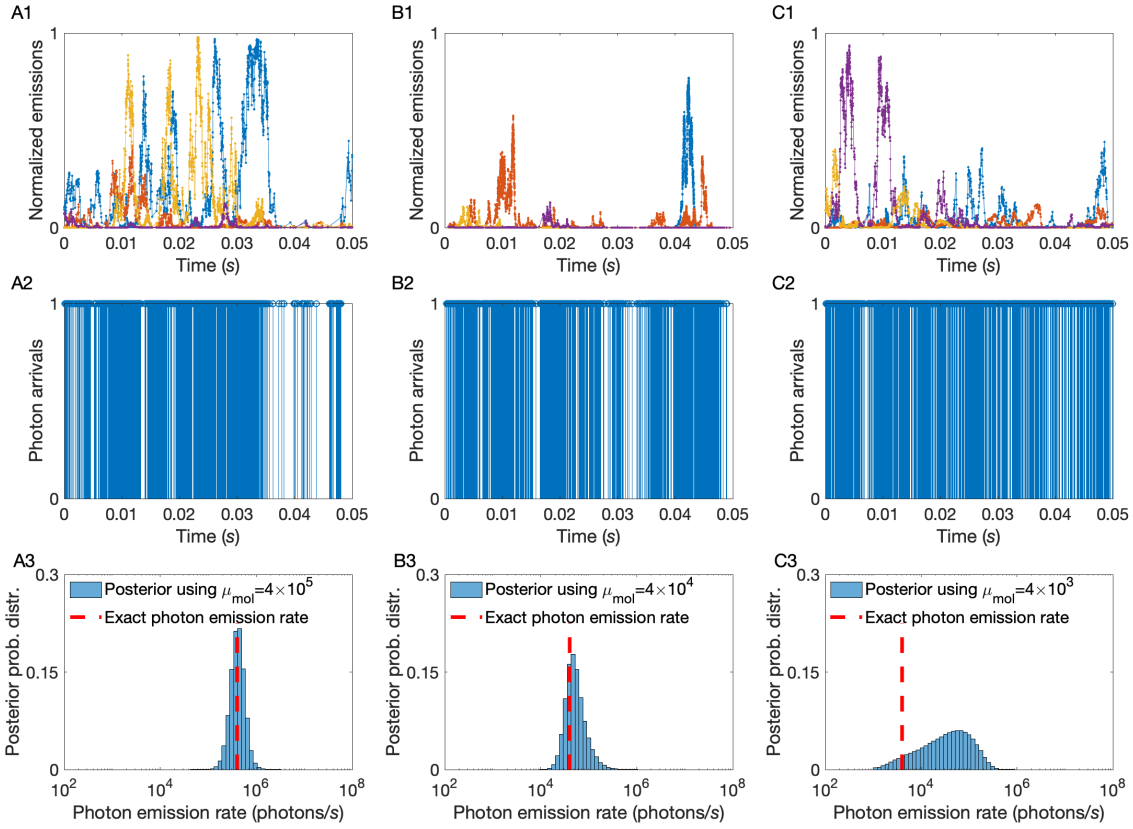

FIG. A4. **A higher molecule photon emission rate provides more photons per unit time and sharper emission rate estimates**. (A1, B1, C1) Instantaneous molecule photon emission rates μ_k^n, normalized by μ_{mol}. (A2, B2, C2) Photon arrival traces resulting from combining photon emissions from every molecule and the background. These traces produced by 10 molecules diffusing at 10 $\mu m^2/s$ for a total time of 50 ms under background photon emission rate of 10^3 photons/s and molecule photon emission rate 4×10^5 photons/s containing ≈ 3000 photon arrivals (A2), molecule photon emission rate 4×10^4 photons/s containing ≈ 2000 photon arrivals (B2), and molecule photon emission rate 4×10^3 photons/s containing ≈ 1000 photon arrivals (C2). (A3, B3, C3) Posterior probability distributions drawn from traces with differing molecule photon emission rates (shown in (A2, B2, C2)). As expected, for the traces with higher molecule photon emission rate, the peak of the posterior sharply matches with the exact value of μ_{mol} (dashed line). Gradually, as we decrease the molecule photon emission rate, the estimation becomes less reliable.

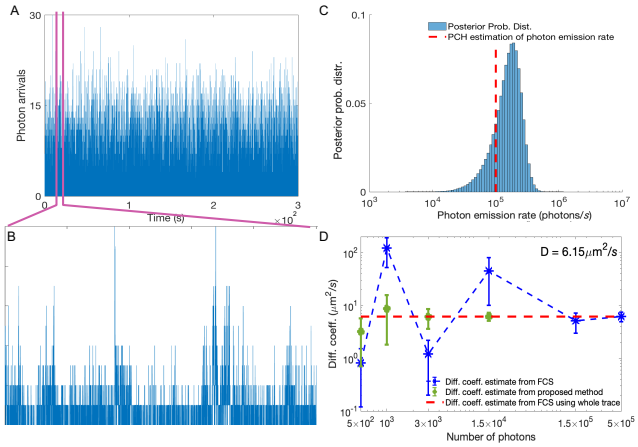

FIG. A5. **Estimation of the diffusion coefficient and molecule photon emission rate for Cy3 dyes**. (A) Experimental intensity trace (binned at 100 μs) with concentration 1 nM of Cy3 dye molecules and 61% glycerol. A background photon emission rate of 600 photons/s is known from calibration. (B) Analyzed portion of the trace containing $\approx$ 3000 photon arrivals. (C) Posterior probability distributions and the value (red dash line) of molecule photon emission rate determined by the photon counting histogram (PCH) method on the entire trace [22]. (D) Similarly to Fig. (1), we compare our method's diffusion coefficient estimate (green dots) to FCS (blue asterisk) as a function of the number of photons used in the analysis. Since by 1.5×10^4 photon arrivals our method has converged, we avoid analyzing larger traces. The red dash line is the FCS estimate obtained from the entire, 5 min, trace containing $\approx 3 \times 10^6$ photon arrivals.

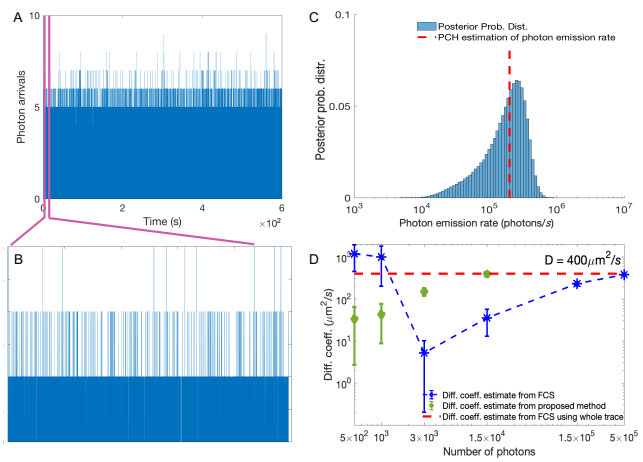

FIG. A6. **Estimation of the diffusion coefficient and molecule photon emission rate for 5-TAMRA dyes**. (A) Experimental intensity trace (binned at 10 μs) with concentration 20 nM of 5-TAMRA dye molecules. A background photon emission rate of 300 photons/s is known from calibration. (B) Analyzed portion of the trace containing $\approx$ 8000 photon arrivals. (C) Posterior probability distributions and the value (red dash line) of molecule photon emission rate determined by the photon counting histogram (PCH) method on the entire trace [22]. (D) Similarly to Fig. (1), we compare our method's diffusion coefficient estimate (green dots) to FCS (blue asterisk) as a function of the number of photons used in the analysis. Since by 1.5×10^4 photon arrivals our method has converged, we avoid analyzing larger traces. The red dash line is the FCS estimate obtained from the entire, 10 min, trace containing $\approx 6 \times 10^6$ photon arrivals.

Detailed Methods Description

Description of Fluorescence Correlation Spectroscopy (FCS)

In FCS the primary quantity of interest is the spontaneously fluctuating fluorescence intensity [29, 91]. Correlations in fluorescence intensities are used to determine physical parameters such as diffusion coefficients. The normalized time autocorrelation function of the fluorescence intensity is defined as

$$G(\tau) = \frac{\langle \delta I(t) \delta I(t+\tau) \rangle}{\langle \delta I(t) \rangle^2} = \frac{\langle I(t) I(t+\tau) \rangle}{\langle I(t) \rangle^2} - 1,$$

where $I(t)$ is the fluorescence intensity, $\delta I(t)$ is intensity fluctuations at time t, and τ is the lag time. The intensity fluctuations of the fluorescence intensity are defined as the deviations from the average of the intensity, $\delta I(t) = I(t) - \langle I(t) \rangle$. For freely diffusing molecules in a 3D Gaussian confocal volume, the autocorrelation, which we use in this study, is

$$G(\tau) = \frac{1}{\langle N \rangle} \frac{1}{1 + \frac{4D\tau}{\omega_{xy}^2}} \frac{1}{\sqrt{1 + \frac{4D\tau}{\omega_z^2}}},$$

where $\langle N \rangle$ is the average number of molecules in the confocal volume, D is the diffusion coefficient, ω_{xy} and ω_z are the confocal volume axes along the xy and z directions. Further details on correlative analysis are contained in the cited literature [15, 29, 30, 35, 69, 91].

