An Alternative Framework for Fluorescence Correlation Spectroscopy

Sina Jazani,[1,2] Ioannis Sgouralis,[1,2] Omer M. Shafraz,[3] Marcia Levitus,[1,4,5] Sanjeevi Sivasankar,[3] and Steve Pressé[1,2,4,a)]

[1)] *Center for Biological Physics, Arizona State University,Tempe, AZ 85287*
[2)] *Department of Physics, Arizona State University, Tempe, AZ 85287*
[3)] *Department of Biomedical Engineering, University of California, Davis, CA 95616*
[4)] *School of Molecular Sciences, Arizona State University, Tempe, AZ 85287*
[5)] *Biodesign Institute, Arizona State University, Tempe, AZ 85287*

ABSTRACT

Fluorescence correlation spectroscopy (FCS), is a flexible and widely used tool routinely exploited for *in vivo* and *in vitro* applications. While FCS provides estimates of dynamical quantities, such as diffusion coefficients, it demands high signal to noise ratios and long time traces, typically in the minute range. In principle, the same information can be extracted from μ-s long time traces; however, an appropriate analysis method is missing. To overcome these limitations, we adapt novel tools inspired by Bayesian non-parametrics, which starts from the direct analysis of the observed photon counts. With this approach, we are able to analyze time traces, which are too short to be analyzed by existing methods, including FCS. Our new analysis extends the capability of single molecule fluorescence confocal microscopy based approaches, to probe processes several orders of magnitude faster in time and permits a reduction of phototoxic effects on living samples induced by long periods of light exposure.

INTRODUCTION

Due to its flexibility and limited invasiveness for *in vivo* applications, single molecule fluorescence confocal microscopy[1–4] is one of the most widely used experimental techniques of modern Biophysics. In this technique, fluorescent molecules are allowed to freely diffuse into a volume illuminated by a tightly focused laser beam of a conventional single-focus confocal setup. Molecular motion inside the illuminated volume generates fluctuations in the emitted fluorescence that is recorded and subsequently temporally autocorrelated[1–4] or, jointly spatiotemporally autocorrelated[5–7], to deduce physical quantities of interest. For example, fluorescence correlation spectroscopy (FCS)[1,2] as well as complementary methods – such as Fluorescence Cross-Correlation Spectroscopy (FCCS)[8], scanning FCS[9,10], spot variation Fluorescence Correlation Spectroscopy (svFCS)[11], Fluorescence Resonance Energy Transfer-Fluorescence Correlation Spectroscopy (FCS-FRET)[12,13], etc – estimate diffusion coefficients, reaction kinetic, binding affinities and concentrations of labeled molecules[14,15].

While single molecule fluorescence confocal microscopy data is acquired on the micro- to millisecond timescales (μs-ms), fluorescence correlation methods typically require the analysis of long time traces, several seconds to tens of minutes long, depending on the molecular concentrations and emission properties of the fluorophores employed[16,17]. These traces, capturing multiple molecule traversals of the confocal volume, provide the statistics needed for the post-processing steps used in traditional FCS analysis[16] (e.g. autocorrelation, and nonlinear fitting to theoretical curves). However, processing steps like these downgrade the quality of the available data and demand either relatively high concentrations or excessively long time traces to yield reliable estimates. The same downgrades are encountered even with less traditional FCS analyses, including Bayesian approaches[18–22], that also rely on auto-correlations.

In principle, within milliseconds, for the fluorophore concentrations and confocal volumes used in most experiments[1,2,23], thousands of data points are already available. Accordingly, if one could, somehow, estimate diffusion coefficients within tens of ms with the same accuracy as FCS, one could hypothetically use tens of minutes worth of data to discriminate between very small differences in diffusion coefficients. Alternatively, one could opt for shorter traces in the first place and, in doing so, reduce the sample's light exposure to only a few ms, thereby minimizing photo-toxic effects which remain a severe limitation of fluorescence microscopy[24–26].

Exploiting data on ms timescales would require a method that, simultaneously and self-consistently, estimates the number of fluorescent molecules at any given time within the (inhomogenously) illuminated volume and deduce their dynamical properties based on their photon emissions, which, in turn, depend on their evolving location within the confocal volume. The mathematics to do so in a rigorous and efficient manner have, so far, been unavailable as analyzing ms traces would demand that we consider all possible populations of molecules responsible for the observed traces, their diffusion coefficients, and every possible location (and, thus, photon emission rate) of those molecules at any given time.

Indeed, with current technology, this global optimiza-

[a)] Email: spresse@asu.edu

tion is prohibitively computationally expensive. To wit, maximum likelihood approaches[15,27], popular in a variety of applications, are excluded since they require that the, otherwise unknown, population of molecules in the confocal volume at any given time be specified in advance by other means. These considerations motivate an entirely new framework for FCS.

Here we introduce a novel approach that exploits Bayesian non-parametrics[15,28,29], a branch of Statistics first suggested[30] in 1973 and only broadly popularized in physical applications over the last few years[15,28,29,31–37]. This approach allows us to account for an arbitrary number of molecules responsible for emitting detected photons. With the proposed method, we are able to estimate physical variables, otherwise determined from FCS, with: (i) significantly shorter time traces; and (ii) nearly single molecule resolution. Furthermore, our overall framework is generalizable, and can estimate not only diffusion coefficients and molecular populations but also track molecules through time as well as determine their molecular brightness and the background photon emission rate.

RESULTS

The method we propose for the analysis of traces from single molecule fluorescence confocal microscopy follows the Bayesian paradigm[15,27,29,38]. Within this paradigm, our goal is to estimate posterior probability distributions over unknown parameters such as diffusion coefficients as well as molecular populations over time.

In this section, we first demonstrate and validate our method by computing posterior distributions using synthetic (simulated) traces mimicking the properties of real single molecule fluorescence confocal experiments. We subsequently benchmark our estimates with traces from control *in vitro* experiments.

A. Demonstration and validation with simulated data

To demonstrate the robustness of our method, we simulate fluorescent time traces under a broad range of: (i) numbers of labeled molecules in the effective volume, Fig. 1; (ii) diffusion coefficients, Fig. 2a; (iii) trace lengths, Fig. 2b; and (iv) molecular brightness, Fig. 3. Since, the majority of our time traces are too short to be meaningfully analyzed with traditional FCS, we compare our posteriors directly to the ground truth that we used in the simulations.

The posteriors we obtain, in all figures, are informed from the analysis of a single trace. In those, the breadth of the posterior (i.e. variance), which is a measure of the accuracy of our estimate, therefore indicates the uncertainty introduced by the finiteness of the data and the inherent noise in this single time trace.

To begin, in Fig. 1 we simulate a 3D Gaussian confocal volume of size ($\omega_{xy} = 0.3\ \mu m$ and $\omega_z = 1.5\ \mu m$) and

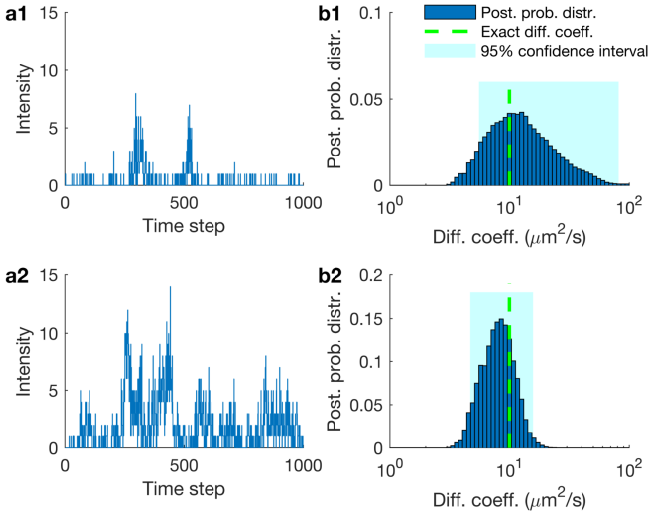

FIG. 1. **Testing the effects of the number of molecules inside the confocal volume on diffusion coefficient estimates**. **(a1)** Synthetic fluorescent intensity trace produced by 1 molecule inside the confocal volume. For this simulation we used a molecular brightness of 5×10^4 photons/s and a background photon emission rate of 10^3 photons/s. **(b1)** Posterior probability distribution over the diffusion coefficient estimated from the trace in (a1). **(a2)** Synthetic fluorescent intensity trace produced by 5 molecules inside the confocal volume otherwise identical to (a1). **(b2)** Posterior probability distribution over the diffusion coefficient estimated from (a2). Traces shown in (a1) and (a2) are acquired at 100 μs for a total of 100 ms and the highlighted regions in (b1) and (b2) represent the 95% confidence intervals. For clarity, the horizontal axis is shown in logarithmic scale.

1 molecule inside the effective volume (Fig. 1a1) or 5 molecules inside the effective volume (Fig. 1a2) diffusing at 10 μm^2/s for a total period of 100 ms.

As can be seen in Fig. 1, a low intensity leads to a wide estimate of the diffusion coefficient. However, the higher the intensity, the sharper (i.e. more conclusive) the estimate of the diffusion coefficient becomes (e.g. note a narrower posterior in Fig. 1b2 as compared to Fig. 1b1). Thus, diffusion coefficients are determined more accurately when the number of labeled molecules are higher. Accordingly, the most difficult data to analyze are those where concentrations of molecules are so low that, on average, only one molecule ventures, albeit rarely, into the effective region of the confocal volume where it can be appreciably excited. Put differently, for an equal length time trace, the posterior estimate over the diffusion coefficient is broader (i.e. less conclusive) for lower numbers of molecules inside the effective volume, Fig. 1b1, than it is for larger numbers of molecules, Fig. 1b2.

Following a similar reasoning, the slower a molecule diffuses, the more photons are collected, leading to a sharper posterior estimate of the corresponding diffusion coefficient, Fig. 2a. Likewise, the longer the trace is, Fig. 2b, or the greater the molecular brightness is, Fig. 3, the sharper the diffusion coefficient estimate becomes.

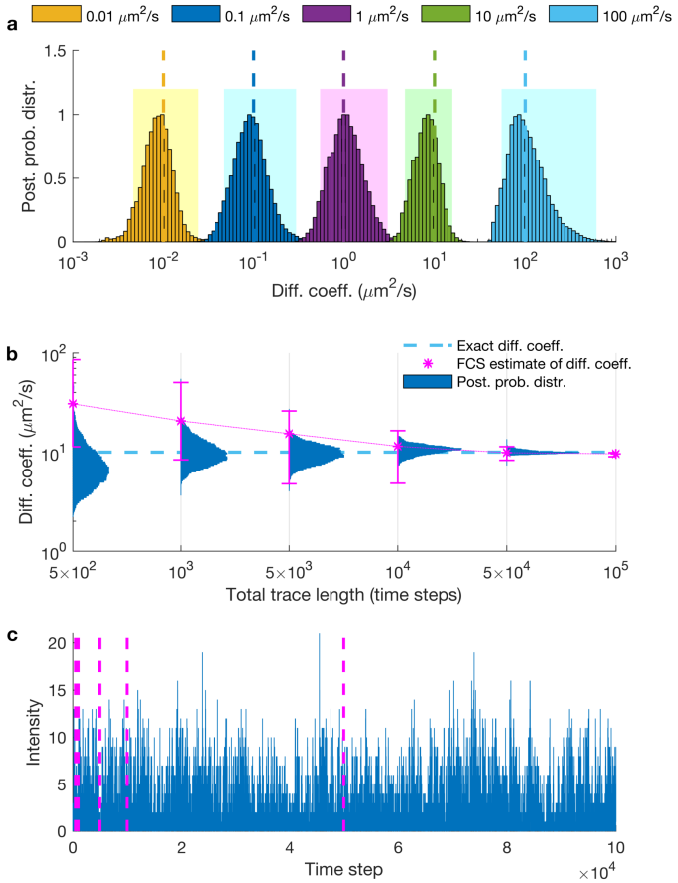

FIG. 2. **Testing the effects of diffusion coefficient and trace length on diffusion coefficient estimates**. **(a)** Posterior probability distributions deduced from traces produced from molecules with diffusion coefficients of 0.01, 0.1, 1, 10 and 100 $\mu m^2/s$. For clarity, posteriors are normalized to maximum 1 and the horizontal axis is shown in logarithmic scale. Shaded regions illustrate the 95% confidence intervals. **(b)** Posterior probability distributions deduced from traces acquired at 100 μs with total trace lengths of 5×10^2, 1×10^3, 5×10^3, 1×10^4, 5×10^4 and 1×10^5 time steps. For comparison, exact values and FCS estimates are also shown and, for clarity, the vertical axis is shown in logarithmic scale. As can be seen from (b), it is typical for FCS to require $\approx$1000$\times$ more data than our method before estimating diffusion coefficients within 2$\times$ of the correct value. **(c)** The entire trace used to deduce diffusion coefficients in (a) and (b). Each segment marked by dashed-lines represents the portion used in (b). The molecular brightness and background photon emission rates used to generate the time traces used are 5×10^4 and 10^3 photons/s.

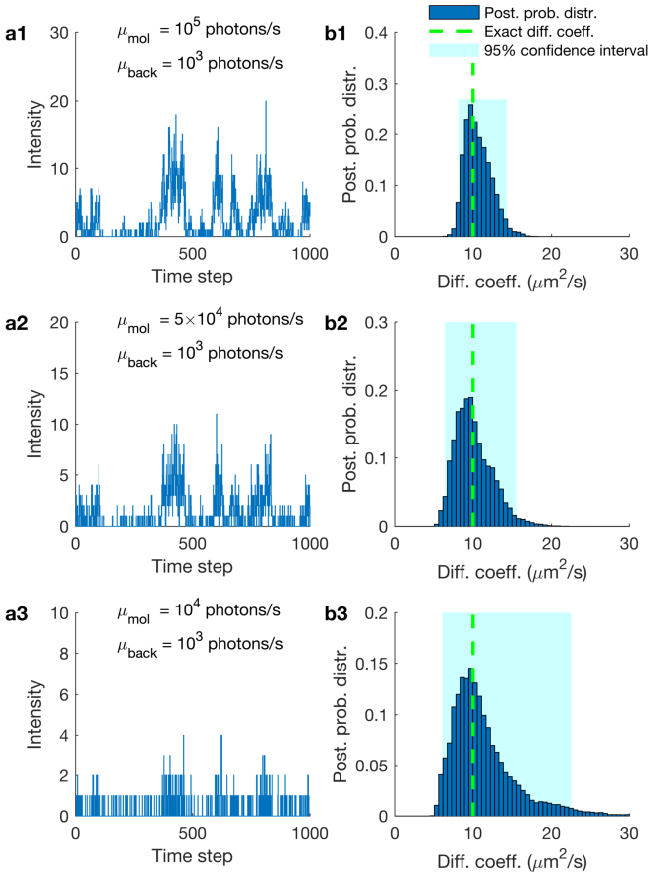

FIG. 3. **Testing the effects of molecular brightness on diffusion coefficient estimates**. **(a)** Intensity traces produced by the same molecular trajectories under a molecular brightness of 10^5, 5×10^4 and 1×10^4 photons/s and background photon emission rate fixed at 10^3 photons/s. **(b)** Posterior probability distributions and exact values of diffusion coefficients obtained from the corresponding traces. Shaded regions illustrate the 95% confidence intervals.

We emphasize that our definition of molecular brightness is based on the the maximum number of detected photons emitted from a single fluorophore when located at the center of the confocal volume and we provide more details in the SUPPLEMENTARY MATERIALS.

In Fig. 3 we demonstrate the robustness of the diffusion coefficient estimates when varying the molecular brightness. While we keep the background photon emission rate fixed, we simulate gradually dimmer fluorophores such as those encountered in an experiment under lower laser powers, until the molecular signature is virtually lost. As can be seen, such traces lead to broader posterior estimates over diffusion coefficients, as one would expect, since these traces are associated with greater uncertainty. Also, as such traces lead to a weaker (i.e. less constraining) likelihood, the posterior resembles more closely the prior (not shown) and naturally starts to deviate from the exact value.