Explanation of Data Simulation

To generate synthetic traces we simulate molecules moving through a three dimensional illuminated volume. The number of moving molecules, N, is predefined in each simulation. We apply periodic boundaries to our volume of L_{xy} and L_z parallel to the focal plane and optical axis, respectively, to keep a relatively stable concentration of molecules near the confocal volume.

We denote the locations of the molecules as x_k^n, y_k^n and z_k^n, where k labels time levels and $n = 1, 2, \ldots, N$ labels molecules. The total trace duration $T_{total} = t_K - t_0$, is predefined. The time intervals between successive recorded photons $\Delta t_{k-1} = t_k - t_{k-1}$, are generated through pseudo-random computer simulations and recorded for subsequent analysis.

The locations of the molecules x_0^n, y_0^n, z_0^n at the first evaluation time t_0 are randomly sampled from the uniform distribution with borders identical to the boundaries $\pm L_{xy}$ and $\pm L_z$ of the prescribed simulation region. Locations x_k^n, y_k^n, z_k^n, for $k = 1, \ldots, K$, at times t_k are generated according to the diffusion model explained above under a predefined diffusion coefficient D.

We obtain photon inter-arrival times, $\boldsymbol{\Delta t} = (\Delta t_1, \Delta t_2, \ldots, \Delta t_{K-1})$, by simulating exponential random variables of rate μ_k. For independent background and molecule photon emission rates, the corresponding exponential emission mean rates μ_k depend on a Gaussian PSF as eqs. (2)–(3). Both background, μ_{back}, and the molecule photon emission rate, μ_{mol}, are predefined.

Definition of Molecule Photon Emission Rate

In this study the emission rate of detected photons for a single fluorophore at position x, y, z is used. This is formulated as the product $\mu(x, y, z) = \mu_0 \varphi_d \varphi_{de} \varphi_f \sigma \mathrm{EXC}(x, y, z) \mathrm{CEF}(x, y, z)$. Here, μ_0 and φ_d are the maximum excitation intensity and the efficiency of the photon collection at the center of the confocal volume, respectively, φ_{de} is the efficiency of the detector, φ_f is the quantum efficiency of the fluorophore, σ is the fluorophore absorption cross-section, $\mathrm{EXC}(x, y, z)$ is the excitation profile and $\mathrm{CEF}(x, y, z)$ is the detection profile [31]. By revising the definition of $\mu(x, y, z)$, we obtain $\mu(x, y, z) = \mu_{mol} \mathrm{PSF}(x, y, z)$ where $\mu_{mol} = \mu_0 \varphi_d \varphi_{de} \varphi_f \sigma$ and $\mathrm{PSF}(x, y, z) = \mathrm{EXC}(x, y, z) \mathrm{CEF}(x, y, z)$.

To relate our *single molecule* photon emission rate μ_{mol} to the average photon count rate typically determined in *bulk* experiments, we compute a spatial average

$$\begin{aligned}\langle \mu(x, y, z) \rangle &= \mu_{mol} \langle \mathrm{PSF}(x, y, z) \rangle = \mu_{mol} \left\langle \exp\left(-2 \frac{x^2 + y^2}{\omega_{xy}^2} - 2 \frac{z^2}{\omega_z^2} \right) \right\rangle \\ &= \mu_{mol} \frac{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp\left(-2 \frac{x^2}{\omega_{xy}^2} - 2 \frac{y^2}{\omega_{xy}^2} - 2 \frac{z^2}{\omega_z^2} \right) dx dy dz}{V} = \mu_{mol} \sqrt{\frac{\pi}{2} \omega_{xy}^2} \sqrt{\frac{\pi}{2} \omega_{xy}^2} \sqrt{\frac{\pi}{2} \omega_z^2} \frac{1}{V}\end{aligned}$$

where V denotes our PSF's effective volume [59, 85] which is equal to $V = \pi^{\frac{3}{2}} \omega_{xy}^2 \omega_z$. As a result, our molecule photon emission rate μ_{mol} is related to $\langle \mu(x, y, z) \rangle$ according to $\mu_{mol} = \sqrt{8}\, \langle \mu(x, y, z) \rangle$.

Description of Wilson-Hiferty Approximation

To perform the necessary computations of the next section, we use a Wilson-Hiferty transform [110] to approximate the probability density of exponential random variables. We use this approximation to sample the locations of the molecules within our overall Gibbs sampler (see next).

To apply the Wilson-Hiferty approximation, first we transform our observation random variable Δt_k to a new random variable ρ_k, where $\rho_k = 2\mu_k \Delta t_k$. A change of variables, indicates that $\Delta t_k|\mu_k \sim \text{Exponential}(\mu_k)$ implies $\rho_k|\mu_k \sim \chi^2(2)$, where $\chi^2(2)$ denotes the chi-square probability distribution with 2 degrees of freedom. By applying another transformation, where $\xi_k = \sqrt[3]{\rho_k/2}$, according to [110], ξ_k follows an *approximately* normal probability distribution $\xi_k \sim \text{Normal}\left(\frac{8}{9}, \frac{1}{9}\right)$. So, by $\xi_k = \sqrt[3]{\rho_k/2}$ and $\rho_k = 2\mu_k \Delta t_k$, we conclude $\sqrt[3]{\Delta t_k} = \xi_k / \sqrt[3]{\mu_k}$. Therefore, since $\sqrt[3]{\Delta t_k} = \xi_k / \sqrt[3]{\mu_k}$, we establish the approximation

$$\begin{aligned} p\left(\sqrt[3]{\Delta t_k}|\mu_k\right) &= p\left(\xi_k = \sqrt[3]{\Delta t_k}\sqrt[3]{\mu_k}\right)\sqrt[3]{\mu_k} \\ &\approx \sqrt[3]{\mu_k}\ \text{Normal}\left(\sqrt[3]{\Delta t_k}\sqrt[3]{\mu_k}; \frac{8}{9}, \frac{1}{9}\right) \\ &= \text{Normal}\left(\sqrt[3]{\Delta t_k}; \frac{8}{9\sqrt[3]{\mu_k}}, \left(\frac{1}{3\sqrt[3]{\mu_k}}\right)^2\right). \end{aligned}$$

Detailed Description of the Inference Framework

Prior Probability Distributions

Within the Bayesian paradigm, all unknown model parameters need priors. These parameters are: the diffusion coefficient D; the molecule photon emission rate μ_{mol}; the initial molecule locations x_1^n,y_1^n,z_1^n; as well as the indicator prior weights q^n.

Prior on the Diffusion Coefficient To make sure that D sampled in our formulation attains only positive values, we choose an Inverse-Gamma prior

$$D \sim \text{InvGamma}\left(\alpha_D, \beta_D\right). \tag{A1}$$

This prior is conjugate to the motion model which simplifies the computations shown below.

Priors on Molecule Photon Emission Rate To guarantee that μ_{mol} sampled in our formulation also attains only positive values, we choose a Gamma prior

$$\mu_{mol} \sim \text{Gamma}\left(\alpha_{mol}, \beta_{mol}\right). \tag{A2}$$

Priors on Initial Molecule Locations Because of the symmetries inherent to the confocal volume, *e.g.*, a molecule at a location (x, y, z) gives rise to the same photon emission rate as a molecule at location, $(-x, -y, -z)$, we use priors on the initial locations that respect these symmetries. To simplify the computations, we use independent symmetric normal distributions

$$x_1^n \sim \text{SymNormal}\left(\mu_{xy}, \sigma_{xy}^2\right), \tag{A3}$$

$$y_1^n \sim \text{SymNormal}\left(\mu_{xy}, \sigma_{xy}^2\right), \tag{A4}$$

$$z_1^n \sim \text{SymNormal}\left(\mu_z, \sigma_z^2\right). \tag{A5}$$

Priors and Hyperpriors for the Indicators To simplify the computations described in the next section, we use a finite, but large, model population of N molecules that contain contributing and noncontributing ones. These molecules are collectively indexed by $n = 1, 2, \ldots, N$. As described in the Methods section, inferring how many molecules are actually warranted by the data analyzed is the same as estimating how many of those N molecules are active, *i.e.*, $b^n = 1$, while the rest are inactive, *i.e.*, $b^n = 0$, and so have no influence and are applied just for computational reasons.

We use a Bernoulli prior of weight q^n to make sure that each indicator b^n takes only values 0 or 1. Moreover, on each weight q^n, we assign a beta hyperprior

$$b^n|q^n \sim \text{Bernoulli}\left(q^n\right), \tag{A6}$$

$$q^n \sim \text{Beta}\left(A_q, B_q\right). \tag{A7}$$

To make sure that the resulting formulation avoids overfitting, we make the specific selections $A_q = \alpha_q/N$ and $B_q = \beta_q(N-1)/N$. For these choices [1, 16, 77, 78], and in the limit that $N \to \infty$ (that is, when the assumed molecule population is large), this prior/hyperprior choice converges to a non-parametric beta-Bernoulli process. Therefore, for $N \gg 1$, the posterior is well defined and becomes *independent* of the selected value of N. In other words, provided N is large enough, its effect on the results is negligible; while its precise value has only computational implications.

Description of the Computational Implementation

Here, $p(D, \mu_{mol}, \{q^n, b^n, \overline{x}^n, \overline{y}^n, \overline{z}^n\}_n|\boldsymbol{\Delta t})$ is the joint probability distribution of our framework where molecular trajectories and measurements are gathered in

$$\begin{aligned} \overline{x}^n &= (x_1^n, x_2^n, \ldots, x_K^n), \\ \overline{y}^n &= (y_1^n, y_2^n, \ldots, y_K^n), \\ \overline{z}^n &= (z_1^n, z_2^n, \ldots, z_K^n), \\ \boldsymbol{\Delta t} &= (\Delta t_1, \Delta t_2, \ldots, \Delta t_{K-1}). \end{aligned}$$

Posterior samples are generated according to Gibbs sampling [37, 66, 86, 100, 107]. We achieve this by sampling a variable conditioned on all other variables and the given photon inter-arrival times $\boldsymbol{\Delta t}$. Conceptually, the steps in the generation of each posterior sample $(D, \mu_{mol}, \{q^n, b^n, \overline{x}^n, \overline{y}^n, \overline{z}^n\}_n)$ are:

1. For each n of the *active* molecules
   (a) Update trajectory $\overline{x}^n$ of active molecule n

(b) Update trajectory $\overline{y}^n$ of active molecule n

(c) Update trajectory $\overline{z}^n$ of active molecule n

2. Update jointly the trajectories $\overline{x}^n, \overline{y}^n, \overline{z}^n$ for all n of the *inactive* molecules
3. Update the diffusion coefficient D
4. Update jointly the prior weights q^n for all model molecules and simultaneously update jointly the indicators b^n for all model molecules
5. Update the molecule photon emission rate μ_{mol}

Sampling Active Molecules Locations To sample the location of an active molecule $(\overline{x}^n, \overline{y}^n, \overline{z}^n)$, we use forward filtering and backward sampling [11, 20, 53, 94]. In particular, we update each dimension sequentially from the following full conditional probability distributions $p(\overline{x}^n|D, \mu_{mol}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \boldsymbol{\Delta t})$, $p(\overline{y}^n|D, \mu_{mol}, \{b^{n'}, \overline{x}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{y}^{n'}\}_{n'\neq n}, \boldsymbol{\Delta t})$, and $p(\overline{z}^n|D, \mu_{mol}, \{b^{n'}, \overline{x}^{n'}, \overline{y}^{n'}\}_{n'}, \{\overline{z}^{n'}\}_{n'\neq n}, \boldsymbol{\Delta t})$. Below, we show in detail the calculation only for sampling $\overline{x}^n$, since for sampling $\overline{y}^n$ and $\overline{z}^n$ they are similar.

To sample the trajectory $\overline{x}^n$, we rely on the factorization

$$\begin{aligned}
&p(\overline{x}^n|D, \mu_{mol}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \boldsymbol{\Delta t}) \\
&= p(x_K^n|D, \mu_{mol}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \boldsymbol{\Delta t}) \\
&\times p(x_{K-1}^n|x_K^n, D, \mu_{mol}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \boldsymbol{\Delta t}) \\
&\times \cdots \\
&\times p(x_2^n|x_1^n, D, \mu_{mol}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \boldsymbol{\Delta t}) \\
&\times p(x_1^n|D, \mu_{mol}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \boldsymbol{\Delta t})
\end{aligned}$$

and, according to this factorization, we sample individual locations x_k^n sequentially

$$\begin{aligned}
x_K^n &\sim p(x_K^n|D, \mu_{mol}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \boldsymbol{\Delta t}) \\
x_k^n &\sim p(x_k^n|x_{k+1}^n, D, \mu_{mol}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \boldsymbol{\Delta t}),
\end{aligned}$$

where, $k = 1, \ldots, K-1$. However, to be able to perform these steps, we first need to compute the involved probability distributions. We describe below a computationally efficient way to do so that proceeds in a forward filtering and a backward sampling step.

Before we start the sampling of the locations, we determine each one of the individual probability distributions that are needed. To do this in a computationally tractable manner [14, 20], we compute filter distributions $p(x_k^n|D, \mu_{mol}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \{\Delta t_{k'}\}_{k'<k})$. In our case, both dynamic (eqs. (7)–(9))) and observation (eq. (1)) probability distributions provide equal probabilities for $+x_k^n$ and $-x_k^n$. Therefore, the filter distribution consists of two modes symmetrically placed across the origin [53]. Accordingly, we compute an *approximate* bimodal symmetric filter of the form

$$\begin{aligned}
&p\left(x_k^n|D, \mu_{mol}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \{\Delta t_{k'}\}_{k'<k}\right) \\
&\approx \text{SymNormal}\left(x_k^n; m_k^n, c_k^n\right)
\end{aligned}$$

where $\text{SymNormal}\left(m_k^n, c_k^n\right)$ describes the symmetric normal distribution. The filter, that is the values of m_k^n and c_k^n, is updated iteratively according to

$$\begin{aligned}
&p\left(x_k^n|D, \mu_{mol}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \{\Delta t_{k'}\}_{k'<k}\right) \\
&\propto p\left(\Delta t_{k-1}|x_k^n, y_k^n, z_k^n, \mu_{mol}, \{b^{n'}, x^{n'}, y^{n'}, z^{n'}\}_n'\right) \\
&\times \int_{x_{k-1}^n} p\left(x_{k-1}^n|D, \mu_{mol}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \{\Delta t_{k'}\}_{k'<k-1}\right) p\left(x_k^n|x_{k-1}^n, D\right) dx_{k-1}^n. \qquad \text{(A8)}
\end{aligned}$$

To be able to carry out these computations efficiently, similar to [53], we work on an approximate model where the exponential emission equation, eq. (1), is replaced by a normal one using the Wilson-Hiferty approximation as we discussed earlier. Our *approximate* emission equation is

$$\begin{aligned}
&T_{data}\left(\Delta t_k\right)|\{x_k^n, y_k^n, z_k^n, b^n\}_n, \mu_{mol} \\
&\sim \text{Normal}\left(T_{mean}(\mu_k), S(\mu_k)^2\right), \quad k = 1, \ldots, K-1.
\end{aligned}$$

where μ_k is given in eq. (2); while $T_{data}\left(\Delta t_k\right)$, $T_{mean}(\mu_k)$ and $S^2(\mu_k)$ are given by the Wilson-Hiferty approximation [110]. As explained earlier, the approximation is given by $T_{data}(\Delta t_k) = \Delta t_k^{1/3}$, $T_{mean}(\mu_k) = 8/(9\mu_k^{1/3})$, and $S^2(\mu_k) = 1/(9\mu_k^{2/3})$. Because of the specific choices of our problem (*i.e.,* diffusive molecules, symmetric normal filter at the proceeding time, and normal likelihood), eq. (A8) reduces to

$$\begin{aligned}
&p\left(x_k^n|D, \mu_{mol}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \{\Delta t_{k'}\}_{k'<k}\right) \\
&= \text{Normal}\left(T_{data}(\Delta t_k); T_{mean}(\mu_k), S(\mu_k)^2\right) \\
&\times \text{SymNormal}\left(x_k^n; m_{k-1}^n, c_{k-1}^n + 2D\Delta t_k\right). \qquad \text{(A9)}
\end{aligned}$$

Finally, to obtain the values of m_k^n and c_k^n, we linearize the product in eq. (A9) as described next. From eq. (A9),

we have

$$\begin{aligned}
&\text{Normal}\left(T_{data}(\Delta t_k); T_{mean}(\mu_k), S(\mu_k)^2\right) \times \text{SymNormal}\left(x_k^n; m_{k-1}^n, c_{k-1}^n + 2D\Delta t_k\right) \\
&\propto \exp\left(\frac{\log\mu(x_k^n)}{3} - \frac{\left(\frac{8}{9} - \sqrt[3]{\mu(x_k^n)\Delta t_{k-1}}\right)^2}{\frac{2}{9}} - \frac{(x_k^n - m_{k-1}^n)^2}{2(c_{k-1}^n + 2D\Delta t_{k-1})}\right) \\
&+ \exp\left(\frac{\log\mu(x_k^n)}{3} - \frac{\left(\frac{8}{9} - \sqrt[3]{\mu(x_k^n)\Delta t_{k-1}}\right)^2}{\frac{2}{9}} - \frac{(x_k^n + m_{k-1}^n)^2}{2(c_{k-1}^n + 2D\Delta t_{k-1})}\right).
\end{aligned} \tag{A10}$$