B. Benchmarking on experimental data with elongated confocal volume shapes

Here we apply our method on experimental traces captured with an elongated confocal volume that we approximate by a cylinder. To do so, we apply our method on fluorescent beads (with average diameter of 45 nm) diffusing in water. We benchmark our estimated diffusion coefficients against the Stokes-Einstein prediction

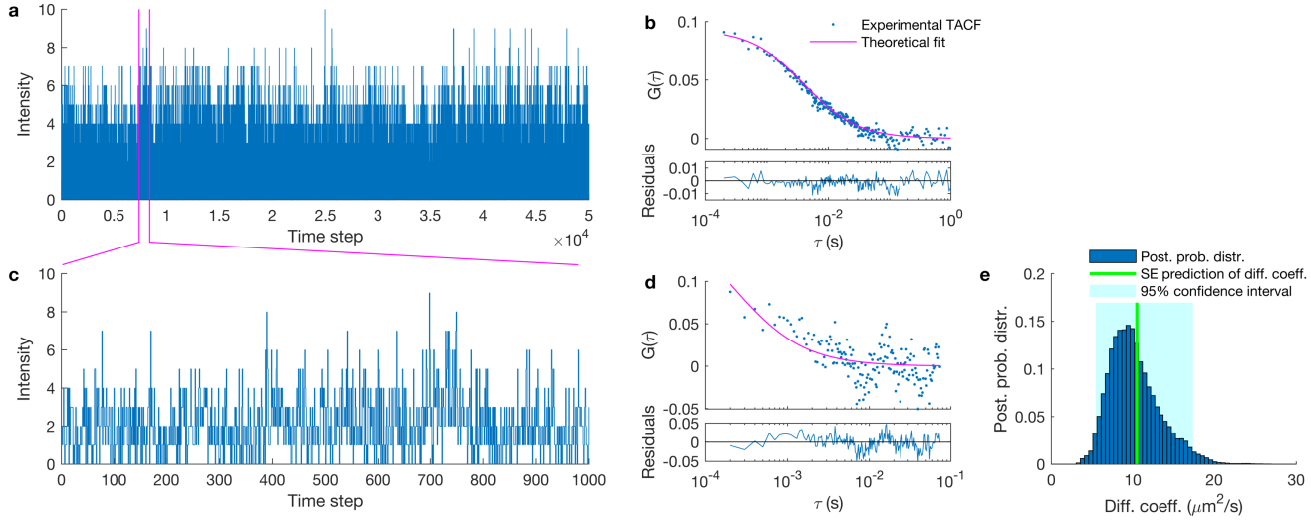

FIG. 4. **Diffusion coefficient estimates from experimental traces of free fluorescent beads using an elongated confocal volume. (a)** Experimental fluorescent intensity trace used in FCS. **(b)** Auto-correlation curve of the trace in (a) and best theoretical fit. **(c)** Portion of the trace in (a) to be used as the input to FCS and our method. **(d)** Auto-correlation curve of trace in (c). **(e)** Posterior probability distribution over diffusion coefficient estimated from the trace in (c). Stokes-Einstein prediction are denoted by a green line. Traces shown in (a) and (c) are acquired in bins of 100 μs for a total period of 5 seconds and 0.1 second respectively. The laser power used to generate the trace (a) is 100 μW (measured before the beam enters the objective). The estimate of diffusion coefficient resulting by autocorrelation fitting in (a) matches with the Stokes-Einstein prediction (i.e. 10.5 $\mu m^2/s$); while in (d) is almost ten fold higher ($\sim$100 $\mu m^2/s$).

and results from FCS. In particular, Fig 4 illustrates our method's performance in the analysis of traces too short to be meaningfully analyzed by FCS. Both the precise FCS formulation used here as well as additional results of our method on traces generated with free Cy3B dyes in glycerol/water mixtures with 70% glycerol can be found in the SUPPLEMENTARY MATERIALS.

C. Benchmarking on experimental data with elliptical confocal volume shapes

1. Fluorescent dyes

Next, we apply our method on experimental time traces derived from single molecule fluorescence confocal microscopy. In our setup, we monitor Cy3 dyes which diffuse freely in a mixture of water and glycerol. We benchmark our estimated diffusion coefficients against two values: those predicted by the Stokes-Einstein formula[40], which is parametrized by physical quantities such as temperature and viscosity; and those estimated by FCS. To analyze the data using FCS, each time we used the full (6 min) trace available. This is by contrast to our the estimates provided by our method which are obtained on traces $\approx$1000$\times$ shorter (100 ms).

In benchmarking, we obtained and analyzed measurements with different: (i) numbers of Cy3 dyes inside the effective volume (tuned by varying Cy3 concentrations); (ii) diffusion coefficients (tuned by adjusting the viscosity of the solution); and (iii) molecular brightness (tuned by adjusting the laser power). For example, in Fig. 5b1-b4, we illustrate the effect of different dye concentrations where a trace with stronger signal, anticipated when concentrations are higher, is expected to lead to better diffusion coefficient estimates (and thus sharper posteriors) on traces of equal length due to the higher number of labeled-molecules inside the confocal volume. Consistent with the synthetic data shown earlier, we obtain a broader posterior over diffusion coefficients when the number of dyes inside the effective volume is low and sharper posteriors for higher numbers of dyes.

Just as before, the slower a molecule diffuses, the more time it spends in the vicinity of the confocal volume, so the more photons are collected, thereby leading to sharper posterior estimates for the diffusion coefficient; see Fig. 5a.

A posterior's sharpness depends strongly on the number of molecules in a time trace, their respective locations, and thus their photon emission rates. As the molecular population near the center of the confocal volume may exhibit strong fluctuations, the width of the posterior may also fluctuate from trace to trace, especially when the individual traces are short. Thus, the individual posteriors become sharper only *on average* as we move to higher numbers of molecules inside the effective volume or molecular brightness.

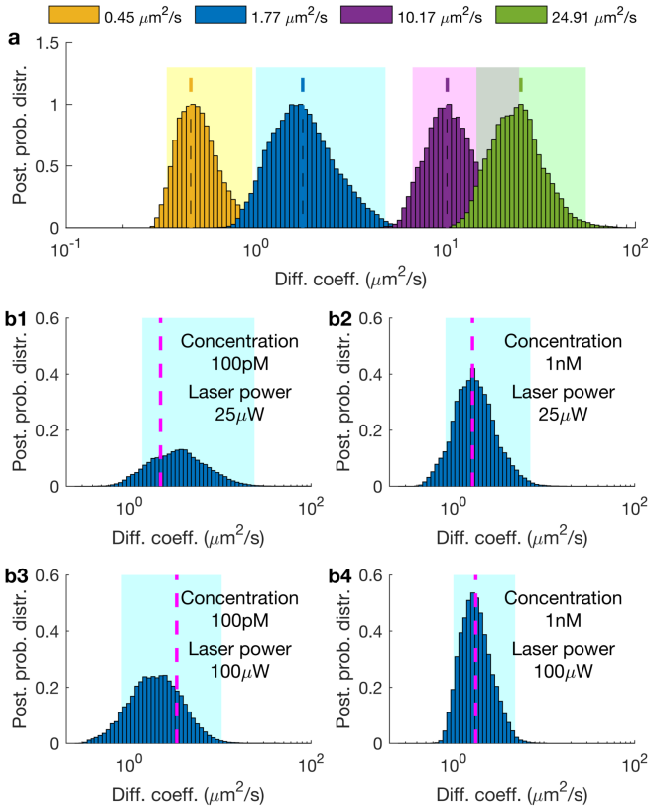

FIG. 5. **Estimating diffusion coefficients of free Cy3.** **(a)** Posterior probability distributions of diffusion coefficients of free Cy3 in different concentrations of glycerol/water mixture. The legend labels the posteriors according to FCS estimates of long time traces. For clarity, posteriors are normalized to maximum 1 and the horizontal axis is shown in logarithmic scale. Also, the 95% confidence intervals are shown by highlighted regions. Posteriors are obtained from the analyses of time traces acquired at 100 μs for total periods of 100 ms. Different diffusion coefficients are obtained by varying the amount of glycerol from 99% to 50% in the glycerol/water mixture. **(b)** Posterior probability distributions over the diffusion coefficients of traces generated by different laser powers (25 and 100 μW, respectively) with different concentrations of Cy3 (100 pM and 1 nM, respectively) in a glycerol/water mixture of 94% glycerol. For comparison, FCS estimates shown by dashed lines are obtained by traces, each of 6 min, i.e. ≈1000× longer than the segments used in our method.

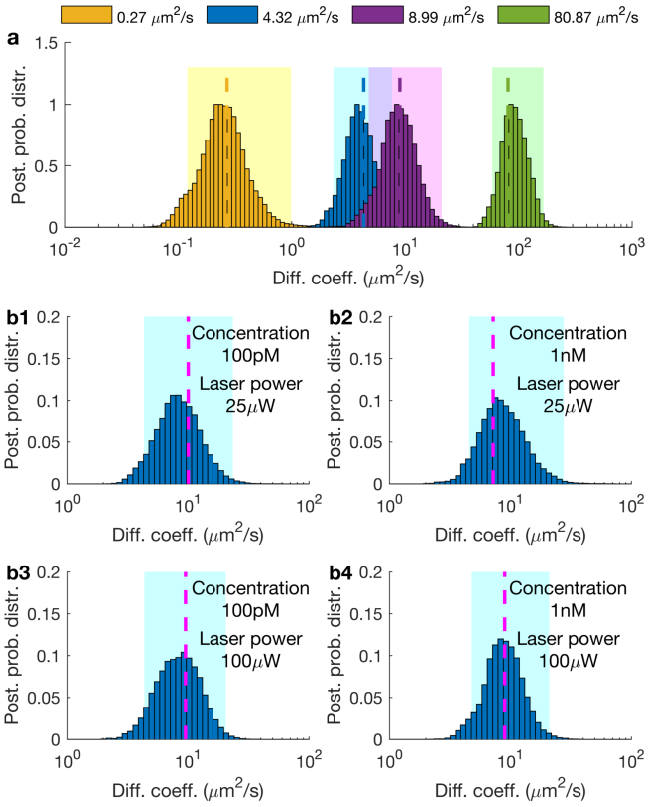

FIG. 6. **Estimating diffusion coefficients of free streptavidin**. **(a)** Posterior probability distributions of diffusion coefficients of free streptavidin labeled by Cy3 in different concentrations of glycerol/water mixture. The legend labels the posteriors according to FCS estimates of long time traces. For clarity, posteriors are normalized to maximum 1 and the horizontal axis is shown in logarithmic scale. Also, the 95% confidence intervals are shown by highlighted regions. Posteriors are obtained from the analyses of time traces acquired at 100 μs for total periods of 100 ms. Different diffusion coefficients are obtained by varying the amount of glycerol from 94% to 0% in the glycerol/water mixture. **(b)** Posterior probability distributions over the diffusion coefficients of traces generated by different laser powers (25 and 100 μW, respectively) with different concentrations of Cy3 (100 pM and 1 nM, respectively) in a glycerol/water mixture of 94% glycerol. For comparison, FCS estimates shown by dashed lines are obtained by traces, each of 6 min, i.e. ≈1000× longer than the segments used in our method.

2. Labeled proteins

To test our method beyond free beads and dyes, we used labeled proteins, namely freely diffusing streptavidin labeled by Cy3. Similar to the previous cases, we tested a range of concentrations, diffusion coefficients, and laser powers. Figure 6 summarizes characteristic results and compares our analyses against the results of FCS applied on longer time traces. As can be seen, our method provides acceptable estimates of the diffusion coefficient even with ≈1000× less datapoints than FCS.

D. Additional results

For all cases described so far, we estimated more than just diffusion coefficients. For example, we also estimate the population of molecules contributing photons to the traces, their instantaneous photon emission rates and locations relative to the center of the confocal volume, as well as the background photon emission rate. A more detailed report of our estimates, with discussions of full joint posterior distributions, can be found in the SUPPLEMENTARY MATERIALS.

In addition to cases involving a single diffusion coeffi-

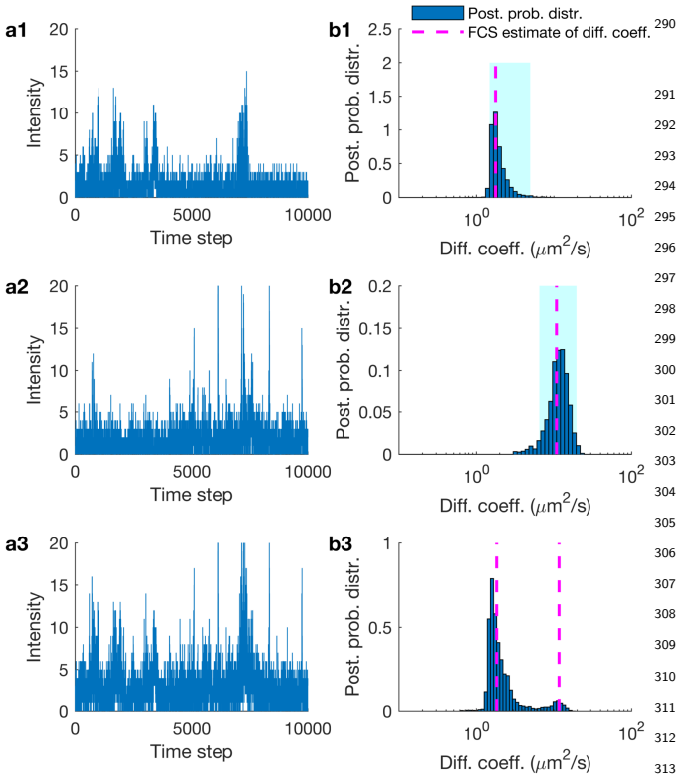

FIG. 7. **Estimating multiple diffusion coefficients in experimental Cy3 traces. (a1)–(a2)** Experimental traces of free Cy3 in glycerol/water mixtures with 94% and 75% glycerol, respectively. **(a3)** Trace resulting by mixing the traces in (a1) and (a2). **(b1)–(b2)** Posterior probability distributions resulting from the analysis of the traces in (a1) and (a2). **(b3)** Posterior probability distribution resulting from the analysis of the trace in (a3). For comparison, FCS estimates shown by dashed lines are produced from five different traces, each of 6 min, i.e. ≈100× longer than the segments shown and analyzed in out method. FCS estimates are highlighted by dashed lines. Posteriors are obtained from the analyses of time traces acquired at 100 μs for a total period of 1 s.

cient that we have considered thus far, our method can be generalized to treat multiple diffusion coefficients as well. To show this, we artificially mixed (summed) and analyzed experimental traces where dyes diffuse in different amounts of glycerol and so they exhibit different diffusion coefficients. On account of the additivity of photon emissions and detections, artificial mixing of traces allows us to obtain realistic traces of different diffusive species that can be analyzed as if they were diffusing simultaneously within the same confocal volume and separately as well. In Fig. 7, we compare the analysis of intensities created by mixing traces containing slow and fast diffusing Cy3 (94% and 75% glycerol/water, respectively). As can be seen, our estimates obtained under simultaneous diffusion compare favorably to the estimates under separate diffusion, indicating that our method can also identify robustly multiple diffusion coefficients at once.

DISCUSSION

Single molecule fluorescence confocal microscopy has the potential to reveal dynamical information at timescales that may be as short as a hundred milliseconds. Here, we have exploited Bayesian non-parametrics to overcome the limitations of specifically fluorescent correlative methods in utilizing short, ≈100 ms, and noisy time traces to deduce molecular properties such as diffusion coefficients. Exploiting novel analysis, to obtain reliable results from such short traces or excessively noisy traces as those obtained under low laser power, is key to minimizing photo-damage inherent to all methods relying on illumination and especially critical to gaining insight on rapid or light-sensitive processes[24–26]. The analysis of similarly short traces may also be required when monitoring non-equilibrium processes that can be resolved only within hundreds of milliseconds. Furthermore, novel analysis with increased sensitivity may reserve longer traces to tease out subtle dynamical features (such as deducing multiple diffusion coefficients at once).

The deep implication of our method is that it places single molecule fluorescence confocal microscopy at a competitive advantage over wide-field techniques used in single particle tracking. Indeed, wide-field techniques provide high, super-resolved, spatial accuracy[15], but with diminished temporal resolution, since molecule localization requires the collection of sufficient photons obtained only after long frame exposures[15]. Such a requirement is especially problematic for photo-sensitive or rapidly diffusing biomolecules[15].

By contrast to wide-field microscopy, single molecule fluorescence confocal microscopy yields minimal spatial resolution. However, as our analysis shows, although spatial resolution may be diminished, reduced photodamage and exceptionally high temporal resolution can be achieved instead.

Since their inception, over half a century ago, correlative methods, such as FCS, have demanded very long traces in order to extract dynamical features from single molecule fluorescence confocal microscopy data[2,11,41–44]. In this study, we have developed a principled framework capable of taking advantage of all spatio-temporal information nested within time traces of photon counts and, together with novel Mathematics, we have reformulated the analysis of single molecule fluorescence confocal microscopy data.

Existing methods, even those that apply Bayesian techniques such as FCS-Bayes[18,22], still utilize autocorrelation functions. Therefore, they demand equally long time traces as FCS and implicitly assume that the physical system under study is at a stationary or equilibrium phase throughout the entire trace. By contrast, our method only requires short traces and therefore it avoids stationarity or equilibrium requirements on timescales longer than those of the data analyzed. In addition, our method also: (i) provide interpretable estimation of errors (i.e. posterior variance) determined exclusively from

the information content of the trace supplied (i.e. length and noise) as opposed to *ad hoc* metric fitting (i.e. chi square); (ii) track instantaneous molecule photon emissions and locations; and (iii) estimate the molecular brightness and background photon emission rates which, if left undetermined, can bias the other estimates.

Since our method is formulated exclusively in the time-domain, it offers a versatile framework for further modifications. For example, it is possible to adapt the present formulation to incorporate scanning-FCS[9,10,45] which involves moving the confocal volume or incorporate demanding illumination profiles, such as those arising in two photon excitation[43,46], TIRF microscopy[42] or even Airy patterns[47] with or without aberrations[48] by changing the specified point spread function (see Methods section). Additionally, it is possible to extend our framework to treat multiple diffusion coefficients (see SUPPLEMENTARY MATERIALS), confining forces or photon emission kinetics as would be relevant for FCS-FRET[49,50] and FLIM[51,52] applications. Also, our method could be extended to handle more complex photophysics[23,53–55], and, since we explicitly track individual molecules over time, extensions appropriate for fast bimolecular reaction kinetics are also conceivable.