The density in eq. (A10) consists of two modes, one on the positive semi-axis of x and one on the negative semi-axis of x. Considering $f(x_k^n) = \sqrt[3]{\mu(x_k^n)\Delta t_{k-1}}$ and $g(x_k^n) = \frac{1}{3}\log\mu(x_k^n)$ and linearizing them around the previous filter's mode, $+m_{k-1}^n$ or $-m_{k-1}^n$, the modes of eq. (A10) are approximated by

$$\begin{aligned}
\exp\left(\frac{\log\mu(x_k^n)}{3} - \frac{\left(\frac{8}{9} - \sqrt[3]{\mu(x_k^n)\Delta t_{k-1}}\right)^2}{\frac{2}{9}} - \frac{(x_k^n - m_{k-1}^n)^2}{2(c_{k-1}^n + 2D\Delta t_{k-1})}\right) &\approx \exp\left(v_1 x_k^n - \frac{(x_k^n - h_1)^2}{\frac{2\sigma^2}{f'(-m_{k-1}^n)^2}} - \frac{(x_k^n - m_{k-1}^n)^2}{2(c_{k-1}^n + 2D\Delta t_{k-1})}\right), \\
\exp\left(\frac{\log\mu(x_k^n)}{3} - \frac{\left(\frac{8}{9} - \sqrt[3]{\mu(x_k^n)\Delta t_{k-1}}\right)^2}{\frac{2}{9}} - \frac{(x_k^n + m_{k-1}^n)^2}{2(c_{k-1}^n + 2D\Delta t_{k-1})}\right) &\approx \exp\left(v_2 x_k^n - \frac{(x_k^n - h_2)^2}{\frac{2\sigma^2}{f'(-m_{k-1}^n)^2}} - \frac{(x_k^n + m_{k-1}^n)^2}{2(c_{k-1}^n + 2D\Delta t_{k-1})}\right).
\end{aligned}$$

Combining both approximations, the density of eq. (A10), is approximated by

$$\begin{aligned}
&\exp\left(\frac{\log\mu(x_k^n)}{3} - \frac{\left(\frac{8}{9} - \sqrt[3]{\mu(x_k^n)\Delta t_{k-1}}\right)^2}{\frac{2}{9}} - \frac{(x_k^n - m_{k-1}^n)^2}{2(c_{k-1}^n + 2D\Delta t_{k-1})}\right) \\
&+ \exp\left(\frac{\log\mu(x_k^n)}{3} - \frac{\left(\frac{8}{9} - \sqrt[3]{\mu(x_k^n)\Delta t_{k-1}}\right)^2}{\frac{2}{9}} - \frac{(x_k^n + m_{k-1}^n)^2}{2(c_{k-1}^n + 2D\Delta t_{k-1})}\right) \\
\approx &\exp\left(v_1 x_k^n - \frac{(x_k^n - h_1)^2}{\frac{2\sigma^2}{f'(+m_{k-1}^n)^2}} - \frac{(x_k^n - m_{k-1}^n)^2}{2(c_{n-1}^k + 2D\Delta t_{k-1})}\right) + \exp\left(v_2 x_k^n - \frac{(x_k^n - h_2)^2}{\frac{2\sigma^2}{f'(-m_{k-1}^n)^2}} - \frac{(x_k^n + m_{k-1}^n)^2}{2(c_{k-1}^n + 2D\Delta t_{k-1})}\right)
\end{aligned} \tag{A11}$$

where $h_1 = \frac{\frac{8}{9} - f(+m_{k-1}^n) + m_{k-1}^n f'(+m_{k-1}^n)}{f'(+m_{k-1}^n)}$, $v_1 = \frac{\mu'(+m_{k-1}^n)}{3\mu(+m_{k-1}^n)}$, and $h_2 = \frac{\frac{8}{9} - f(-m_{k-1}^n) - m_{k-1}^n f'(-m_{k-1}^n)}{f'(-m_{k-1}^n)}$, $v_2 = \frac{\mu'(-m_{k-1}^n)}{3\mu(-m_{k-1}^n)}$.

Equation (A11) describes a symmetric normal distribution. Equating this distribution with our filter, *i.e.*, $\text{SymNormal}(m_k^n, c_k^n)$, we obtain $c_k^n = \left(\frac{1}{c_{k-1}^n + 2D\Delta t_{k-1}} + \frac{9}{2} f'(m_{k-1}^n)^2\right)^{-1}$, and $m_k^n = \left(v_1 + 9h_1 f'(m_{k-1}^n)^2 + \frac{m_{k-1}^n}{c_{k-1}^n + 2D\Delta t_{k-1}}\right) c_k^n$. These apply for $k = 2, \ldots, K$ and are used to update the filter. To begin, we use eq. (A3)–(A5) and set $c_1^n = \sigma_{xy}^2$ and $m_1^n = \mu_{xy}$.

Having computed the filter distributions above, we are able to sample the individual locations by starting from x_K^n and moving backward towards x_1^n. In particular, by applying the Bayes' rule, each one of the individual distributions factorize as

$$p(x_k^n|x_{k+1}^n, D, \mu_{mol}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \boldsymbol{\Delta t}) \propto \tag{A12}$$
$$p\left(x_k^n|D, \mu_{mol}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \{\Delta t_{k'}\}_{k'<k}\right) \times p\left(x_k^n|x_{k+1}^n, D\right).$$

The first term is given by the filter distribution which is replaced by our approximate SymNormal(m_k^n, c_k^n), and the second term is our motion model Normal$(x_k^n, 2D\Delta t_{k-1})$, all of which are known at this stage. Therefore, backward sampling starts at x_K^n and continues for x_{k-1}^n with

$$\begin{aligned}
x_K^n &\sim \frac{1}{2}\text{Normal}\left(+m_K^n, c_K^n\right) + \frac{1}{2}\text{Normal}\left(-m_K^n, c_K^n\right) \\
x_{k-1}^n &\sim \frac{\exp\left[\frac{(x_k^n - m_k^n)^2}{c_k^n + 2D\Delta t_{k-1}}\right]}{2\sqrt{2\pi(c_k^n + 2D\Delta t_{k-1})}}\text{Normal}\left(\frac{x_k^n c_k^n + 2D\Delta t_{k-1} m_k^n}{c_k^n + 2D\Delta t_{k-1}}, \frac{2D\Delta t_{k-1} c_k^n}{c_k^n + 2D\Delta t_{k-1}}\right) \\
&+ \frac{\exp\left[\frac{(x_k^n + m_k^n)^2}{c_k^n + 2D\Delta t_{k-1}}\right]}{2\sqrt{2\pi(c_k^n + 2D\Delta t_{k-1})}}\text{Normal}\left(\frac{x_k^n c_k^n - 2D\Delta t_{k-1} m_k^n}{c_k^n + 2D\Delta t_{k-1}}, \frac{2D\Delta t_{k-1} c_k^n}{c_k^n + 2D\Delta t_{k-1}}\right).
\end{aligned}$$

Sampling Inactive Molecule Trajectories To update the trajectories of the inactive molecules, we sample from the corresponding conditionals $p(\overline{x}^n, \overline{y}^n, \overline{z}^n|D, \mu_{mol}, \{q^n, b^n\}_n, \boldsymbol{\Delta t})$. Since the inactive molecules are not associated with the observations in $\boldsymbol{\Delta t}$, these reduce to $p(\overline{x}^n, \overline{y}^n, \overline{z}^n|D, \{q^n, b^n\}_n)$ which we simulate as standard 3D Brownian motion [64].

Sampling the Diffusion Coefficient We sample the diffusion coefficient from the conditional probability distribution $p(D|\mu_{mol}, \{q^n, b^n, \overline{x}^n, \overline{y}^n, \overline{z}^n\}_n, \boldsymbol{\Delta t})$. Because of the specific dependencies of the variables in this formulation, *e.g.*, eq. (A1) and eqs. (7)–(9), the conditional distribution simplifies to $p(D|\{\overline{x}^n, \overline{y}^n, \overline{z}^n\}_n, \boldsymbol{\Delta t})$. Using Bayes' rule, this distribution becomes $D \sim \text{InvGamma}(\alpha'_D, \beta'_D)$ where $\alpha'_D = \alpha_D + \frac{3N(K-1)}{2}$ and $\beta'_D = \beta_D + \sum_{k=2}^K \frac{\sum_{n=1}^N \left(\left(x_k^n - x_{k-1}^n\right)^2 + \left(y_k^n - y_{k-1}^n\right)^2 + \left(z_k^n - z_{k-1}^n\right)^2\right)}{4\Delta t_{k-1}}$.

Sampling Molecule Indicators For each molecule n we sample its indicator prior weight, q^n, from the corresponding conditional distribution $p(q^n|\mu_{mol}, D, \{b^n, \overline{x}^n, \overline{y}^n, \overline{z}^n\}_n, \boldsymbol{\Delta t})$, which simplifies to $p(q^n|b^n)$. For this we use eq. (A7) and eq. (A6). According to Bayes' rule, the latter distribution becomes $q^n \sim \text{Beta}(\alpha', \beta')$ where $\alpha' = \frac{\alpha}{N} + b^n$ and $\beta' = \beta\frac{N-1}{N} + 1 - b^n$. Subsequently, we update the indicators b^n by sampling from the corresponding conditional distribution $p(\{b^n\}_n|D, \mu_{mol}, \{q^n, \overline{x}^n, \overline{y}^n, \overline{z}^n\}_n, \boldsymbol{\Delta t})$ using a Methropolis-Hasting algorithm [18, 23]. For this, we use a proposal $b_{new}^n \sim \text{Bernoulli}(q^n)$. With this proposal, the acceptance ratio becomes

$$r = \prod_{k=1}^{K-1} \frac{\mu_{back} + \mu_{mol}\sum_{n=1}^N b_{new}^n \exp\left(-2\frac{(x_k^n)^2 + (y_k^n)^2}{\omega_{xy}^2} - 2\frac{(z_k^n)^2}{\omega_z^2}\right)}{\mu_{back} + \mu_{mol}\sum_{n=1}^N b_{old}^n \exp\left(-2\frac{(x_k^n)^2 + (y_k^n)^2}{\omega_{xy}^2} - 2\frac{(z_k^n)^2}{\omega_z^2}\right)} \frac{e^{-\left[\mu_{back} + \mu_{mol}\sum_{n=1}^N b_{new}^n \exp\left(-2\frac{(x_k^n)^2 + (y_k^n)^2}{\omega_{xy}^2} - 2\frac{(z_k^n)^2}{\omega_z^2}\right)\right]\Delta t_k}}{e^{-\left[\mu_{back} + \mu_{mol}\sum_{n=1}^N b_{old}^n \exp\left(-2\frac{(x_k^n)^2 + (y_k^n)^2}{\omega_{xy}^2} - 2\frac{(z_k^n)^2}{\omega_z^2}\right)\right]\Delta t_k}}.$$

Sampling the Molecule Photon Emission Rate In the last step, after updating the locations and indicators of the molecules, we sample the molecule photon emission rate from the corresponding conditional distribution $p(\mu_{mol}|D, \{q^n, b^n, \overline{x}^n, \overline{y}^n, \overline{z}^n\}_n, \boldsymbol{\Delta t})$. To sample this distribution, we also use a Metropolis-Hastings step. For this, we use proposals of the form $\mu_{mol}^{new} \sim \text{Gamma}\left(\alpha, \mu_{mol}^{old}/\alpha\right)$ where μ_{mol}^{old} denotes the current sampled value. Using both eqs. (1) and (2), the acceptance ratio becomes

$$
\begin{aligned}
r = \exp &\left[\sum_{k=1}^{K-1} \log\left(\frac{\mu_{back} + \mu_{mol}^{new} \sum_{n=1}^{N} b^n \exp\left(-2\frac{(x_k^n)^2+(y_k^n)^2}{\omega_{xy}^2} - 2\frac{(z_k^n)^2}{\omega_z^2}\right)}{\mu_{back} + \mu_{mol}^{old} \sum_{n=1}^{N} b^n \exp\left(-2\frac{(x_k^n)^2+(y_k^n)^2}{\omega_{xy}^2} - 2\frac{(z_k^n)^2}{\omega_z^2}\right)}\right)\right] \\
&\times \exp\left[\sum_{k=1}^{K-1}\left[(\mu_{mol}^{old} - \mu_{mol}^{new})\Delta t_k \sum_{n=1}^{N} b^n \exp\left(-2\frac{(x_k^n)^2 + (y_k^n)^2}{\omega_{xy}^2} - 2\frac{(z_k^n)^2}{\omega_z^2}\right)\right]\right] \\
&\times \exp\left[(\alpha_{mol} - 1)\log\left(\frac{\mu_{mol}^{new}}{\mu_{mol}^{old}}\right) + \frac{1}{\beta_{mol}}(\mu_{mol}^{old} - \mu_{mol}^{new}) + (2\alpha - 1)\log\left(\frac{\mu_{mol}^{old}}{\mu_{mol}^{new}}\right)\right] \times \exp\left[\alpha\left(\frac{\mu_{mol}^{new}}{\mu_{mol}^{old}} - \frac{\mu_{mol}^{old}}{\mu_{mol}^{new}}\right)\right].
\end{aligned}
$$

TABLE I. Here, we list point estimates of our analyses, which we obtain from the marginal posterior probability distributions $p(D|\boldsymbol{\Delta t})$ and $p(\mu_{mol}|\boldsymbol{\Delta t})$. Estimates are listed according to figure.

	D		μ_{mol}	
	mean	std	mean	std
	$\mu m^2/s$	$\mu m^2/s$	photons/s	photons/s
Fig. (2A)	4.54	4.49	-	-
Fig. (2B)	4.17	4.11	-	-
Fig. (2C)	1.14	1.12	-	-
Fig. (2D)	1.02	1.01	-	-
Fig. (2E)	4.75	4.64	-	-
Fig. (4B1)	1.03	0.25	-	-
Fig. (4B2)	0.95	0.63	-	-
Fig. (4B3)	0.75	0.68	-	-
Fig. (4B4)	0.45	0.77	-	-
Fig. (5A3)	1.01	0.27	-	-
Fig. (5B3)	1.09	0.51	-	-
Fig. (5C3)	1.65	1.59	-	-
Fig. (6)	1.05×10^{-2}	0.22×10^{-2}	-	-
	1.21×10^{-1}	0.34×10^{-1}	-	-
	1.06	0.19	-	-
	9.87	2.33	-	-
	117.62	35.13	-	-
Fig. (7A)	0.99	0.34	-	-
Fig. (7B)	0.96	0.51	-	-
Fig. (7C)	2.92	2.68	-	-
Fig. (7D)	3.26	2.95	-	-
Fig. (A3A3)	10.02	1.17	-	-
Fig. (A3B3)	9.96	2.19	-	-
Fig. (A4A3)	-	-	4.11×10^5	1.61×10^3
Fig. (A4B3)	-	-	4.37×10^4	2.84×10^3
Fig. (A4C3)	-	-	1.28×10^5	1.25×10^4

TABLE II. Summary of notation.

Description	Variable	Units
Diffusion coefficient	D	μm^2/s
α parameter of the diffusion coefficient prior	α_D	-
β parameter of the diffusion coefficient prior	β_D	μm^2/s
Photon inter-arrival time	Δt	s
Total trace duration	T_{total}	s
molecule photon emission rate (maximum)	μ_{mol}	photons/s
α parameter of the molecule photon emission rate's prior	α_{mol}	-
β parameter of the molecule photon emission rate's prior	β_{mol}	photons/s
Emission rate of molecule n at time t_k	μ_k^n	photons/s
Combined photon emission rate at time t_k	μ_k	photons/s
Background photon emission rate	μ_{back}	photons/s
Minor semi-axis of confocal PSF (focal plane)	ω_{xy}	μm
Major semi-axis of confocal PSF (optical axis)	ω_z	μm
Location of molecule n at time t_k in x-coordinate	x_k^n	μm
Location of molecule n at time t_k in y-coordinate	y_k^n	μm
Location of molecule n at time t_k in z-coordinate	z_k^n	μm
Recorded photon inter-arrival time between t_k and t_{k-1}	Δt_k	s
Indicator variable for molecule n	b^n	-
Prior weight for b_n	q^n	-
α parameter of prior weight q^n	α_q	-
β parameter of prior weight q^n	β_q	-
Upper bound for the number of model molecules	N	-
Mean value of initial molecule position's prior in the xy-plane	μ_{xy}	μm
Mean value of initial molecule position's prior on the z-axis	μ_z	μm
Variance of the initial molecule position's prior in the xy-plane	σ_{xy}^2	μm
Variance of the initial molecule position's prior on the z-axis	σ_z^2	μm
Periodic boundary in the xy-plane	L_{xy}	μm
Periodic boundary on the z-axis	L_z	μm

TABLE III. Probability distributions used and their densities. Here, the corresponding random variables are denoted by x. We use ";" to separate random variables from parameters. For example, Normal$(x;\mu,\sigma^2)$ means that x is the random variable (e.g. $\int_{-\infty}^{+\infty} dx\text{Normal}(x;\mu,\sigma^2)=1$), and μ and σ^2 are parameters characterizing this density.

Distribution	Notation	Probability density function	Mean	Variance
Normal	Normal(μ,σ^2)	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Symmetric Normal	SymNormal(μ,σ^2)	$\frac{1}{2}\frac{e^{-\frac{(x+\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}+\frac{1}{2}\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$	0	$\mu^2+\sigma^2$
Exponential	Exponential(μ)	$\mu e^{-\mu x}$	$\frac{1}{\mu}$	$\frac{1}{\mu^2}$
Chi-square	$\chi^2(\alpha,2)$	$\frac{1}{\Gamma(\frac{\alpha}{2})2^{\frac{\alpha}{2}}}x^{\frac{\alpha}{2}-1}e^{-\frac{x}{2}}$	α	2α
Gamma	Gamma(α,β)	$\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-\frac{x}{\beta}}$	$\alpha\beta$	$\alpha\beta^2$
Inverse-Gamma	InvGamma(α,β)	$\frac{\beta^\alpha}{\Gamma(\alpha)}x^{-\alpha-1}e^{-\frac{\beta}{x}}$	$\frac{\beta}{\alpha-1}$	$\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$
Beta	Beta(α,β)	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Bernoulli	Bernoulli(q)	$(q-1)\delta_0(x)+q\delta_1(x)$	q	$q(1-q)$

TABLE IV. Parameter values used in the generation of the synthetic traces. Choices are listed according to figures.