METHODS

Here we describe the formulation and mathematical foundation of our model. Our overarching goal is to start from an experimental time series of photon counts, $\overline{w} = (w_1, w_2, ..., w_K)$ where w_k denotes the photon intensity assessed at time t_k (which includes both background photons as well as photons derived from the labeled molecules of interest), and derive estimates of kinetic quantities such as molecular locations with respect to the center of the confocal volume as well as diffusion coefficients.

To derive estimates for the desired quantities, we need to compute intermediate quantities which include: (i) molecular brightness; (ii) background photon emission rate; and, most importantly, (iii) the unknown population of moving molecules and their relative location with respect to the center of the confocal volume. Below we explain each one of these in detail. Computational details and a working implementation of the entire method are available in the SUPPLEMENTARY MATERIALS.

E. Model description

The starting point of our analysis is the raw data, namely the photon counts. As our current focus is on deducing dynamical information on timescales exceeding $\approx 1\ \mu$s, we ignore triplet state and photon anti-bunching effects which occur on vastly different timescales[16,56,57].

At the timescale of interest, individual photon detections happen stochastically and independently from each other. Accordingly, the total number of photon counts w_k between successive assessments follows Poisson[15,27] (shot noise) statistics

$$w_k \sim \text{Poisson}\left((t_k - t_{k-1})\left(\mu_{back} + \sum_n \mu_k^n\right)\right) \quad (1)$$

where μ_{back} is a background photon emission rate and $\sum_n \mu_k^n$ gathers the photon emission rates μ_k^n from individual fluorescent molecules that we index with $n = 1, 2, \ldots$. The the total number of molecules involved in the summation above is to be determined. This is the key reason we invoke Bayesian non-parametrics in the model inference section (see below). Since we only collect a small fraction of the total photons emitted by the fluorescent molecules, in our framework μ_k^n coincides with the emission rate of detected photons, as opposed to the true photon emission rate which might be larger.

Each rate μ_k^n depends on the position (x_k^n, y_k^n, z_k^n) of the corresponding molecule relative to the center of the confocal volume as well as other features such as laser intensity, laser wavelength, quantum yield and camera pinhole size[58]. Similar to other studies[41,59,60], we combine all these effects into a characteristic point spread function (PSF) that combines excitation and emission PSFs

$$\mu_k^n = \mu_{mol}\text{PSF}(x_k^n, y_k^n, z_k^n). \quad (2)$$

The parameter μ_{mol} represents the molecular brightness and as we discuss in the SUPPLEMENTARY MATERIALS it is related to the maximum photon emission rate of a single molecule that is located at the center of the confocal volume. Specific choices of PSF models, such as Gaussian or Gaussian-Lorentzian, are also detailed in the SUPPLEMENTARY MATERIALS.

Finally, we associate individual molecular locations across time by adopting a motion model. Here we assume that molecules are purely diffusive and arrive at

$$\begin{aligned} x_k^n &\sim \text{Normal}\left(x_{k-1}^n, 2(t_k - t_{k-1})D\right) \\ y_k^n &\sim \text{Normal}\left(y_{k-1}^n, 2(t_k - t_{k-1})D\right) \\ z_k^n &\sim \text{Normal}\left(z_{k-1}^n, 2(t_k - t_{k-1})D\right) \end{aligned} \quad (3)$$

where D denotes the diffusion coefficient, which we assume is the same for all molecules. As we explain in the SUPPLEMENTARY MATERIALS, these probabilities result directly from the diffusion equation. Additionally, in the SUPPLEMENTARY MATERIALS, we illustrate how this motion model can be generalized to capture more than one diffusion coefficients.

A graphical summary of the entire formulation is shown on Fig. 8.

F. Model inference

The quantities which we want to estimate, for example the diffusion coefficient D, molecular locations through

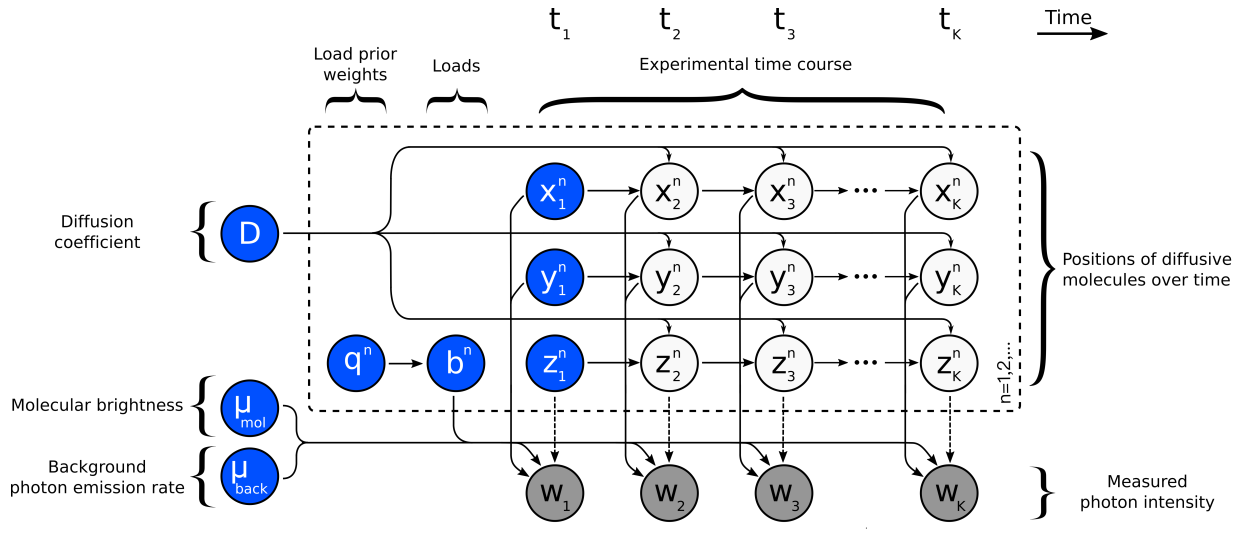

FIG. 8. **Graphical representation of the formulation used in the analysis of fluorescence time traces**. A population of model molecules, labeled by $n = 1, 2, \ldots$, evolves over the course of the experiment which is marked by $k = 1, 2, \ldots, K$. Here, x_k^n,y_k^n,z_k^n denote the location in Cartesian space of molecule n at time t_k; μ_{mol} denotes the brightness of an individual molecule; and μ_{back} denotes the background photon emission rate. During the experiment, only a single observation w_k, combining photon emissions between t_{k-1} and t_k from every molecule and background is recorded at every time step. The diffusion coefficient D determines the evolution of the molecular locations which, in turn, influence the photon emission rates and ultimately the recorded photon intensity w_k. Auxiliary variables b^n, or "loads", and corresponding prior weights q^n, are introduced in order to estimate the unknown population size. The dashed arrows apply for the 3D-Gaussian and 2D-Gaussian-Lorentzian PSFs; while in the case of the 2D-Gaussian-Cylindrical there is no dependency of the measurements w_k on the z_k^n coordinates of the molecules (see the SUPPLEMENTARY MATERIALS for the definitions of these PSFs).

time (x_k^n, y_k^n, z_k^n), molecular brightness μ_{mol} and background photon emission rate μ_{back} or the molecular population are introduced as model variables in the preceding formulation. To estimate values for these variables, we follow the Bayesian paradigm[15,28,38,60].

Variables such as D, μ_{mol} and μ_{back} are parameters of the model and, as such, require priors. Choices for these priors are straightforward and, for interpretational and computational convenience, we adopt the distributions described in the SUPPLEMENTARY MATERIALS.

Additionally, we must place priors on the initial molecular locations, (x_1^n, y_1^n, z_1^n), i.e. the locations of the molecules at the onset of the measurement period. Specifying a prior on initial molecular locations also entails specifying a prior on the molecular population.

In particular, to allow the dimensionality or, alternatively, the complexity of our model to fluctuate based on the number of molecules that contribute to the fluorescent trace, we abandon traditional Bayesian parametric priors and turn to the non-parametric formulation described below.

Before we proceed any further, we recast equation (2) as

$$\mu_k^n = b^n \mu_{mol} \mathrm{PSF}(x_k^n, y_k^n, z_k^n). \quad (4)$$

The newly introduced variables b^n, one for each model molecule, may take only values 1 or 0. In particular, the possibility that $b^n = 0$, coinciding with the case where molecules do not contribute to the observation, allows us to introduce an arbitrarily large number of molecules, technically an infinite number. With the introduction of b^n, we can estimate the number of molecules that contribute photons (termed "active" to distinguish them from those that do not contribute termed "inactive") simultaneously with the rest of the parameters simply by treating each b^n as a separate parameter and estimating its value (of 1 for active molecules and 0 for inactive ones).

To estimate b^n, we place a prior $b^n \sim \mathrm{Bernoulli}(q^n)$ and subsequently a hyperprior on q^n in order to learn precisely how many model molecules are active. For the latter, we choose $q^n \sim \mathrm{Beta}(A_q, B_q)$ with hyperparameters A_q and B_q. Both steps can be combined by invoking the newly developed Beta-Bernoulli process[36,61] which is described in more detail in the SUPPLEMENTARY MATERIALS.

Once the choices for the priors above are made, we form a joint posterior probability $p(D, \mu_{mol}, \mu_{back}, \{x_k^n, y_k^n, z_k^n, b^n, q^n\}_k^n | \overline{w})$ encompassing all unknown variables which we may wish to determine.

The nonlinearities in the PSF, with respect to variables $\{x_k^n, y_k^n, z_k^n\}_k^n$, and the non-parametric prior on $\{b^n, q^n\}^n$ exclude analytic forms for our posterior. For this reason, we develop a computational scheme exploiting Markov chain Monte Carlo[38,62] that can be used to generate pseudo-random samples from this posterior.

The main bottleneck of a naive implementation of our

method, as compared to correlative methods, is its higher computational cost. As we explain in the SUPPLEMENTARY MATERIALS, to have computations run on an average desktop computer, we adopt mathematical approximations (e.g. photon binning, Anscombe transform[63] and filter updates[64,65]) that are tested on the synthetic data presented.

A working implementation (source code and GUI) is given in the SUPPORTING MATERIALS.

G. Data acquisition

1. Synthetic data

We obtain the synthetic data presented in the Results section by standard pseudo-random computer simulations[66–68] that mimic the common single molecule fluorescence confocal setup. We provide details and complete parameter choices in the SUPPLEMENTARY MATERIALS.

2. Experimental data

For the experimental data acquired with elongated confocal volumes, a stock solution of Cy3B (monoreactive NHS ester, GE Healthcare) was prepared by dissolving a small amount of solid in 1 mL of doubly-distilled water, and its concentration was determined from the absorbance of the solution using the extinction coefficient provided by the vendors. A 10 nM solution was then prepared by appropriate dilution of the stock and measured on a silicone perfusion chamber mounted on a glass coverslip. Fluorescent beads were purchased from ThermoFisher (Catalog number: F8792. Lot number: 1604237). The average diameter was 0.046 μm as indicated in the certificate of analysis provided by the vendors. Suspensions for FCS measurements were prepared by adding 3 μL of stock solution (9.4 x 1014 particles/mL) to 1 mL of water and sonicating the mixture for 20 minutes. Measurements were carried out using a home-built instrument. A 532 nm continuous-wave laser (Compass 215M-10, Coherent, Santa Clara, CA) was attenuated to 100 μW and focused onto an PlanApo 100x, 1.4 NA, oil-immersion, objective (Olympus, Center Valley, PA). Emitted fluorescence was collected using the same objective and then passed through a 50 μm pinhole to reject the out-of-focus light. The signal was detected using a silicon avalanche photodiode (SPCM-AQR-14; Perkin-Elmer, Fremont, CA). A band-pass filter (Omega 3RD560-620) in front of the detector was employed to further reduce the background signal and an ALV correlator card (ALV 5000/EPP, ALV-GmbH, Langen, Germany) was used to correlate the detected fluorescence signal. Data for our analysis were acquired with 100 μs resolution using a PCI-6602 acquisition card (National Instruments, Austin, TX).

For the experimental data acquired with elliptical confocal volumes, Cy3 dye and Cy3-labeled streptavidin solutions were prepared by suspending Cy3 or streptavidin in glycerol/buffer (pH 7.5, 10 mM Tris-HCl,100 mM NaCl and 10 mM KCl, 2.5 mM $CaCl_2$) at different v/v, to a final concentration of either 100 pM or 1 nM. The solutions were added onto a glass-bottomed fluid-cell, mounted on a custom designed single molecule fluorescence confocal microscope[69,70] 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 NA, oil-immersion objective (Olympus). the laser power was measured before the objective and the beam is reflected by a dichroic and focused by the objective on to the sample. The dichroic reflected 95% of the intensity on to the objective. Emitted fluorescence was collected by the same objective and focused onto the detection face of a Single Photon Avalanche Diode (SPAD, Micro Photon Devices) that has a maximum count rate of 11.8 Mc/s. A bandpass filter was placed in front of the detector to transmit only the fluorescence from Cy3 and to block the back-scattered excitation light. TTL pulses, triggered by the arrival of individual photons on the SPAD, were timestamped and recorded at 80 MHz by a field programmable gated array (FPGA, NI Instruments) using custom LabVIEW software and initially binned at 100 μs[70].

REFERENCES

[1] E. L. Elson and D. Magde, "Fluorescence correlation spectroscopy. i. conceptual basis and theory," Biopolymers **13**, 1–27 (1974).

[2] D. Magde, E. L. Elson, and W. W. Webb, "Fluorescence correlation spectroscopy. ii. an experimental realization," Biopolymers **13**, 29–61 (1974).

[3] G. R. Bright, G. W. Fisher, J. Rogowska, and D. L. Taylor, "Fluorescence ratio imaging microscopy," Methods in cell biology **30**, 157–192 (1989).

[4] J. A. Fitzpatrick and B. F. Lillemeier, "Fluorescence correlation spectroscopy: linking molecular dynamics to biological function in vitro and in situ," Current opinion in structural biology **21**, 650–660 (2011).

[5] M. A. Digman and E. Gratton, "Lessons in fluctuation correlation spectroscopy," Annual review of physical chemistry **62**, 645–668 (2011).

[6] C. Di Rienzo, E. Gratton, F. Beltram, and F. Cardarelli, "Fast spatiotemporal correlation spectroscopy to determine protein lateral diffusion laws in live cell membranes," Proceedings of the National Academy of Sciences **110**, 12307–12312 (2013).

[7] M. A. Digman, C. M. Brown, P. Sengupta, P. W. Wiseman, A. R. Horwitz, and E. Gratton, "Measuring fast dynamics in solutions and cells with a laser scanning microscope," Biophysical journal **89**, 1317–1327 (2005).

[8] P. Schwille, F.-J. Meyer-Almes, and R. Rigler, "Dual-color fluorescence cross-correlation spectroscopy for multicomponent diffusional analysis in solution," Biophysical journal **72**, 1878–1886 (1997).

[9] Z. Petrášek and P. Schwille, "Precise measurement of diffusion coefficients using scanning fluorescence correlation spectroscopy," Biophysical Journal **94**, 1437–1448 (2008).

[10] Z. Petrášek and P. Schwille, "Scanning fluorescence correlation

spectroscopy," in *Single molecules and nanotechnology* (Springer, 2008) pp. 83–105.

[11] V. Ruprecht, S. Wieser, D. Marguet, and G. J. Schütz, "Spot variation fluorescence correlation spectroscopy allows for super-resolution chronoscopy of confinement times in membranes," Biophysical journal **100**, 2839–2845 (2011).

[12] K. Remaut, B. Lucas, K. Braeckmans, N. Sanders, S. De Smedt, and J. Demeester, "Fret-fcs as a tool to evaluate the stability of oligonucleotide drugs after intracellular delivery," Journal of controlled release **103**, 259–271 (2005).

[13] T. Torres and M. Levitus, "Measuring conformational dynamics: a new fcs-fret approach," The Journal of Physical Chemistry B **111**, 7392–7400 (2007).

[14] K. Tsekouras, A. P. Siegel, R. N. Day, and S. Pressé, "Inferring diffusion dynamics from fcs in heterogeneous nuclear environments," Biophysical journal **109**, 7–17 (2015).

[15] A. Lee, K. Tsekouras, C. Calderon, C. Bustamante, and S. Pressé, "Unraveling the thousand word picture: An introduction to super-resolution data analysis," Chemical Reviews **117**, 7276–7330 (2017).

[16] R. Rigler and E. S. Elson, *Fluorescence correlation spectroscopy: theory and applications*, Vol. 65 (Springer Science & Business Media, 2012).

[17] J. Enderlein, I. Gregor, D. Patra, and J. Fitter, "Statistical analysis of diffusion coefficient determination by fluorescence correlation spectroscopy," Journal of fluorescence **15**, 415–422 (2005).

[18] S.-M. Guo, J. He, N. Monnier, G. Sun, T. Wohland, and M. Bathe, "Bayesian approach to the analysis of fluorescence correlation spectroscopy data ii: application to simulated and in vitro data," Analytical chemistry **84**, 3880–3888 (2012).

[19] J. He, S.-M. Guo, and M. Bathe, "Bayesian approach to the analysis of fluorescence correlation spectroscopy data i: theory," Analytical chemistry **84**, 3871–3879 (2012).