	L_{xy}	L_z	ω_{xy}	ω_z	N	D	μ_{mol}	μ_{back}	T_{total}
Units	μm	μm	μm	μm	-	$\mu m^2/s$	photons/s	photons/s	s
Fig. (2A)	1	2	0.3	1.5	4	1	4×10^4	10^3	0.03
Fig. (2B)	1	2	0.3	1.5	4	1	4×10^4	10^3	0.03
Fig. (2C)	1	2	0.3	1.5	4	1	4×10^4	10^3	0.03
Fig. (2D)	1	2	0.3	1.5	4	1	4×10^4	10^3	0.03
Fig. (2E)	1	2	0.3	1.5	4	1	4×10^4	10^3	0.03
Fig. (4A)	1	2	0.3	1.5	4	1	4×10^4	10^3	0.03
Fig. (5A)	1	2	0.3	1.5	10	1	4×10^4	10^3	0.03
Fig. (5C)	1	2	0.3	1.5	4	1	4×10^4	10^3	0.03
Fig. (5E)	1	2	0.3	1.5	1	1	4×10^4	10^3	0.03
Fig. (6)	1	2	0.3	1.5	4	10^{-2}	4×10^4	10^3	0.03
Fig. (6)	1	2	0.3	1.5	4	10^{-1}	4×10^4	10^3	0.03
Fig. (6)	1	2	0.3	1.5	4	1	4×10^4	10^3	0.03
Fig. (6)	1	2	0.3	1.5	4	10	4×10^4	10^3	0.03
Fig. (6)	1	2	0.3	1.5	4	100	4×10^4	10^3	0.03
Fig. (7A)	1	2	0.3	1.5	4	1	4×10^5	10^3	0.03
Fig. (7B)	1	2	0.3	1.5	4	1	4×10^4	10^3	0.03
Fig. (7C)	1	2	0.3	1.5	4	1	4×10^3	10^3	0.03
Fig. (7D)	1	2	0.3	1.5	4	1	10^3	10^3	0.03
Fig. (A3A)	1	2	0.3	1.5	10	10	4×10^5	10^3	0.05
Fig. (A3C)	1	2	0.3	1.5	10	10	4×10^4	10^3	0.05
Fig. (A4A)	1	2	0.3	1.5	10	10	4×10^5	10^3	0.05
Fig. (A4C)	1	2	0.3	1.5	10	10	4×10^4	10^3	0.05
Fig. (A4E)	1	2	0.3	1.5	10	10	4×10^3	10^3	0.05

TABLE V. Parameter values used in the analyses of the traces. Choices are listed according to figures.

	ω_{xy}	ω_z	N	α_D	β_D	α_{mol}	β_{mol}	α_q	β_q	μ_{xy}	μ_z	σ^2_{xy}	σ^2_z
Units	μm	μm	-	-	$\mu m^2/s$	-	phts/s	-	-	-	μm	μm	μm^2
Fig. (2A)	0.3	1.5	-	1	1	1	10^5	-	-	0.1	0.1	1	1
Fig. (2B)	0.3	1.5	-	1	1	1	10^5	-	-	0.1	0.1	1	1
Fig. (2C)	0.3	1.5	-	1	1	1	10^5	-	-	0.1	0.1	1	1
Fig. (2D)	0.3	1.5	-	1	1	1	10^5	-	-	0.1	0.1	1	1
Fig. (2E)	0.3	1.5	-	1	1	1	10^5	-	-	0.1	0.1	1	1
Fig. (4)	0.3	1.5	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (5A3)	0.3	1.5	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (5B3)	0.3	1.5	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (5C3)	0.3	1.5	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (6)	0.3	1.5	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (6)	0.3	1.5	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (6)	0.3	1.5	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (6)	0.3	1.5	20	1	100	1	10^5	1	1	0.1	0.1	1	1
Fig. (7A)	0.3	1.5	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (7B)	0.3	1.5	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (7C)	0.3	1.5	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (7D)	0.3	1.5	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (8)	0.23	0.55	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (9)	0.23	0.55	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (10)	0.23	0.55	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (11)	0.23	0.55	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (12)	0.27	4.51	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (13)	0.22	3.90	20	1	100	1	10^5	1	1	0.1	0.1	1	1
Fig. (A3)	0.3	1.5	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (A4)	0.3	1.5	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (A5)	0.23	0.55	20	1	1	1	10^5	1	1	0.1	0.1	1	1
Fig. (A6)	0.22	3.90	20	1	100	1	10^5	1	1	0.1	0.1	1	1

[1] Luai Al Labadi and Mahmoud Zarepour. On approximations of the beta process in latent feature models: Point processes approach. *Sankhya A*, 80(1):59–79, 2018.

[2] Michael Andrec, Ronald M Levy, and David S Talaga. Direct determination of kinetic rates from single-molecule photon arrival trajectories using hidden markov models. *The Journal of Physical Chemistry A*, 107(38):7454–7464, 2003.

[3] JO Andrews, W Conway, W-K Cho, A Narayanan, J-H Spille, N Jayanth, T Inoue, S Mullen, J Thaler, and II Cissé. qSR: A quantitative super-resolution analysis tool reveals the cell-cycle dependent organization of RNA polymerase i in live human cells. *Scientific Reports*, 8(1):7424, 2018.

[4] Matthew Antonik, Suren Felekyan, Alexander Gaiduk, and Claus AM Seidel. Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *The Journal of Physical Chemistry B*, 110(13):6970–6978, 2006.

[5] Daniel Axelrod, DE Koppel, J Schlessinger, Ei Elson, and Watt W Webb. Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophysical Journal*, 16(9):1055–1069, 1976.

[6] Andre C Barato and Udo Seifert. Cost and precision of brownian clocks. *Physical Review X*, 6(4):041053, 2016.

[7] A Benda, M Beneš, V Marecek, A Lhotskỳ, W Th Hermens, and M Hof. How to determine diffusion coefficients in planar phospholipid systems by confocal fluorescence correlation spectroscopy. *Langmuir*, 19(10): 4120–4126, 2003.

[8] Howard C Berg. *Random walks in biology*. Princeton University Press, 1993.

[9] Andrew J Berglund. Statistics of camera-based single-particle tracking. *Physical Review E*, 82(1):011917, 2010.

[10] Robert B Best and Gerhard Hummer. Reaction coordinates and rates from transition paths. *Proceedings of the National Academy of Sciences*, 102(19):6732–6737, 2005.

[11] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[12] Max Born and Emil Wolf. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013.

[13] GJ Brakenhoff, K Visscher, and HTM Van der Voort. Size and shape of the confocal spot: control and relation to 3D imaging and image processing. In *Handbook of biological confocal microscopy*, pages 87–91. Springer, 1990.

[14] Mark Briers, Arnaud Doucet, and Simon Maskell. Smoothing algorithms for state–space models. *Annals of the Institute of Statistical Mathematics*, 62(1):61, 2010.

[15] Gary R Bright, Gregory W Fisher, Jadwiga Rogowska, and D Lansing Taylor. Fluorescence ratio imaging microscopy. *Methods in Cell Biology*, 30:157–192, 1989.

[16] Tamara Broderick, Michael I Jordan, Jim Pitman, et al. Beta processes, stick-breaking and power laws. *Bayesian Analysis*, 7(2):439–476, 2012.

[17] Jonathan E Bronson, Jingyi Fei, Jake M Hofman, Ruben L Gonzalez Jr, and Chris H Wiggins. Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. *Biophysical Journal*, 97(12):3196–3205, 2009.

[18] Ben Calderhead. A general construction for parallelizing metropolis- hastings algorithms. *Proceedings of the National Academy of Sciences*, 111(49):17408–17413, 2014.

[19] Christopher P Calderon and Kerry Bloom. Inferring latent states and refining force estimates via hierarchical dirichlet process modeling in single particle tracking experiments. *PloS one*, 10(9):e0137633, 2015.

[20] Olivier Cappé, Eric Moulines, and Tobias Rydén. Inference in hidden markov models. In *Proceedings of EUSFLAT Conference*, pages 14–16, 2009.

[21] Aleksei V Chechkin, Flavio Seno, Ralf Metzler, and Igor M Sokolov. Brownian yet non-gaussian diffusion: from superstatistics to subordination of diffusing diffusivities. *Physical Review X*, 7(2):021002, 2017.

[22] Yan Chen, Joachim D Müller, Peter TC So, and Enrico Gratton. The photon counting histogram in fluorescence fluctuation spectroscopy. *Biophysical Journal*, 77(1):553–567, 1999.

[23] Siddhartha Chib and Edward Greenberg. Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335, 1995.

[24] Daniel T Chiu, Noo Li Jeon, Sui Huang, Ravi S Kane, Christopher J Wargo, Insung S Choi, Donald E Ingber, and George M Whitesides. Patterned deposition of cells and proteins onto surfaces by using three-dimensional microfluidic systems. *Proceedings of the National Academy of Sciences*, 97(6):2408–2413, 2000.

[25] Won-Ki Cho, Jan-Hendrik Spille, Micca Hecht, Choongman Lee, Charles Li, Valentin Grube, and Ibrahim I Cisse. Mediator and rna polymerase II clusters associate in transcription-dependent condensates. *Science*, 361(6400):412–415, 2018.

[26] Hoi Sung Chung and Irina V Gopich. Fast single-molecule FRET spectroscopy: theory and experiment. *Physical Chemistry Chemical Physics*, 16(35):18644–18657, 2014.