[20] W. Kügel, A. Muschielok, and J. Michaelis, "Bayesian-inference-based fluorescence correlation spectroscopy and single-molecule burst analysis reveal the influence of dye selection on dna hairpin dynamics," ChemPhysChem **13**, 1013–1022 (2012).

[21] K. R. Murphy, C. A. Stedmon, P. Wenig, and R. Bro, "Openfluor–an online spectral library of auto-fluorescence by organic compounds in the environment," Analytical Methods **6**, 658–661 (2014).

[22] G. Sun, S.-M. Guo, C. Teh, V. Korzh, M. Bathe, and T. Wohland, "Bayesian model selection applied to the analysis of fluorescence correlation spectroscopy data of fluorescent proteins in vitro and in vivo," Analytical Chemistry **87**, 4326–4333 (2015).

[23] T. J. Stasevich, F. Mueller, A. Michelman-Ribeiro, T. Rosales, J. R. Knutson, and J. G. McNally, "Cross-validating frap and fcs to quantify the impact of photobleaching on in vivo binding estimates," Biophysical journal **99**, 3093–3101 (2010).

[24] Z. Liu, L. D. Lavis, and E. Betzig, "Imaging live-cell dynamics and structure at the single-molecule level," Molecular cell **58**, 644–659 (2015).

[25] M. Purschke, N. Rubio, K. D. Held, and R. W. Redmond, "Phototoxicity of hoechst 33342 in time-lapse fluorescence microscopy," Photochemical & Photobiological Sciences **9**, 1634–1639 (2010).

[26] V. Magidson and A. Khodjakov, "Circumventing photodamage in live-cell microscopy," in *Methods in cell biology*, Vol. 114 (Elsevier, 2013) pp. 545–560.

[27] U. Von Toussaint, "Bayesian inference in physics," Reviews of Modern Physics **83**, 943 (2011).

[28] M. Tavakoli, J. N. Taylor, C.-B. Li, T. Komatsuzaki, and S. Pressé, "Single molecule data analysis: An introduction," in *Advances in Chemical Physics* (John Wiley & Sons, 2017) Chap. 4, pp. 205–305.

[29] K. E. Hines, "A primer on bayesian inference for biophysical systems," Biophysical journal **108**, 2103–2113 (2015).

[30] T. S. Ferguson, "A bayesian analysis of some nonparametric problems," The annals of statistics , 209–230 (1973).

[31] I. Sgouralis and S. Pressé, "An introduction to infinite hmms for single-molecule data analysis," Biophysical Journal **112**, 2021–2029 (2017).

[32] I. Sgouralis and S. Pressé, "Icon: an adaptation of infinite hmms for time traces with drift," Biophysical journal **112**, 2117–2126 (2017).

[33] I. Sgouralis, M. Whitmore, L. Lapidus, M. J. Comstock, and S. Pressé, "Single molecule force spectroscopy at high data acquisition: A bayesian nonparametric analysis," The Journal of Chemical Physics **148**, 123320 (2018).

[34] C. P. Calderon and K. Bloom, "Inferring latent states and refining force estimates via hierarchical dirichlet process modeling in single particle tracking experiments," PloS one **10**, e0137633 (2015).

[35] K. E. Hines, J. R. Bankston, and R. W. Aldrich, "Analyzing single-molecule time series via nonparametric bayesian inference," Biophysical journal **108**, 540–556 (2015).

[36] J. Paisley and L. Carin, "Nonparametric factor analysis with beta process priors," in *Proceedings of the 26th Annual International Conference on Machine Learning* (ACM, 2009) pp. 777–784.

[37] I. Sgouralis, A. Nebenführ, and V. Maroulas, "A bayesian topological framework for the identification and reconstruction of subcellular motion," SIAM Journal on Imaging Sciences **10**, 871–899 (2017).

[38] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*, Vol. 2 (CRC press Boca Raton, FL, 2014).

[39] Y. Chen, J. D. Müller, Q. Ruan, and E. Gratton, "Molecular brightness characterization of egfp in vivo by fluorescence fluctuation spectroscopy," Biophysical journal **82**, 133–144 (2002).

[40] H. C. Berg, *Random walks in biology* (Princeton University Press, 1993).

[41] T. Wohland, R. Rigler, and H. Vogel, "The standard deviation in fluorescence correlation spectroscopy," Biophysical journal **80**, 2987–2999 (2001).

[42] K. Hassler, M. Leutenegger, P. Rigler, R. Rao, R. Rigler, M. Gösch, and T. Lasser, "Total internal reflection fluorescence correlation spectroscopy (tir-fcs) with low background and high count-rate per molecule," Optics Express **13**, 7415–7423 (2005).

[43] K. M. Berland, P. So, and E. Gratton, "Two-photon fluorescence correlation spectroscopy: method and application to the intracellular environment," Biophysical Journal **68**, 694–701 (1995).

[44] R. Rigler, Ü. Mets, J. Widengren, and P. Kask, "Fluorescence correlation spectroscopy with high count rate and low background: analysis of translational diffusion," European Biophysics Journal **22**, 169–175 (1993).

[45] J. Enderlein, I. Gregor, D. Patra, and J. Fitter, "Art and artefacts of fluorescence correlation spectroscopy," Current pharmaceutical biotechnology **5**, 155–161 (2004).

[46] T. Dertinger, V. Pacheco, I. von der Hocht, R. Hartmann, I. Gregor, and J. Enderlein, "Two-focus fluorescence correlation spectroscopy: A new tool for accurate and absolute diffusion measurements," ChemPhysChem **8**, 433–443 (2007).

[47] M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light* (Elsevier, 2013).

[48] S. F. Gibson and F. Lanni, "Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy," JOSA A **9**, 154–166 (1992).

[49] M. Pirchi, R. Tsukanov, R. Khamis, T. E. Tomov, Y. Berger, D. C. Khara, H. Volkov, G. Haran, and E. Nir, "Photon-by-photon hidden markov model analysis for microsecond single-molecule fret kinetics," The Journal of Physical Chemistry B **120**, 13065–13075 (2016).

[50] J. E. Bronson, J. Fei, J. M. Hofman, R. L. Gonzalez Jr, and C. H. Wiggins, "Learning rates and states from biophysical time series: a bayesian approach to model selection and single-molecule fret data," Biophysical journal **97**, 3196–3205 (2009).

[51] E. B. van Munster and T. W. Gadella, "Fluorescence lifetime imaging microscopy (flim)," in *Microscopy techniques* (Springer, 2005) pp. 143–175.

[52]P. I. Bastiaens and A. Squire, "Fluorescence lifetime imaging microscopy: spatial resolution of biochemical processes in the cell," Trends in cell biology **9**, 48–52 (1999).

[53]L. Song, C. Varma, J. Verhoeven, and H. J. Tanke, "Influence of the triplet excited state on the photobleaching kinetics of fluorescein in microscopy," Biophysical journal **70**, 2959–2968 (1996).

[54]D. Soumpasis, "Theoretical analysis of fluorescence photobleaching recovery experiments," Biophysical journal **41**, 95–97 (1983).

[55]I. Sgouralis, S. Madaan, F. Djutanta, R. Kha, R. F. Hariadi, and S. Pressé, "A bayesian nonparametric approach to single molecule förster resonance energy transfer," The Journal of Physical Chemistry B (2018).

[56]J. Widengren, U. Mets, and R. Rigler, "Fluorescence correlation spectroscopy of triplet states in solution: a theoretical and experimental study," The Journal of Physical Chemistry **99**, 13368–13379 (1995).

[57]K. S. Grußmayer, A. Kurz, and D.-P. Herten, "Single-molecule studies on the label number distribution of fluorescent markers," Chemphyschem **15**, 734–742 (2014).

[58]G. Brakenhoff, K. Visscher, and H. 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* (Springer, 1990) pp. 87–91.

[59]Y. Chen, J. D. Müller, P. T. So, and E. Gratton, "The photon counting histogram in fluorescence fluctuation spectroscopy," Biophysical journal **77**, 553–567 (1999).

[60]S. Jazani, I. Sgouralis, and S. Pressé, "A method for single molecule tracking using a conventional single-focus confocal setup," The Journal of Chemical Physics **150**, 123320 (2019).

[61]K. Tsekouras, T. C. Custer, H. Jashnsaz, N. G. Walter, and S. Pressé, "A novel method to accurately locate and count large numbers of steps by photobleaching," Molecular Biology of the Cell **27**, 3601–3615 (2016).

[62]C. Robert and G. Casella, *Introducing Monte Carlo Methods with R* (Springer Science & Business Media, 2009).

[63]F. J. Anscombe, "The transformation of poisson, binomial and negative-binomial data," Biometrika **35**, 246–254 (1948).

[64]H. M. Menegaz, J. Y. Ishihara, G. A. Borges, and A. N. Vargas, "A systematization of the unscented kalman filter theory," IEEE Transactions on automatic control **60**, 2583–2598 (2015).

[65]E. A. Wan and R. Van Der Merwe, "The unscented kalman filter for nonlinear estimation," in *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000* (Ieee, 2000) pp. 153–158.

[66]O. C. Ibe, *Elements of Random Walk and Diffusion Processes* (John Wiley & Sons, 2013).

[67]D. J. Higham, "An algorithmic introduction to numerical simulation of stochastic differential equations," SIAM review **43**, 525–546 (2001).

[68]R. Erban and S. J. Chapman, "Stochastic modelling of reaction–diffusion processes: algorithms for bimolecular reactions," Physical biology **6**, 046001 (2009).

[69]H. Li, C.-F. Yen, and S. Sivasankar, "Fluorescence axial localization with nanometer accuracy and precision," Nano letters **12**, 3731–3735 (2012).

[70]P. D. Schmidt, B. H. Reichert, J. G. Lajoie, and S. Sivasankar, "Method for high frequency tracking and sub-nm sample stabilization in single molecule fluorescence microscopy," Scientific reports **8**, 13912 (2018).

ACKNOWLEDGMENTS

SP acknowledges support from NSF CAREER grant MCB-1719537. SS acknowledge support from the National Institute of Health grant R01GM121885. ML thanks Anirban Purohit for assistance with the experiments.

AUTHOR CONTRIBUTIONS

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

CODE AVAILABILITY.

Source code and GUI versions of the methods developed herewith are available through the SUPPORTING MATERIALS.

An Alternative Framework for Fluorescence Correlation Spectroscopy — Supplementary Material —

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

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

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

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

Here we provide supplementary materials and technical details that complement the main text. These include: (i) Additional analysis results that demonstrate the estimation of molecular brightness and background photon emission rates, joint posterior probability distributions, molecule locations, and additional results for multiple diffusive species. These results are repeated for simulated and experimental data. (ii) Additional details of the methods used including descriptions of the motion model, the Stokes-Einstein model, point spread functions (PSFs), and time trace preparation. (iii) A complete description of the inference framework developed that includes choices for the prior probability distributions and a computational implementation. (iv) A description of the modifications necessary for the model with multiple diffusive species. (v) Summary of notation and other conventions used throughout this study as well as detailed parameter choices for the simulations and analyses.

Contents

S1. Additional results

S1.1. Analysis of additional simulated data

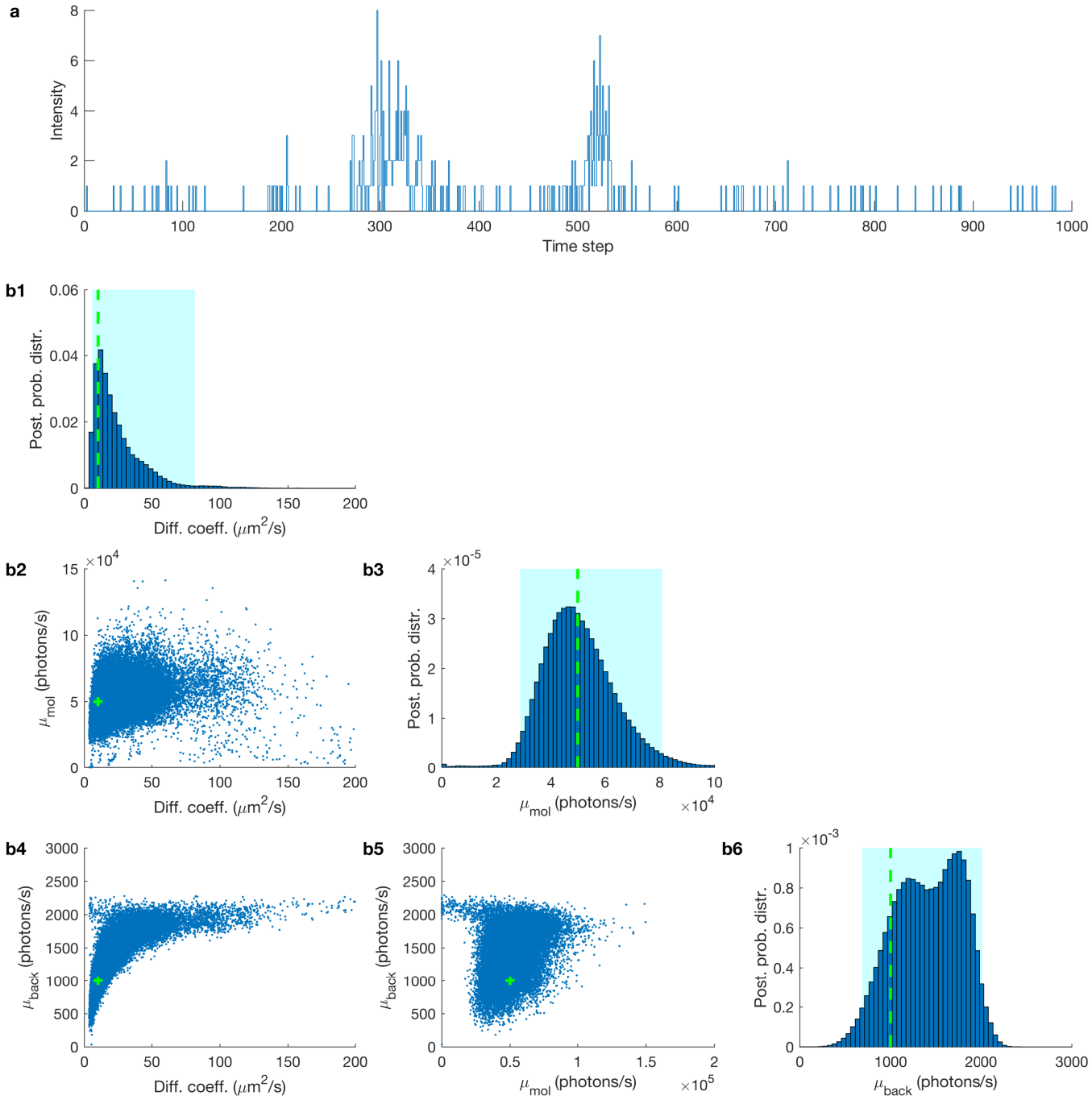

FIG. S1. **Estimated joint posterior probability distribution. (a)** Synthetic fluorescent intensity trace used in Fig. 1a1 with a length of 1000 data points and time step 100 μs. The true values of the diffusion coefficient, molecular brightness and background emission rates are, 10 $\mu m^2/s$, 5×10^4 photons/s and 10^3 photons/s (shown by green dashed lines). **(b1)** The posterior of the diffusion coefficient. **(b2)** The joint probability distribution of the diffusion coefficient and molecular brightness. **(b3)** The posterior probability distribution of the molecular brightness. **(b4)** The joint probability distribution of the diffusion coefficient and molecular brightness. **(b5)** The joint probability distribution of the molecular brightness and background photon emission rates. **(b6)** The posterior probability distribution of the background photon emission rate. The 95% confidence intervals are shown with pink highlighted regions.

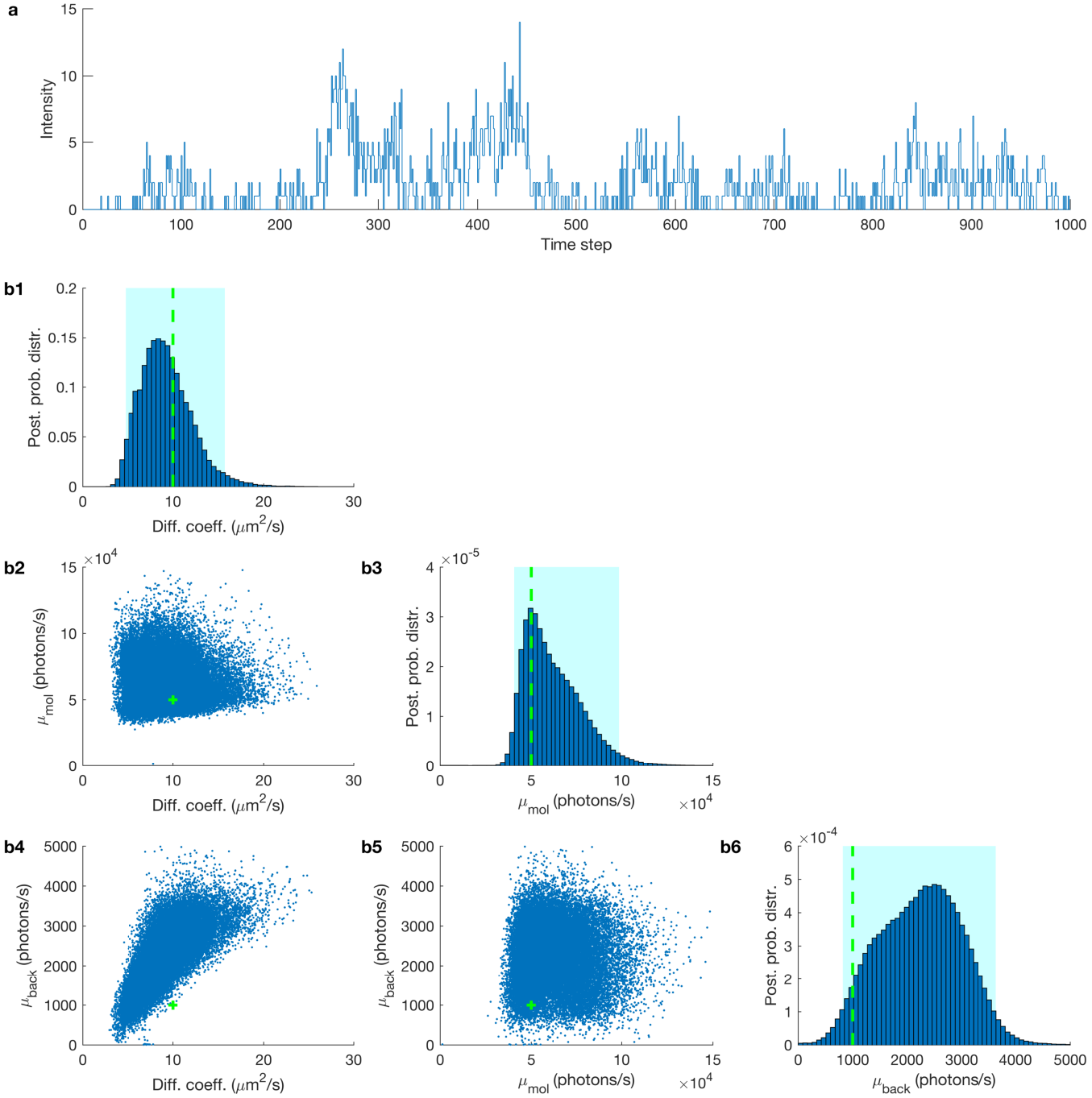

FIG. S2. **Estimated joint posterior probability distribution. (a)** Synthetic fluorescent intensity trace used in Fig. 1a2 with a length of 1000 data points and time step 100 μs. The true values of the diffusion coefficient, molecular brightness and background emission rates are, 10 $\mu m^2/s$, 5×10^4 photons/s and 10^3 photons/s (shown by green dashed lines). **(b1)** The posterior of the diffusion coefficient. **(b2)** The joint probability distribution of diffusion coefficient and molecular brightness. **(b3)** The posterior probability distribution of the molecular brightness. **(b4)** The joint probability distribution of the diffusion coefficient and molecular brightness. **(b5)** The joint probability distribution of the molecular brightness and background photon emission rates. **(b6)** The posterior probability distribution of the background photon emission rate. The 95% confidence intervals of the posterior over the number of molecules is highlighted in cyan.

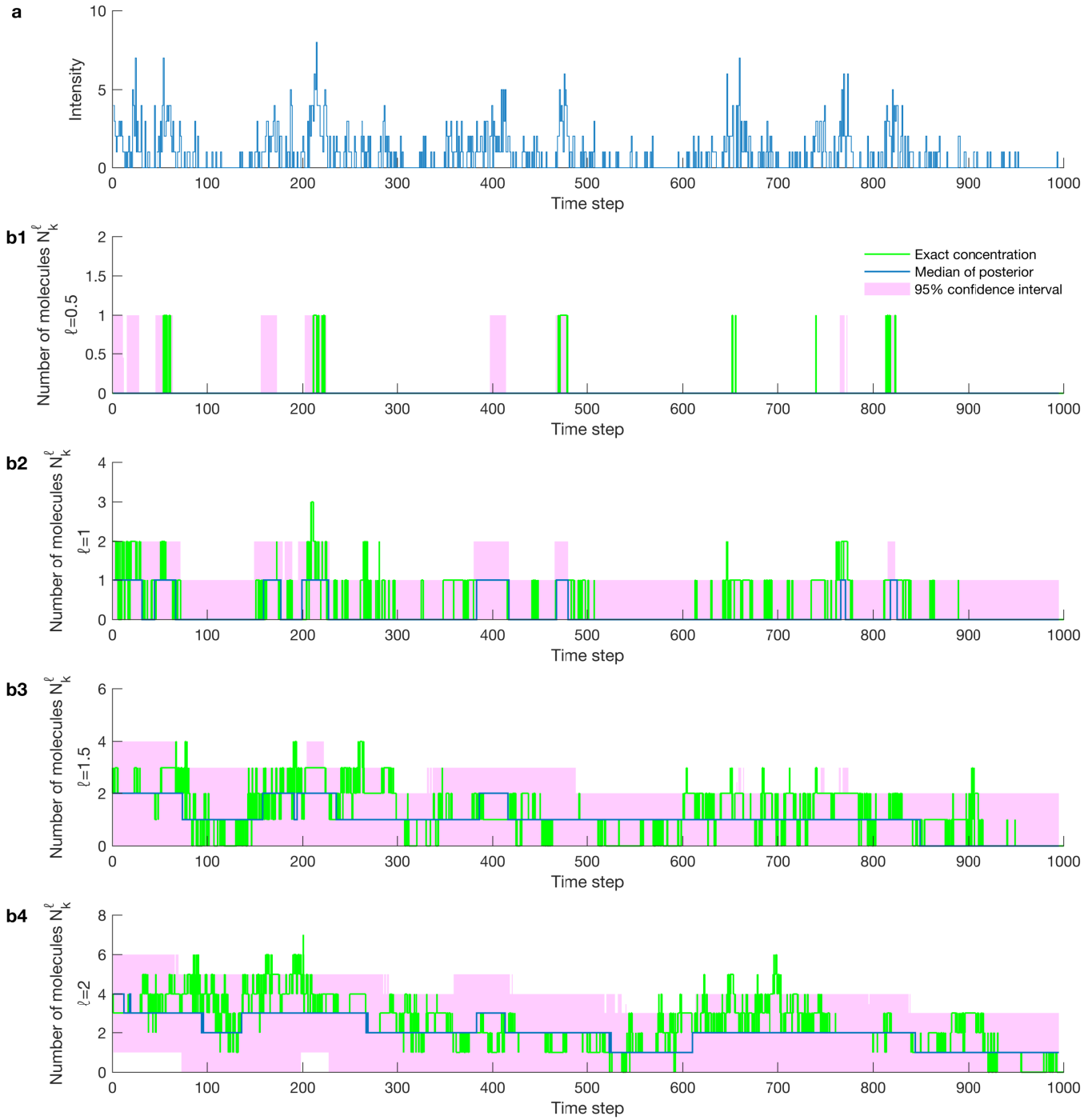

FIG. S3. **Estimated number of molecules/concentrations. (a)** Synthetic fluorescent intensity trace produced with a molecular brightness of 5×10^4 photons/s and a background photon emission rate of 10^3 photons/s, diffusion coefficient of 10 μm^2/s and 50 molecules. **(b1)–(b4)** Number of molecules estimated from the trace in (a) corresponding to normalized distances from the confocal center of $\ell = 0.5, 1, 1.5, 2$. The exact number of molecules is shown by the green lines, the median of the posterior over the number of the molecules is shown by the blue lines, and the 95% confidence intervals of the posteriors over the number of the molecules are highlighted in pink. For details of the definition of number of molecules N_k^ℓ and the normalized distance ℓ, see Eq. (S23), below.

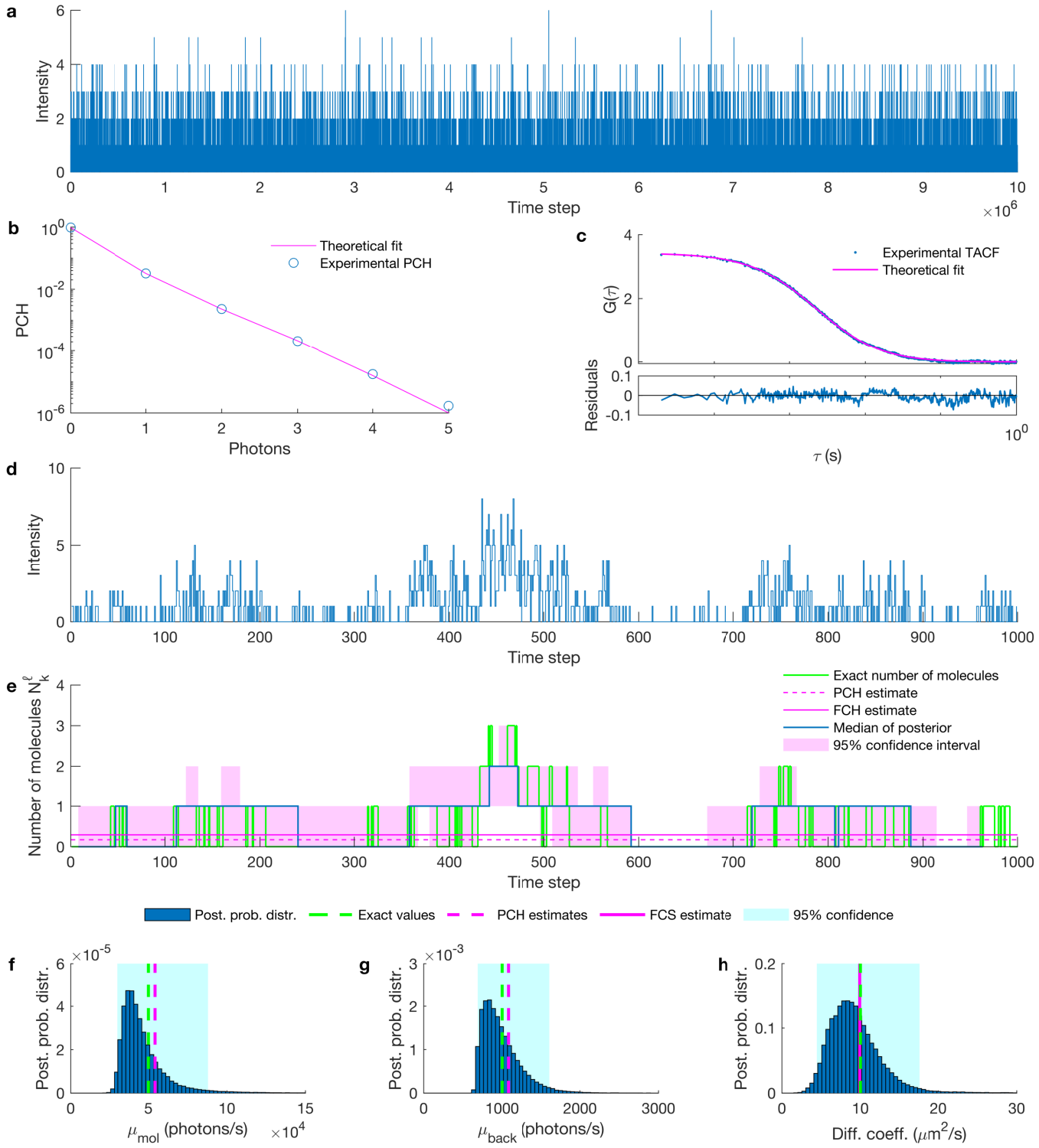

FIG. S4. **Comparison of estimated photon emission rates and concentration with FCS and PCH. (a)** Targeted synthetic fluorescent intensity trace. The time step is 10 μs and the total duration of the trace is 100 s. **(b)** PCH curve and the theoretical fit. **(c)** FCS curve and best theoretical fit. **(d)** The portion of the trace analyzed by our method rebinned at 100 μs. **(e)** The number of molecules in the effective volume with ℓ =1, arising from the trace in (d). Exact value of the number of molecules is shown by the green line and the PCH and FCS estimates are shown by the dashed and solid pink lines. **(g)** On the posterior probability distribution of the molecular brightness we superpose the PCH estimate of the molecular brightness (pink dashed line) and the true value (green dashed line). **(h)** On the posterior probability distribution of the background photon emission rate we superpose the PCH estimate of the background photon emission rate (pink dashed line) and the true value (green dashed line). **(f)** The posterior probability distribution of the diffusion coefficient obtained by analyzing the trace in (d). The FCS estimate of the diffusion coefficient obtained by analyzing the total trace, shown in (a), illustrated by a pink dashed line with the exact value (green dashed line). The targeted synthetic trace is generated by freely diffusive molecules with diffusion coefficient, molecular brightness and background photon emission rates of of 10 μm^2/s, 5×10^4 photons/s and 10^3 photons/s, respectively.

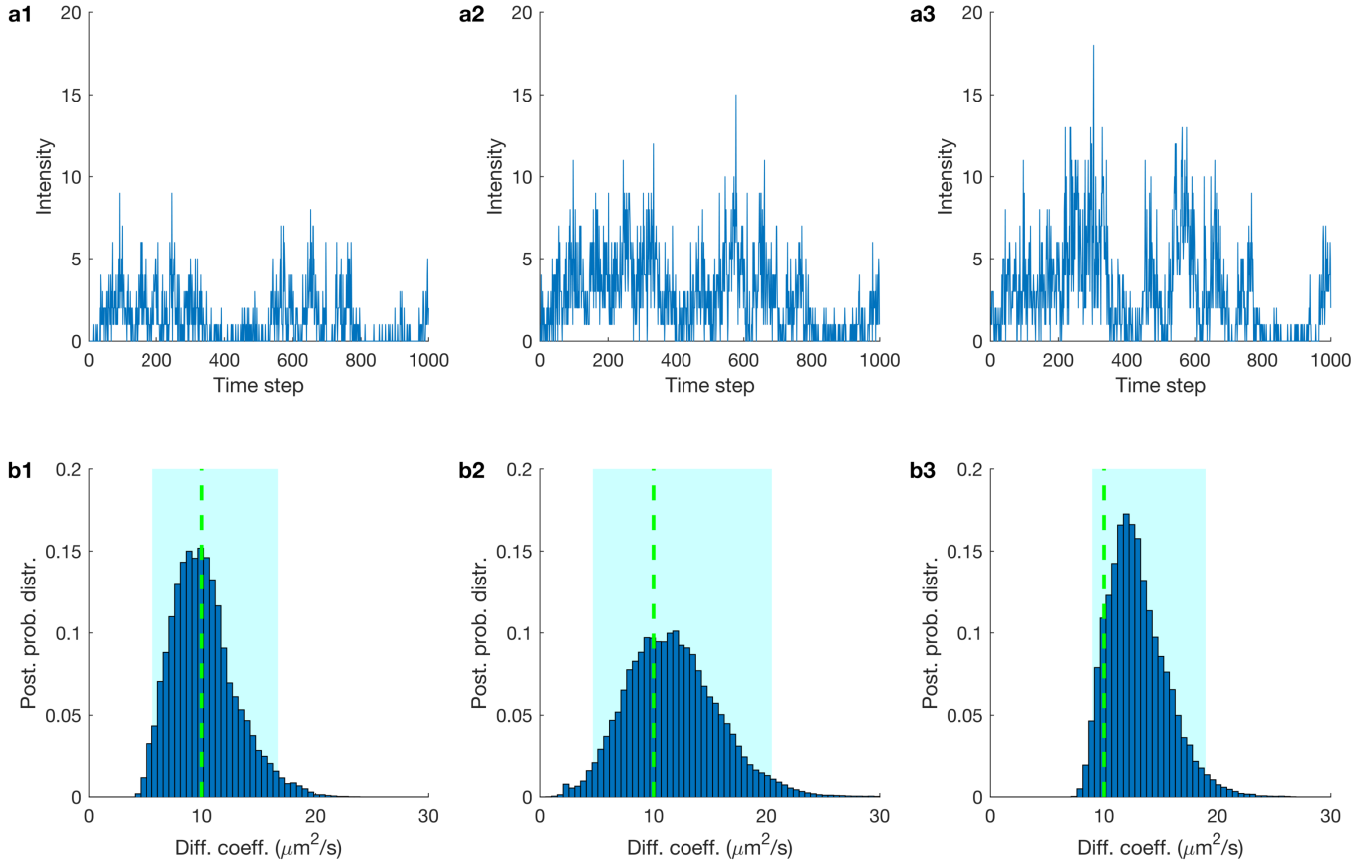

FIG. S5. **Comparison of the posterior on the diffusion coefficient obtained from synthetic fluorescent intensity traces under different PSF models**. **(a1)** Synthetic fluorescent intensity trace produced with a 3DG PSF, Eq. (S16). **(a2)** Synthetic fluorescent intensity trace produced with a 2DGL PSF, Eq. (S18). **(a3)** Synthetic fluorescent intensity trace produced with a 2DGC PSF, Eq. (S17). **(b1)** Posterior of the diffusion coefficient using a 3DG PSF model on the trace in (a1). **(b2)** Posterior of the diffusion coefficient using a 2DGL PSF model on the trace in (a2). **(b3)** Posterior of the diffusion coefficient using a 2DGC PSF model on the trace in (a3). To facilitate the comparison both traces analyzed are generated using the same underlying molecule trajectories with molecular brightness and background photon emission rates of 5×10^4 photons/s and 10^3 photons/s, diffusion coefficient of 1 μm^2/s (shown by green dashed lines).