[27] Hoi Sung Chung, Fanjie Meng, Jae-Yeol Kim, Kevin McHale, Irina V Gopich, and John M Louis. Oligomerization of the tetramerization domain of p53 probed by two-and three-color single-molecule FRET. *Proceedings of the National Academy of Sciences*, page 201700357, 2017.

[28] Petra Dittrich, Flaminia Malvezzi-Campeggi, Michael Jahnz, and Petra Schwille. Accessing molecular dynamics in cells by fluorescence correlation spectroscopy. *Biological Chemistry*, 382(3):491–494, 2001.

[29] Elliot L Elson. Fluorescence correlation spectroscopy: past, present, future. *Biophysical Journal*, 101(12): 2855–2870, 2011.

[30] Elliot L Elson and Douglas Magde. Fluorescence correlation spectroscopy. I. conceptual basis and theory. *Biopolymers*, 13(1):1–27, 1974.

[31] Jörg Enderlein and W Patrick Ambrose. Optical collection efficiency function in single-molecule detection experiments. *Applied Optics*, 36(22):5298–5302, 1997.

[32] Jorg Enderlein, Ingo Gregor, Digambara Patra, and Jorg Fitter. Art and artefacts of fluorescence correlation

spectroscopy. *Current Pharmaceutical Biotechnology*, 5 (2):155–161, 2004.

[33] Radek Erban and S Jonathan Chapman. Stochastic modelling of reaction–diffusion processes: algorithms for bimolecular reactions. *Physical Biology*, 6(4):046001, 2009.

[34] Thomas S Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, pages 209–230, 1973.

[35] James AJ Fitzpatrick and Björn F Lillemeier. Fluorescence correlation spectroscopy: linking molecular dynamics to biological function in vitro and in situ. *Current Opinion in Structural Biology*, 21(5):650–660, 2011.

[36] Andreas Gahlmann and WE Moerner. Exploring bacterial cell biology with single-molecule tracking and superresolution imaging. *Nature Reviews Microbiology*, 12(1): 9, 2014.

[37] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*, volume 2. CRC press Boca Raton, FL, 2014.

[38] Sarah Frisken Gibson and Frederick Lanni. Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy. *JOSA A*, 9(1):154–166, 1992.

[39] Irina V Gopich. Accuracy of maximum likelihood estimates of a two-state model in single-molecule FRET. *The Journal of Chemical Physics*, 142(3):034110, 2015.

[40] Irina V Gopich and Attila Szabo. Single-molecule FRET with diffusion and conformational dynamics. *The Journal of Physical Chemistry B*, 111(44):12925–12932, 2007.

[41] Irina V Gopich and Attila Szabo. Decoding the pattern of photon colors in single-molecule FRET. *The Journal of Physical Chemistry B*, 113(31):10965–10973, 2009.

[42] Irina V Gopich and Attila Szabo. Theory of the energy transfer efficiency and fluorescence lifetime distribution in single-molecule FRET. *Proceedings of the National Academy of Sciences*, 109(20):7747–7752, 2012.

[43] Taekjip Ha, Alice Y Ting, Joy Liang, W Brett Caldwell, Ashok A Deniz, Daniel S Chemla, Peter G Schultz, and Shimon Weiss. Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism. *Proceedings of the National Academy of Sciences*, 96(3):893–898, 1999.

[44] JM Haile, Ian Johnston, A John Mallinckrodt, Susan McKay, et al. Molecular dynamics simulation: elementary methods. *Computers in Physics*, 7(6):625–625, 1993.

[45] Marta Hajdziona and Andrzej Molski. Maximum likelihood-based analysis of single-molecule photon arrival trajectories. *The Journal of Chemical Physics*, 134 (5):054112, 2011.

[46] Jun He, Syuan-Ming Guo, and Mark Bathe. Bayesian approach to the analysis of fluorescence correlation spectroscopy data i: theory. *Analytical Chemistry*, 84(9): 3871–3879, 2012.

[47] Desmond J Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43(3):525–546, 2001.

[48] Keegan E Hines. A primer on Bayesian inference for biophysical systems. *Biophysical Journal*, 108(9):2103–2113, 2015.

[49] Keegan E Hines, John R Bankston, and Richard W Aldrich. Analyzing single-molecule time series via nonparametric Bayesian inference. *Biophysical Journal*, 108 (3):540–556, 2015.

[50] Nils Lid Hjort, Chris Holmes, Peter Müller, and Stephen G Walker. *Bayesian nonparametrics*, volume 28. Cambridge University Press, 2010.

[51] Bo Huang, Mark Bates, and Xiaowei Zhuang. Superresolution fluorescence microscopy. *Annual Review of Biochemistry*, 78:993–1016, 2009.

[52] Oliver C Ibe. *Elements of Random Walk and Diffusion Processes*. John Wiley & Sons, 2013.

[53] Sina Jazani, Ioannis Sgouralis, and Steve Pressé. A method for single molecule tracking using a conventional single-focus confocal setup. *The Journal of Chemical Physics*, 150(11):114108, 2019.

[54] Sina Jazani, Ioannis Sgouralis, Omer M Shafraz, Marcia Levitus, Sanjeevi Sivasankar, and Steve Pressé. An alternative framework for fluorescence correlation spectroscopy. *Nature Communications*, 10, 2019.

[55] Margaret E Johnson and Gerhard Hummer. Free-propagator reweighting integrator for single-particle dynamics in reaction-diffusion models of heterogeneous protein-protein interaction systems. *Physical Review X*, 4(3):031037, 2014.

[56] Peet Kask, Rolf Günther, and Peter Axhausen. Statistical accuracy in fluorescence fluctuation experiments. *European Biophysics Journal*, 25(3):163–169, 1997.

[57] Oleg Krichevsky and Grégoire Bonnet. Fluorescence correlation spectroscopy: the technique and its applications. *Reports on Progress in Physics*, 65(2):251, 2002.

[58] Lucas Lacasa, Inés P Mariño, Joaquin Miguez, Vincenzo Nicosia, Édgar Roldán, Ana Lisica, Stephan W Grill, and Jesús Gómez-Gardeñes. Multiplex decomposition of non-markovian dynamics and the hidden layer reconstruction problem. *Physical Review X*, 8(3):031038, 2018.

[59] Joseph R Lakowicz. *Principles of fluorescence spectroscopy*. Springer, 2006.

[60] Antony Lee, Konstantinos Tsekouras, Christopher Calderon, Carlos Bustamante, and Steve Pressé. Unraveling the thousand word picture: An introduction to super-resolution data analysis. *Chemical Reviews*, 2017.

[61] Sang-Hyuk Lee, Jae Yen Shin, Antony Lee, and Carlos Bustamante. Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM). *Proceedings of the National Academy of Sciences*, 109(43):17436–17441, 2012.

[62] Eitan Lerner, Antonino Ingargiola, and Shimon Weiss. Characterizing highly dynamic conformational states: the transcription bubble in RNAP-promoter open complex as an example. *The Journal of Chemical Physics*, 148(12):123315, 2018.

[63] Hui Li, Chi-Fu Yen, and Sanjeevi Sivasankar. Fluorescence axial localization with nanometer accuracy and precision. *Nano Letters*, 12(7):3731–3735, 2012.

[64] Lawrence C-L Lin and Frank LH Brown. Brownian dynamics in fourier space: membrane simulations over long length and time scales. *Physical Review Letters*, 93 (25):256001, 2004.

[65] Hongchao Liu, Bo Jiu, Hongwei Liu, and Zheng Bao. Superresolution isar imaging based on sparse bayesian learning. *IEEE Transactions on Geoscience and Remote Sensing*, 52(8):5005–5013, 2014.

[66] Huan Liu and Hiroshi Motoda. *Computational methods of feature selection*. CRC Press, 2007.

[67] Zhe Liu, Luke D Lavis, and Eric Betzig. Imaging live-cell dynamics and structure at the single-molecule level. *Molecular cell*, 58(4):644–659, 2015.

[68] Radek Macháň and Martin Hof. Lipid diffusion in planar membranes investigated by fluorescence correlation spectroscopy. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1798(7):1377–1391, 2010.

[69] Douglas Magde, Elliot L Elson, and Watt W Webb. Fluorescence correlation spectroscopy. II. an experimental realization. *Biopolymers*, 13(1):29–61, 1974.

[70] Valentin Magidson and Alexey Khodjakov. Circumventing photodamage in live-cell microscopy. In *Methods in cell biology*, volume 114, pages 545–560. Elsevier, 2013.

[71] Sean P Meyn and Richard L Tweedie. Stability of markovian processes II: Continuous-time processes and sampled chains. *Advances in Applied Probability*, 25(3): 487–517, 1993.

[72] Sean P Meyn and Richard L Tweedie. Stability of markovian processes III: Foster–lyapunov criteria for continuous-time processes. *Advances in Applied Probability*, 25(3):518–548, 1993.

[73] X Michalet, OHW Siegmund, JV Vallerga, P Jelinsky, JE Millaud, and S Weiss. Detectors for single-molecule fluorescence imaging and spectroscopy. *Journal of Modern Optics*, 54(2-3):239–281, 2007.

[74] Arjun Narayanan, Anatoli B Meriin, Michael Y Sherman, and Ibrahim I Cisse. A first order phase transition underlies the formation of sub-diffractive protein aggregates in mammalian cells. *bioRxiv*, page 148395, 2017.