60

S1.2. Analysis of additional experimental data

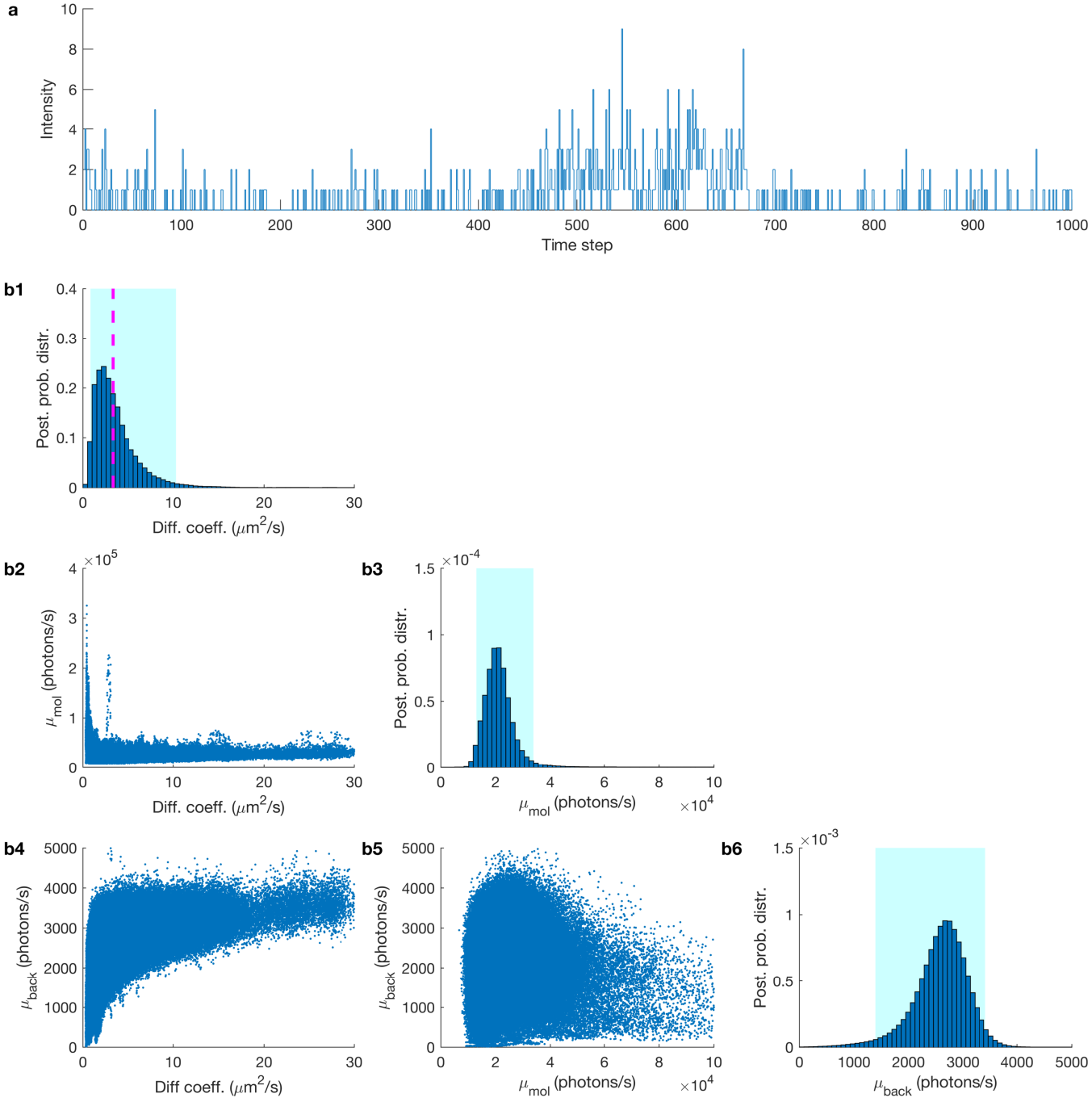

FIG. S6. **Estimated joint posterior probability distribution.** **(a)** Experimental fluorescent intensity trace used in Fig. 5b3 with a length of 1000 data points and time step 100 μs. **(b1)** The posterior of the diffusion coefficient. The FCS estimate is shown by a magenta dashed line. **(b2)** The joint probability distribution of the diffusion coefficient and the molecular brightness. **(b3)** The posterior probability distribution of the molecular brightness. **(b4)** The joint probability distribution of the diffusion coefficient and molecular brightness. **(b5)** The joint probability distribution of the molecular brightness and background photon emission rates. **(b5)** The posterior probability distribution of the background photon emission rate and the 95% confidence intervals of the posteriors are highlighted in cyan. The experimental fluorescent intensity trace was produced with a concentration of 100 pM of Cy3 in a 94% glycerol/water mixture.

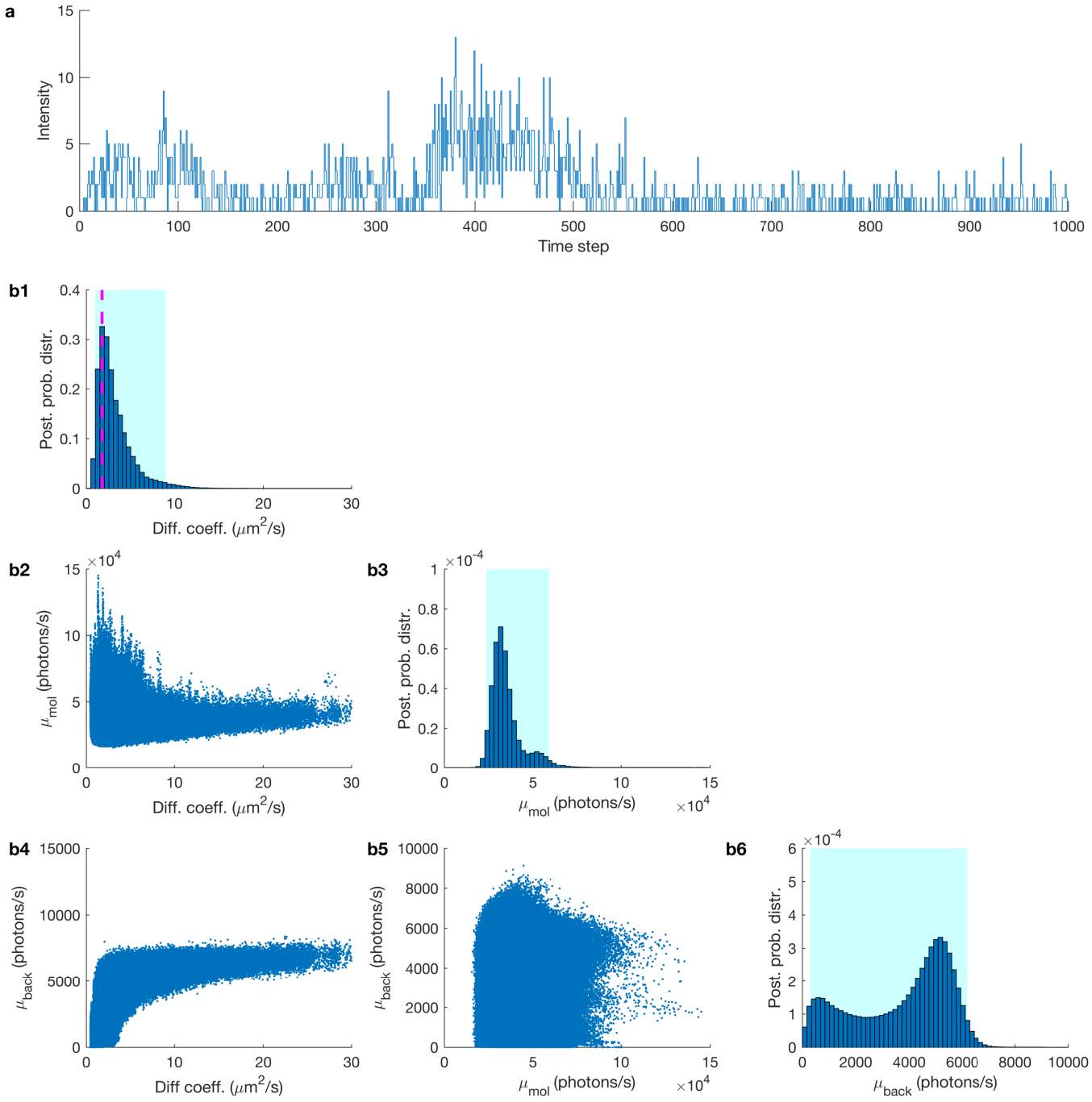

FIG. S7. **Estimated joint posterior probability distribution. (a)** Experimental fluorescent intensity trace used in Fig. 5b4 with a length of 1000 data points and time step 100 μs. **(b1)** The posterior of the diffusion coefficient. The FCS estimate is shown by a magenta dashed line. **(b2)** The joint probability distribution of the diffusion coefficient and the molecular brightness. **(b3)** The posterior probability distribution of the molecular brightness. **(b4)** The joint probability distribution of the diffusion coefficient and the molecular brightness. **(b5)** The joint probability distribution of the molecular brightness and background photon emission rates. **(b5)** The posterior probability distribution of the background photon emission rate and the 95% confidence intervals of the posteriors are highlighted in cyan. The experimental fluorescent intensity trace produced is with a concentration of 1 nM of Cy3 in a 94% glycerol/water mixture.

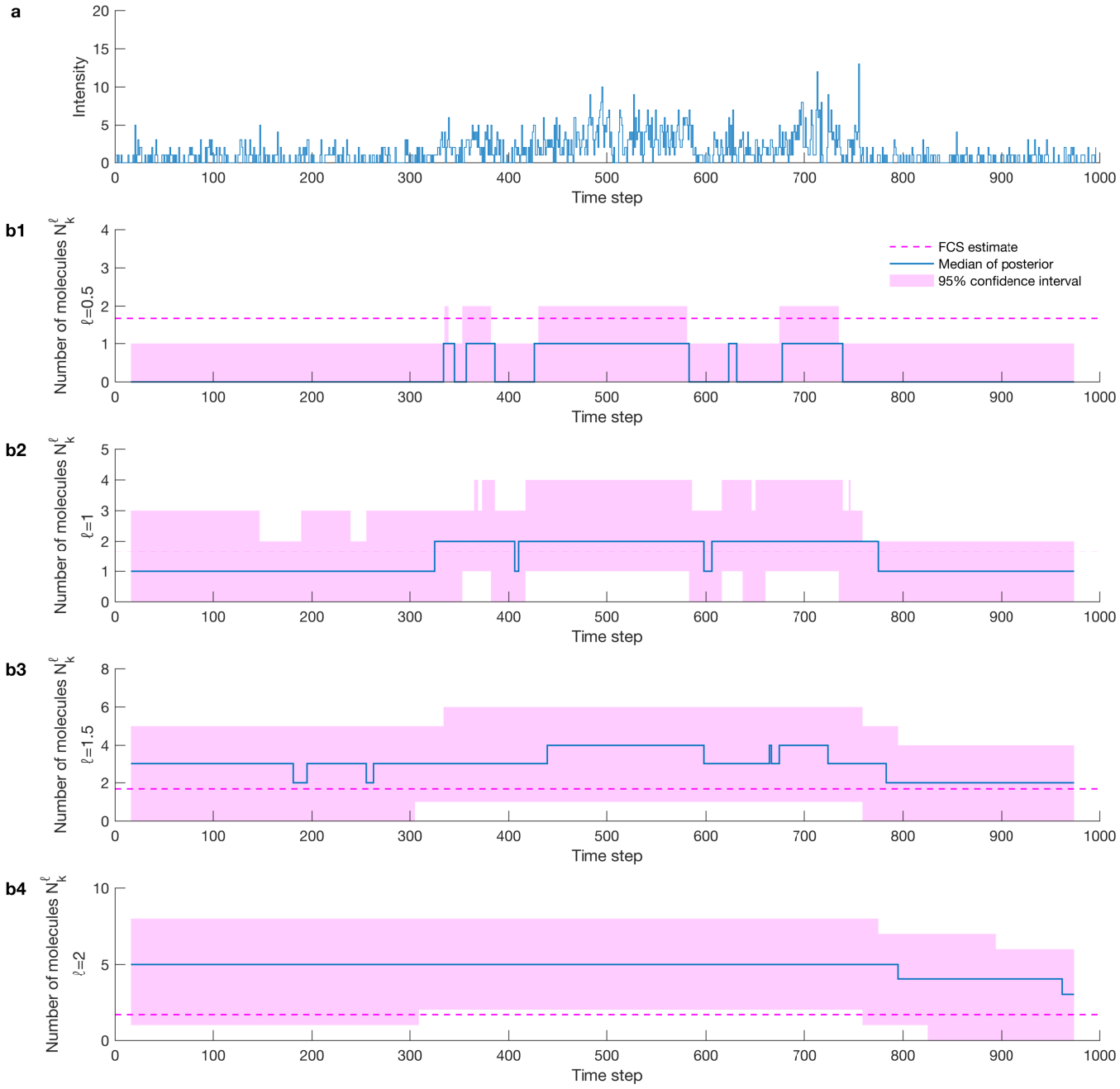

FIG. S8. **Estimated number of molecules/concentrations**. **(a)** Experimental fluorescent intensity trace produced with a concentration of 1 nM of Cy3 in a 94% glycerol/water mixture. **(b1)–(b4)** Number of molecules estimated from the trace in (a) with $\ell = 0.5, 1, 1.5, 2$, respectively. The FCS estimate of the average number of molecules in the effective volume ($\sim$1.68 *molecules*) by analyzing a 3 minutes long time trace, is shown by the magenta dashed lines and the median of the posterior over the number of the molecules is shown by a blue line. The 95% confidence interval of the posterior over the number of the molecules is highlighted in pink. For details of the definition of number of the molecules N_k^ℓ and the normalized distance ℓ, see Eq. (S23), below.

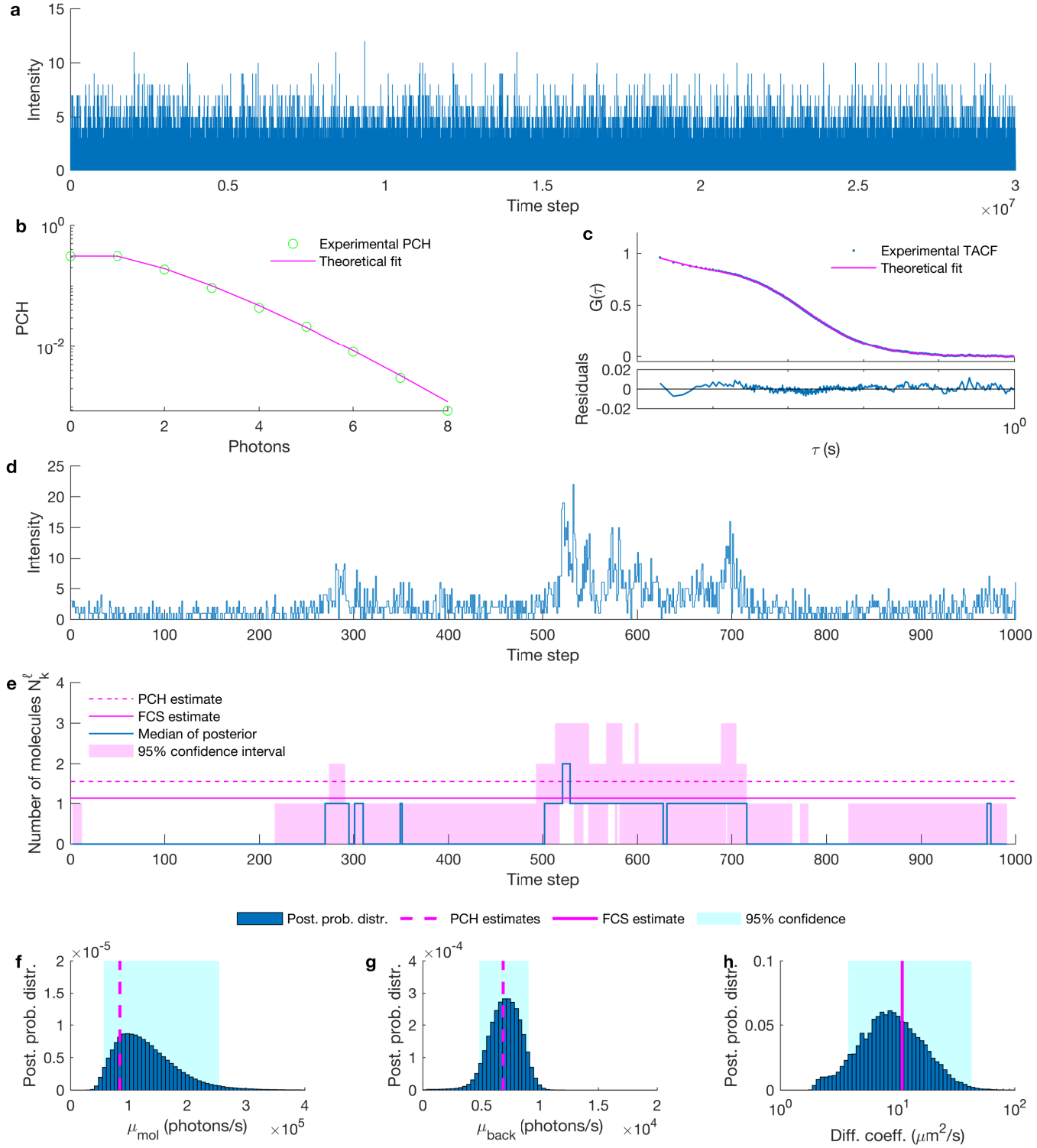

FIG. S9. **Comparison of estimated photon emission rates and concentration with FCS and PCH. (a)** Targeted experimental fluorescent intensity trace. The time step is 10 μs with a total time of 5 min. **(b)** PCH curve and the theoretical fit. **(c)** FCS curve and the best theoretical fit. **(d)** The portion of the trace analyzed by our method rebinned at 100 μs. **(e)** The concentration of Cy3 in the effective volume with ℓ =1, arising from the trace in (d). The experimental concentration is shown by the green line and the PCH estimated is shown by the pink line. **(g)** The posterior probability distribution of the molecular brightness with the PCH estimated of the molecular brightness shown by a solid green line. **(h)** The posterior probability distribution of the background photon emission rate with the PCH estimate of the background photon emission rate shown by a solid green line. **(f)** The posterior probability distribution of the diffusion coefficient obtained by analyzing the trace in (d). The FCS estimate of the diffusion coefficient obtained by analyzing the total time trace, shown in (a), is denoted by a pink solid line. The targeted experimental trace is generated by free diffusive Cy3 in a mixture of water and glycerol with 75% glycerol, a laser power of 100 μW and a concentration of Cy3 at 1 nM, excitation wavelength, NA and refractive index used are 532 nm, 1.42 and 1.4, respectively.