[75] Eyal Nir, Xavier Michalet, Kambiz M Hamadani, Ted A Laurence, Daniel Neuhauser, Yevgeniy Kovchegov, and Shimon Weiss. Shot-noise limited single-molecule fret histograms: comparison between theory and experiments. *The Journal of Physical Chemistry B*, 110(44): 22103–22124, 2006.

[76] Kenji Okamoto and Yasushi Sako. Variational bayes analysis of a photon-based hidden markov model for single-molecule FRET trajectories. *Biophysical Journal*, 103(6):1315–1324, 2012.

[77] John Paisley and Lawrence Carin. Nonparametric factor analysis with beta process priors. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 777–784. ACM, 2009.

[78] John Paisley and Michael I Jordan. A constructive definition of the beta process. *arXiv preprint arXiv:1604.00685*, 2016.

[79] Konstantina Palla, David A Knowles, and Zoubin Ghahramani. A reversible infinite HMM using normalised random measures. In *International Conference on Machine Learning*, 2014.

[80] Robert D Phair and Tom Misteli. Kinetic modelling approaches to in vivo imaging. *Nature Reviews Molecular Cell Biology*, 2(12):898, 2001.

[81] Rob Phillips, Julie Theriot, Jane Kondev, and Hernan Garcia. *Physical biology of the cell*. Garland Science, 2012.

[82] Menahem Pirchi, Roman Tsukanov, Rashid Khamis, Toma E Tomov, Yaron Berger, Dinesh C Khara, Hadas Volkov, Gilad Haran, and Eyal Nir. Photon-by-photon hidden markov model analysis for microsecond single-molecule FRET kinetics. *The Journal of Physical Chemistry B*, 120(51):13065–13075, 2016.

[83] Martin Purschke, Noemi Rubio, Kathryn D Held, and Robert W Redmond. Phototoxicity of hoechst 33342 in time-lapse fluorescence microscopy. *Photochemical & Photobiological Sciences*, 9(12):1634–1639, 2010.

[84] Katrien Remaut, Bart Lucas, Kevin Braeckmans, NN Sanders, SC De Smedt, and Jo Demeester. FRET-FCS as a tool to evaluate the stability of oligonucleotide drugs after intracellular delivery. *Journal of Controlled Release*, 103(1):259–271, 2005.

[85] Rudolf Rigler and Elliot S Elson. *Fluorescence correlation spectroscopy: theory and applications*, volume 65. Springer Science & Business Media, 2012.

[86] Christian Robert and George Casella. *Introducing Monte Carlo Methods with R*. Springer Science & Business Media, 2009.

[87] Saveez Saffarian and Elliot L Elson. Statistical analysis of fluorescence correlation spectroscopy: the standard deviation and bias. *Biophysical Journal*, 84(3):2030–2042, 2003.

[88] Dag Scherfeld, Nicoletta Kahya, and Petra Schwille. Lipid dynamics and domain formation in model membranes composed of ternary mixtures of unsaturated and saturated phosphatidylcholines and cholesterol. *Biophysical Journal*, 85(6):3758–3768, 2003.

[89] Patrick D Schmidt, Benjamin H Reichert, John G Lajoie, and Sanjeevi Sivasankar. Method for high frequency tracking and sub-nm sample stabilization in single molecule fluorescence microscopy. *Scientific Reports*, 8(1):13912, 2018.

[90] Benjamin Schuler. Perspective: Chain dynamics of unfolded and intrinsically disordered proteins from nanosecond fluorescence correlation spectroscopy combined with single-molecule FRET. *The Journal of Chemical Physics*, 149(1):010901, 2018.

[91] Petra Schwille and Elke Haustein. Fluorescence correlation spectroscopy: an introduction to its concepts and applications. *Biophysics textbook online*, 1(3), 2001.

[92] Petra Schwille, Franz-Josef Meyer-Almes, and Rudolf Rigler. Dual-color fluorescence cross-correlation spectroscopy for multicomponent diffusional analysis in solution. *Biophysical Journal*, 72(4):1878–1886, 1997.

[93] Petra Schwille, Ulrich Haupts, Sudipta Maiti, and Watt W Webb. Molecular dynamics in living cells observed by fluorescence correlation spectroscopy with one-and two-photon excitation. *Biophysical Journal*, 77 (4):2251–2265, 1999.

[94] Steven L Scott. Bayesian methods for hidden markov models: Recursive computing in the 21st century. *Journal of the American Statistical Association*, 97(457): 337–351, 2002.

[95] Ioannis Sgouralis and Steve Pressé. ICON: an adaptation of infinite hmms for time traces with drift. *Biophysical Journal*, 112(10):2117–2126, 2017.

[96] Ioannis Sgouralis and Steve Pressé. An introduction to infinite HMMs for single-molecule data analysis. *Biophysical Journal*, 112(10):2021–2029, 2017.

[97] Ioannis Sgouralis, Shreya Madaan, Franky Djutanta, Rachael Kha, Rizal F Hariadi, and Steve Pressé. A bayesian nonparametric approach to single molecule forster resonance energy transfer. *The Journal of Physical Chemistry B*, 123(3):675–688, 2018.

[98] Ioannis Sgouralis, Miles Whitmore, Lisa Lapidus, Matthew J Comstock, and Steve Pressé. Single molecule

force spectroscopy at high data acquisition: A bayesian nonparametric analysis. *The Journal of Chemical Physics*, 148(12):123320, 2018.

[99] Jesse Stricker, Paul Maddox, ED Salmon, and Harold P Erickson. Rapid assembly dynamics of the escherichia coli FtsZ-ring demonstrated by fluorescence recovery after photobleaching. *Proceedings of the National Academy of Sciences*, 99(5):3171–3175, 2002.

[100] Meysam Tavakoli, J Nicholas Taylor, Chun-Biu Li, Tamiki Komatsuzaki, and Steve Pressé. Single molecule data analysis: an introduction. *arXiv preprint arXiv:1606.00403*, 2016.

[101] Meysam Tavakoli, Konstantinos Tsekouras, Richard Day, Kenneth W Dunn, and Steve Presse. Quantitative kinetic models from intravital microcopy: A case study using hepatic transport. *The Journal of Physical Chemistry B*, 2019.

[102] Yee W Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Sharing clusters among related groups: Hierarchical dirichlet processes. In *Advances in neural information processing systems*, pages 1385–1392, 2005.

[103] Jean-Yves Tinevez, Joe Dragavon, Lamya Baba-Aissa, Pascal Roux, Emmanuelle Perret, Astrid Canivet, Vincent Galy, and Spencer Shorte. A quantitative method for measuring phototoxicity of a live cell imaging microscope. In *Methods in enzymology*, volume 506, pages 291–309. Elsevier, 2012.

[104] Tedman Torres and Marcia Levitus. Measuring conformational dynamics: a new FCS-FRET approach. *The Journal of Physical Chemistry B*, 111(25):7392–7400, 2007.

[105] Konstantinos Tsekouras, Amanda P Siegel, Richard N Day, and Steve Pressé. Inferring diffusion dynamics from FCS in heterogeneous nuclear environments. *Biophysical Journal*, 109(1):7–17, 2015.

[106] Stephan Uphoff, Kristofer Gryte, Geraint Evans, and Achillefs N Kapanidis. Improved temporal resolution and linked hidden markov modeling for switchable single-molecule FRET. *ChemPhysChem*, 12(3):571–579, 2011.

[107] Udo Von Toussaint. Bayesian inference in physics. *Reviews of Modern Physics*, 83(3):943, 2011.

[108] Marta Waligórska and Andrzej Molski. Maximum likelihood-based analysis of photon arrival trajectories in single-molecule FRET. *Chemical Physics*, 403:52–58, 2012.

[109] Shimon Weiss. Fluorescence spectroscopy of single biomolecules. *Science*, 283(5408):1676–1683, 1999.

[110] Edwin B Wilson and Margaret M Hilferty. The distribution of chi-square. *Proceedings of the National Academy of Sciences*, 17(12):684–688, 1931.

[111] Thorsten Wohland, Rudolf Rigler, and Horst Vogel. The standard deviation in fluorescence correlation spectroscopy. *Biophysical Journal*, 80(6):2987–2999, 2001.

[112] Guanghua Wu, Haifeng Ji, Karolyn Hansen, Thomas Thundat, Ram Datar, Richard Cote, Michael F Hagan, Arup K Chakraborty, and Arunava Majumdar. Origin of nanomechanical cantilever motion generated from biomolecular interactions. *Proceedings of the National Academy of Sciences*, 98(4):1560–1564, 2001.

[113] Shiwei Wu, Gang Han, Delia J Milliron, Shaul Aloni, Virginia Altoe, Dmitri V Talapin, Bruce E Cohen, and P James Schuck. Non-blinking and photostable upconverted luminescence from single lanthanide-doped nanocrystals. *Proceedings of the National Academy of Sciences*, 106(27):10917–10921, 2009.

[114] Bo Zhang, Josiane Zerubia, and Jean-Christophe Olivo-Marin. Gaussian approximations of fluorescence microscope point-spread function models. *Applied Optics*, 46 (10):1819–1829, 2007.