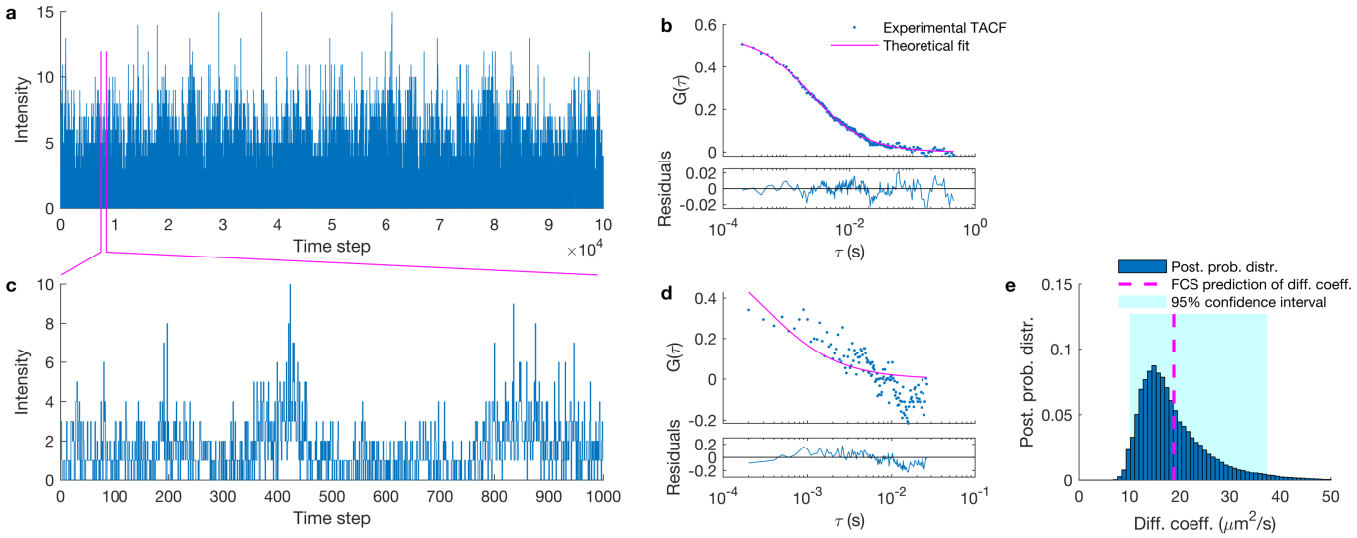

FIG. S10. **Testing diffusion coefficient estimates in experimental traces of free Cy3B dyes using an elongated confocal volume. (a)** Experimental fluorescent intensity trace used in FCS. The time trace is generated by 2.5 nM Cy3B dyes in glycerol/water mixture with 70% glycerol and laser power of 100 μW. **(b)** Auto-correlation curve of the trace in (a) and best theoretical fit. **(c)** Portion of the trace in (a) to be used as the input to FCS and our method. **(d)** Auto-correlation curve of trace in (c). **(e)** Posterior probability distribution over diffusion coefficient estimated from the trace in (c). Traces shown in (a) and (c) are acquired at 100 μs for a total of 10 second and 0.1 second respectively. The laser power use to generate the signal (a) is 100 μW measured before the beam enters the objective. The estimation of the diffusion coefficient as the results of autocorrelation fitting in (a) matched with Stokes-Einstein prediction, equal to 18.79 $\mu m^2/s$ and in (d) is 145.75 $\mu m^2/s$.

61 ## S1.3. Analysis of additional data for multiple diffusive species

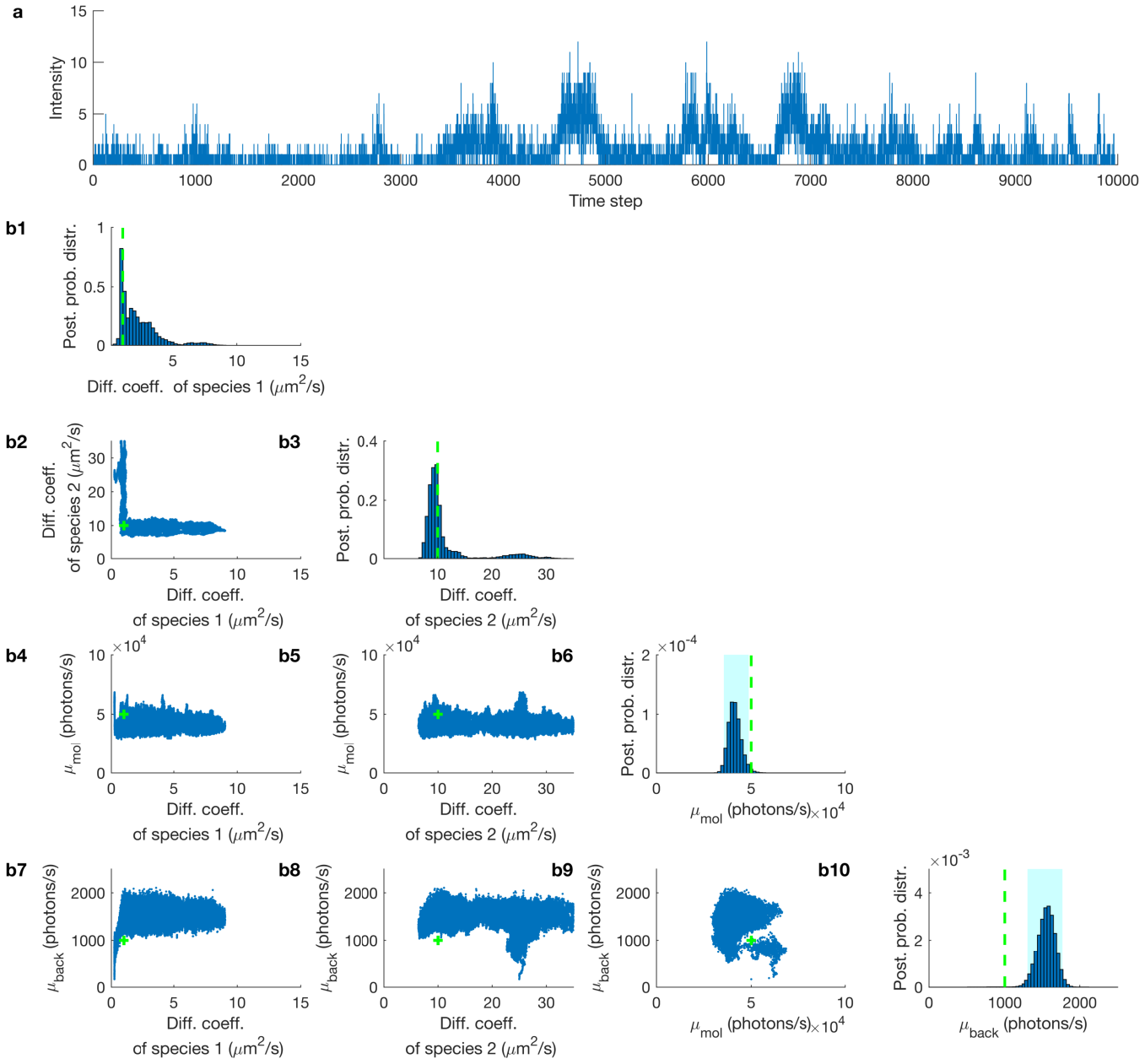

FIG. S11. **Estimated joint posterior probability distribution of multiple diffusive species**. **(a)** A mixed fluorescent intensity trace was obtained by combining the traces from two different synthetic signals with molecular brightness and background emission rates of 5×10^4 and 10^3 photons/s, respectively, and diffusion coefficients of 1 and 10 $\mu m^2/s$. **(b1)** The posterior probability distribution of the diffusion coefficient of diffusive species 1. **(b2)** The joint probability distribution of the diffusion coefficient for diffusive species 1 and diffusive species 2. **(b3)** The posterior probability distribution of the diffusion coefficient of diffusive species 2. **(b4)** The joint probability distribution of diffusion coefficient of diffusive species 1 along with the molecular brightness. **(b5)** The joint probability distribution of diffusion coefficient of diffusive species 2 along with the molecular brightness. **(b6)** The posterior probability distribution of the molecular brightness. **(b7)** The joint probability distribution of diffusion coefficient for diffusive species 1 along with the background photon emission rate. **(b8)** The joint probability distribution of the diffusion coefficient of diffusive species 2 and the background photon emission rate. **(b9)** The joint probability distribution of the molecular brightness and background photon emission rate. **(b10)** The posterior probability distribution of the background photon emission rate. The trace is binned at 100 μs with a total trace duration of 1 s. The exact values of the parameters are shown by green dashed lines and the 95% confidence intervals of the posteriors are highlighted in cyan.

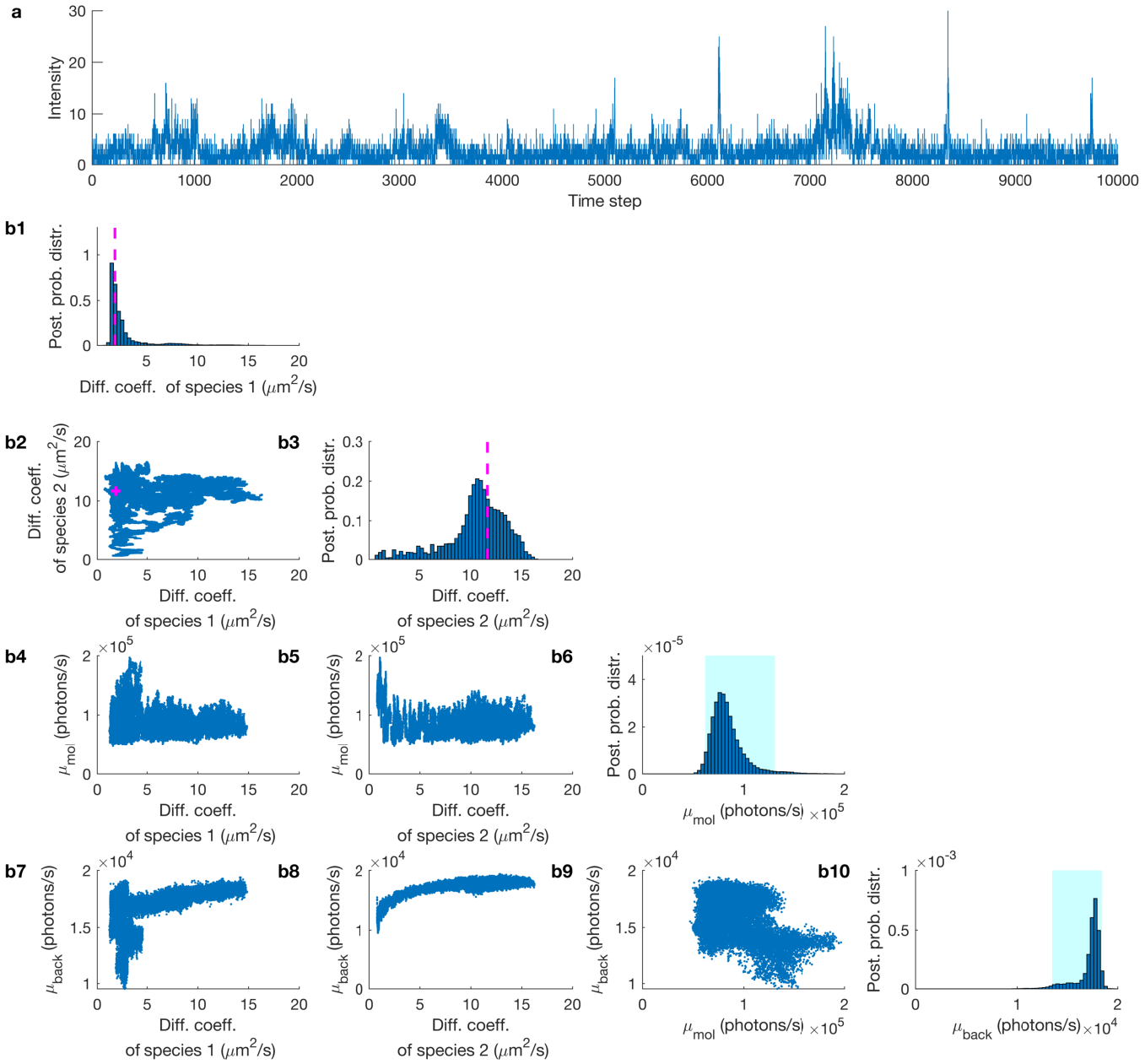

FIG. S12. **Estimated joint posterior probability distribution of multiple diffusive species. (a)** Experimental fluorescent intensity trace used in Fig. 7 a3 with length 10^4 data points and step $100\mu s$. **(b1)** The posterior probability distribution of the diffusion coefficient of diffusive species 1. **(b2)** The joint probability distribution of diffusion coefficient of diffusive species 1 and diffusive species 2. **(b3)** The posterior probability distribution of the diffusion coefficient of diffusive species 2. **(b4)** The joint probability distribution of diffusion coefficient of diffusive species 1 along with the molecular brightness. **(b5)** The joint probability distribution of diffusion coefficient of diffusive species 2 along with the molecule photon emission rate. **(b6)** The posterior probability distribution of the molecular brightness. **(b7)** The joint probability distribution of diffusion coefficient of diffusive species 1 and background photon emission rate. **(b8)** The joint probability distribution of diffusion coefficient of diffusive species 2 and background photon emission rate. **(b9)** The joint probability distribution of the molecular brightness and background photon emission rate. **(b10)** The posterior probability distribution of the background photon emission rate. The trace is generated by mixing two experimental traces of concentration 1 nM of freely diffusive Cy3 in a water/glycerol mixtures with 94% and 75% glycerol each. The laser power, wavelength, NA and refractive index are 100 μW, 532 nm, 1.42 and 1.4, respectively. The FCS estimates are shown by a magenta dashed lines and the 95% confidence intervals of the posteriors are highlighted in cyan.

S2. Summary of point estimates

TABLE S1. Here we list characteristic values (point estimates) summarizing the posterior probability distributions of this study. Mean and std refer to posterior mean value and standard deviation (i.e. square root of variance). Values are listed according to figures.

	D		μ_{mol}		μ_{back}	
	mean	std	mean	std	mean	std
	$\mu m^2/s$	$\mu m^2/s$	photons/s	photons/s	photons/s	photons/s
Fig. 1(b1)	25.07	20.61	5.06×10^4	1.35×10^4	1.52×10^3	0.40×10^3
Fig. 1(b2)	9.21	2.82	5.96×10^4	1.49×10^4	2.32×10^3	0.79×10^3
Fig. 2(a)	1.07×10^{-2}	0.48×10^{-2}	4.69×10^4	1.31×10^4	1.11×10^3	0.79×10^3
	1.23×10^{-1}	0.64×10^{-1}	4.11×10^4	2.01×10^4	1.46×10^3	1.13×10^3
	1.41	0.67	6.53×10^4	1.50×10^4	3.85×10^3	0.96×10^3
	9.27	2.82	5.96×10^4	1.35×10^4	5.63×10^3	1.48×10^3
	180.45	173.25	6.11×10^4	5.18×10^4	3.87×10^3	0.31×10^3
Fig. 2(b)	8.75	7.04	3.64×10^4	8.14×10^3	4.64×10^3	2.17×10^3
	10.22	2.98	3.55×10^4	5.54×10^4	4.16×10^3	1.13×10^3
	9.73	1.97	5.64×10^4	8.36×10^3	6.12×10^3	0.56×10^3
	10.53	1.18	5.08×10^4	4.53×10^3	4.87×10^3	0.37×10^3
	10.05	0.42	4.89×10^4	2.13×10^3	1.35×10^3	0.23×10^3
Fig. 3(a1)	10.65	1.68	9.85×10^4	1.36×10^4	2.17×10^3	0.91×10^3
Fig. 3(a2)	10.04	2.38	5.48×10^4	6.47×10^3	2.09×10^3	0.53×10^3
Fig. 3(a3)	11.29	4.14	1.02×10^4	6.97×10^3	2.66×10^3	0.19×10^3
Fig. 4(c)	10.17	3.09	1.46×10^4	1.42×10^3	2.87×10^3	0.43×10^3
Fig. 5(a)	0.55	0.16	2.18×10^4	1.44×10^4	1.79×10^3	2.21×10^3
	2.28	1.03	3.51×10^4	9.21×10^3	3.55×10^3	1.85×10^3
	12.80	4.51	1.30×10^5	5.17×10^4	1.00×10^4	1.96×10^3
	27.96	10.30	9.15×10^4	2.37×10^4	7.01×10^3	0.46×10^3
Fig. 5(b1)	7.50	6.82	2.20×10^4	9.55×10^3	1.06×10^3	0.22×10^3
Fig. 5(b2)	3.68	2.73	4.32×10^4	2.13×10^4	2.59×10^3	0.52×10^3
Fig. 5(b3)	2.70	2.68	2.31×10^4	5.47×10^3	2.94×10^3	2.14×10^3
Fig. 5(b4)	2.28	1.03	3.51×10^4	9.21×10^3	3.55×10^3	1.85×10^3
Fig. 6(a)	0.36	0.22	7.19×10^3	2.31×10^3	1.42×10^3	1.98×10^3
	4.48	1.35	2.75×10^4	3.03×10^3	0.98×10^3	0.94×10^3
	10.66	4.06	6.61×10^4	1.33×10^4	2.33×10^3	1.51×10^3
	102.26	27.74	2.36×10^5	5.70×10^4	9.33×10^3	0.77×10^3
Fig. 6(b1)	10.86	4.91	3.42×10^4	2.35×10^4	2.59×10^3	0.59×10^3
Fig. 6(b2)	10.54	4.22	5.03×10^4	2.03×10^4	3.78×10^3	0.62×10^3
Fig. 6(b3)	11.76	6.42	3.12×10^4	6.09×10^3	8.41×10^3	1.21×10^3
Fig. 6(b4)	10.66	4.06	6.61×10^4	1.33×10^4	2.33×10^3	1.51×10^3
Fig. 7(a1)	1.98	0.90	2.09×10^5	2.10×10^4	1.06×10^4	1.64×10^3
Fig. 7(a2)	12.65	3.22	4.22×10^4	1.54×10^4	2.57×10^3	0.49×10^3
Fig. 7(a3)	-	-	6.79×10^4	4.58×10^3	1.46×10^4	0.71×10^3
Fig. S4	9.44	3.97	4.62×10^4	1.59×10^4	0.98×10^3	0.25×10^3
Fig. S5(a1)	10.16	2.81	2.71×10^4	3.63×10^3	2.26×10^3	0.62×10^3
Fig. S5(a2)	11.71	4.04	4.89×10^4	1.06×10^4	6.78×10^3	1.79×10^3
Fig. S5(a3)	12.93	2.95	6.05×10^4	1.31×10^4	3.34×10^3	0.98×10^3
Fig. S9	15.73	10.43	1.29×10^5	5.17×10^4	7.21×10^3	1.15×10^3
Fig. S10	18.72	7.18	2.63×10^4	5.93×10^3	7.26×10^3	6.46×10^3
Fig. S11	-	-	4.14×10^4	3.38×10^3	1.55×10^3	0.13×10^3

S3. Detailed methods description

S3.1. Representation of molecular diffusive motion

Consider a particle moving in *1D diffusion*. The probability distribution $p(x,t)$ of the particle's location obeys Fick's second law [1–3] and is given by the diffusion equation

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2} \tag{S1}$$

where D is the particle's diffusion coefficient. Assuming the particle is located at x_{k-1} at a time t_{k-1}, i.e. assuming the initial condition $p(x,t_{k-1}) = \delta(x - x_{k-1})$, and a free space boundary, i.e. $\lim_{x\to\pm\infty} p(x,t) = 0$, we can solve this equation to obtain $p(x,t)$ for any later time t. The solution is

$$p(x,t) = \frac{\exp\left(-\frac{(x-x_{k-1})^2}{4(t-t_{k-1})D}\right)}{\sqrt{4\pi(t-t_{k-1})D}} \tag{S2}$$

which equals to the probability density of a normal random variable with mean x_{k-1} and variance $2(t-t_{k-1})D$, see Table S5. At time $t = t_k$, we therefore have

$$x_k|x_{k-1} \sim \textbf{Normal}\left(x_{k-1}, 2(t_k - t_{k-1})D\right). \tag{S3}$$

Similarly, solving the diffusion equation for particles following isotropic *3D diffusion* in free space, we have

$$\begin{aligned} x_k|x_{k-1} &\sim \textbf{Normal}\left(x_{k-1}, 2(t_k - t_{k-1})D\right) \\ y_k|y_{k-1} &\sim \textbf{Normal}\left(y_{k-1}, 2(t_k - t_{k-1})D\right) \\ z_k|z_{k-1} &\sim \textbf{Normal}\left(z_{k-1}, 2(t_k - t_{k-1})D\right) \end{aligned} \tag{S4}$$

which constitute the molecular motion model used throughout this study.

S3.2. Description of Stokes-Einstein model

For the experimental data, we benchmark our estimates of the diffusion coefficient against the Stokes-Einstein prediction [2, 3]. Namely, for a spherical particle in a quiescent fluid at uniform temperature

$$D = \frac{kT}{6\pi r\eta} \tag{S5}$$

where, D is the diffusion coefficient, k is Boltzmann's constant, T is the solution's absolute temperature, r is the hydrodynamic radius of the particle [4] and η is the solution's dynamic viscosity [5].

S3.3. FCS formulation

The formulation we used in this study to autocorrelate the synthetic and experimental time traces is

$$G_{ex}(\tau) = \frac{\langle \delta I(t+\tau)\delta I(t)\rangle}{\langle \delta I(t)\rangle^2} = \frac{\langle I(t+\tau)I(t)\rangle}{\langle I(t)\rangle^2} - 1 \tag{S6}$$

where the $I(t)$ is the number of detected photons at time t. The computational implementation uses the Wiener-Khinchin Theorem [6].

Also, the theoretical function [7–10] used to fit the autocorrelation curves (using a 3DG PSF) is

$$G_{th}(\tau) = \frac{1}{\langle N\rangle}\frac{1 - F + Fe^{-\frac{\tau}{\tau_F}}}{(1-F)}\frac{1}{1+\frac{4D\tau}{\omega_{xy}^2}}\frac{1}{(1+\frac{4D\tau}{\omega_z^2})^{\frac{1}{2}}} \tag{S7}$$

and for the 2DGL PSFs is

$$G_{th}(\tau) = \frac{1}{\langle N \rangle} \frac{1 - F + Fe^{-\frac{\tau}{\tau_F}}}{(1-F)} \frac{1}{1 + \frac{4D\tau}{\omega_{xy}^2}} + 1 \tag{S8}$$

where, $\langle N \rangle$ is the average number of molecule in the effective volume, D is the diffusion coefficient, τ_F is the triplet state relaxation time and F is the fraction molecules populating the triplet state.

To find the best fit, we use χ^2 minimization

$$\chi^2 = \sum_{\tau} (G_{th}(\tau) - G_{ex}(\tau))^2. \tag{S9}$$

S3.4. Definition of molecular brightness

As the definition of molecular brightness, we use the *emission rate of detected photons* of a single fluorophore, for example Eq. (2). For a fluorophore located at (x, y, z) this is formulated as

$$\mu(x, y, z) = \mu_0 \; \varphi_d \; \varphi_{qe} \; \varphi_f \; \sigma \; \mathrm{EXC}(x, y, z) \; \mathrm{CEF}(x, y, z) \tag{S10}$$

where, μ_0 is the maximum excitation intensity which occurs at the center of the confocal volume, φ_d is the efficiency of the photon collection at the center of the confocal volume, φ_{qe} is the quantum efficiency of the detector, φ_f is the quantum efficiency of the fluorophore (i.e. quantum yield), σ is the absorption cross-section of the fluorophore, $\mathrm{EXC}(x, y, z)$ is the excitation profile and $\mathrm{CEF}(x, y, z)$ is the detection profile, i.e. collection efficiency function, which equals the fraction of the detected photons to the total photons emitted by a point source [11].

To obtain Eq. (2), we cast Eq. (S10) in the simplified form

$$\mu(x, y, z) = \mu_{mol} \mathrm{PSF}(x, y, z) \tag{S11}$$

where $\mu_{mol} = \mu_0 \, \varphi_d \, \varphi_{qe} \, \varphi_f \, \sigma$, which we term *molecular brightness* at the center of the confocal volume [12], and $\mathrm{PSF}(x, y, z) = \mathrm{EXC}(x, y, z) \, \mathrm{CEF}(x, y, z)$, which we term the PSFPSF.

To relate the parameter μ_{mol} to the *average photon count rate*, which is commonly estimated in bulk experiments [13**?** , 14], we consider the spatial average of $\mu(x, y, z)$ as follows

$$\langle \mu(x, y, z) \rangle = \mu_{mol} \langle \mathrm{PSF}(x, y, z) \rangle. \tag{S12}$$

For the specific choice of a 3DG PSF (see below), the average is computed as follows

$$\langle \mathrm{PSF}(x, y, z) \rangle = \frac{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp\left(-2\frac{x^2}{\omega_{xy}^2} - 2\frac{y^2}{\omega_{xy}^2} - 2\frac{z^2}{\omega_z^2}\right) dxdydz}{V_{eff}} = \sqrt{\frac{\pi}{2}\omega_{xy}^2} \sqrt{\frac{\pi}{2}\omega_{xy}^2} \sqrt{\frac{\pi}{2}\omega_z^2} \frac{1}{V_{eff}} \tag{S13}$$

where V_{eff} denotes the effective volume of 3DG PSF [9, 15] and it is given by

$$V_{eff} = \pi^{\frac{3}{2}} \omega_{xy}^2 \omega_z. \tag{S14}$$

Consequently, Eq. (S13) implies

$$\mu_{mol} = \sqrt{8} \, \langle \mu(x, y, z) \rangle. \tag{S15}$$

In other words, the molecular brightness is, by definition, approximately 2.8 times larger than the average photon count rate of single molecule [13**?** , 14].

S3.5. Definition of point spread function models

In this study we use three different point spread functions as approximations to the more realistic Airy function [16–18], namely a 3D-Gaussian (3DG) [19], a 2D-Gaussian-Cylindrical (2DGC) [19] and a 2D-Gaussian-Lorentzian (2DGL) [20–23].

The definition of the PSF for the 3DG case is

$$\mathrm{PSF}_{\mathrm{3DG}}(x,y,z) = \exp\left(-2\frac{x^2+y^2}{\omega_{xy}^2} - 2\frac{z^2}{\omega_z^2}\right) \tag{S16}$$

while, the definition of the PSF for the 2DGC case is

$$\mathrm{PSF}_{\mathrm{2DGC}}(x,y,z) = \exp\left(-2\frac{x^2}{\omega_{xy}^2} - 2\frac{y^2}{\omega_{xy}^2}\right). \tag{S17}$$

For both cases, ω_{xy} and ω_z are the semi-axes lateral and parallel to the optical axis. These are represented in terms of the excitation wavelength λ_{exc}, solution refraction index n_{sol}, and numerical aperture NA of the microscope as $\omega_{xy} = 0.61\lambda_{\mathrm{exc}}/\mathrm{NA}$ and $\omega_z = 1.5n_{\mathrm{sol}}\lambda_{\mathrm{exc}}/\mathrm{NA}^2$; for example see [24, 25]. For more realistic representations, ω_{xy} and ω_z can be estimated directly based on calibration experiments with known diffusion coefficients; for example see [26].

The definition of the PSF for the 2DGL case is

$$\mathrm{PSF}_{\mathrm{2DGL}}(x,y,z) = \frac{1}{1+\left(\frac{z}{z_R}\right)^2}\exp\left(\frac{-2\frac{x^2+y^2}{\omega_{xy}^2}}{1+\left(\frac{z}{z_R}\right)^2}\right) \tag{S18}$$

where ω_{xy}, λ_{exc}, and n_{sol} are similar to the 3DG cas or 2DG cases and $z_R = n_{\mathrm{sol}}\pi\omega_{xy}^2/\lambda_{\mathrm{exc}}$.

S3.6. Description of the data simulation

To generate fluorescence intensity time traces that mimic a realistic confocal setup, we simulate molecules moving [1?] through an illuminated 3D volume. The number of moving molecules N is prescribed in each simulation. To maintain a relatively stable concentration of molecules near the confocal volume, and so to avoid generating traces where every molecule eventually strays into un-illuminated regions, we impose periodic rectangular boundaries to our volume. The boundaries are placed at $\pm L_{xy}$ perpendicular to the focal plane and $\pm L_z$ perpendicular to the optical axis.

We assess the locations of the molecules x_k^n, y_k^n, z_k^n, where $k = 1,\ldots,K$ label time levels and $n = 1,\ldots,N$ label molecules, at equidistant time intervals $t_1, t_2, \ldots, t_K$. The time interval between successive assessments $\delta t = t_k - t_{k-1}$, as well as the total trace duration $T_{total} = t_K - t_0$, are prescribed.

Molecule locations at the first assessment x_1^n, y_1^n, z_1^n are sampled randomly from a uniform distribution with limits equal to the boundaries $\pm L_{xy}$ and $\pm L_z$ of our pre-specified volume. Subsequent locations are generated according to the diffusion model described above under a prescribed diffusion coefficient D.

Finally, we obtain individual photon emissions w_k by simulating Bernoulli random variables of success probability $q_k = 1 - e^{-\mu_k \delta t}$, where the rate μ_k gathers single photon contributions from the background and the entire molecule population according to

$$\mu_k = \mu_{back} + \mu_{mol}\sum_{n=1}^{N}\mathrm{PSF}(x_k^n, y,_k^n z_k^n) \tag{S19}$$

where both background and molecular brightness, μ_{back} and μ_{mol}, are prescribed.

The PSF model is also prescribed. To avoid artifacts induced by the periodic boundaries we impose in our volume, we ensure that $L_{xy} \gg \omega_{xy}$, $L_z \gg \omega_z$, or $L_z \gg z_R$, where ω_{xy}, ω_z and z_R characterize the geometry of the confocal volume, see Eqs. (S16)–(S18), above.

Detailed parameter choices for all simulations performed are listed in Table S6.

S3.7. Definition of normalized distance and numbers of molecules

As we need to estimate the positions of the molecules with respect to the center of the confocal volume, which is the point of origin, in order to ultimately estimate the number of molecules as a proxy for molecule concentration, for example Figs. S3 and S8, we must address difficulties associated with symmetries of the confocal PSF with respect to

rotations around the optical axis or the focal plane. [27] For this, for a molecule at (x_k^n, y_k^n, z_k^n), when the 3DG PSF is used, Eq. (S16), we rely on

$$d_k^n = \sqrt{\left(\frac{x_k^n}{\omega_{xy}}\right)^2 + \left(\frac{y_k^n}{\omega_{xy}}\right)^2 + \left(\frac{z_k^n}{\omega_z}\right)^2} \tag{S20}$$

while, when the 2DGL PSF is used, Eq. (S18), we rely on

$$d_k^n = \sqrt{\frac{1}{2}\log\left(1 + \left(\frac{z_k^n}{z_R}\right)^2\right) + \frac{\left(\frac{x_k^n}{\omega_{xy}}\right)^2 + \left(\frac{y_k^n}{\omega_{xy}}\right)^2}{1 + \left(\frac{z_k^n}{z_R}\right)^2}} \tag{S21}$$

where d_k^n is the normalized distance with respect to the center of the confocal volume of molecule n at time k. Similarly, when a 2DGC PSF is used, Eq. (S17), we rely on

$$d_k^n = \sqrt{\left(\frac{x_k^n}{\omega_{xy}}\right)^2 + \left(\frac{y_k^n}{\omega_{xy}}\right)^2} \tag{S22}$$

where d_k^n is the normalized distance with respect to the optical axis of molecule n at time k.

These distances are obtained by setting the respective PSFs equal to $\exp(-(d_k^n)^2)$ and are unaffected by the aforementioned symmetries, i.e. $x_k^n \mapsto -x_k^n$, $y_k^n \mapsto -y_k^n$, and $z_k^n \mapsto -z_k^n$.

For a given normalized distance ℓ, we define the number of molecules N_k^ℓ as the number of estimated (active) molecules within the corresponding distance. That is

$$N_k^\ell = \sum_n b^n H\left(1 - \frac{d_k^n}{\ell}\right) \tag{S23}$$

where H is the Heaviside step function, b^n is the load of molecule n, and V_ℓ is the volume of a designated effective region chosen to agree with the effective volume V_{eff} used in FCS [9].

S3.8. Description of the time trace preparation

The initial time trace consists of single photon arrival times which are computationally too expensive to analyze. Our method instead operates on photon intensity traces which are either obtained directly during an experiment or obtained from individual photon arrival time traces after binning. To transform single photon arrival time traces into intensity time traces, we use time bins of fixed size (main size) that typically span multiple photon arrival times. To speed up the computations further, as some bins have none of very few photons, over certain portions of the trace we use larger bins (auxiliary size).

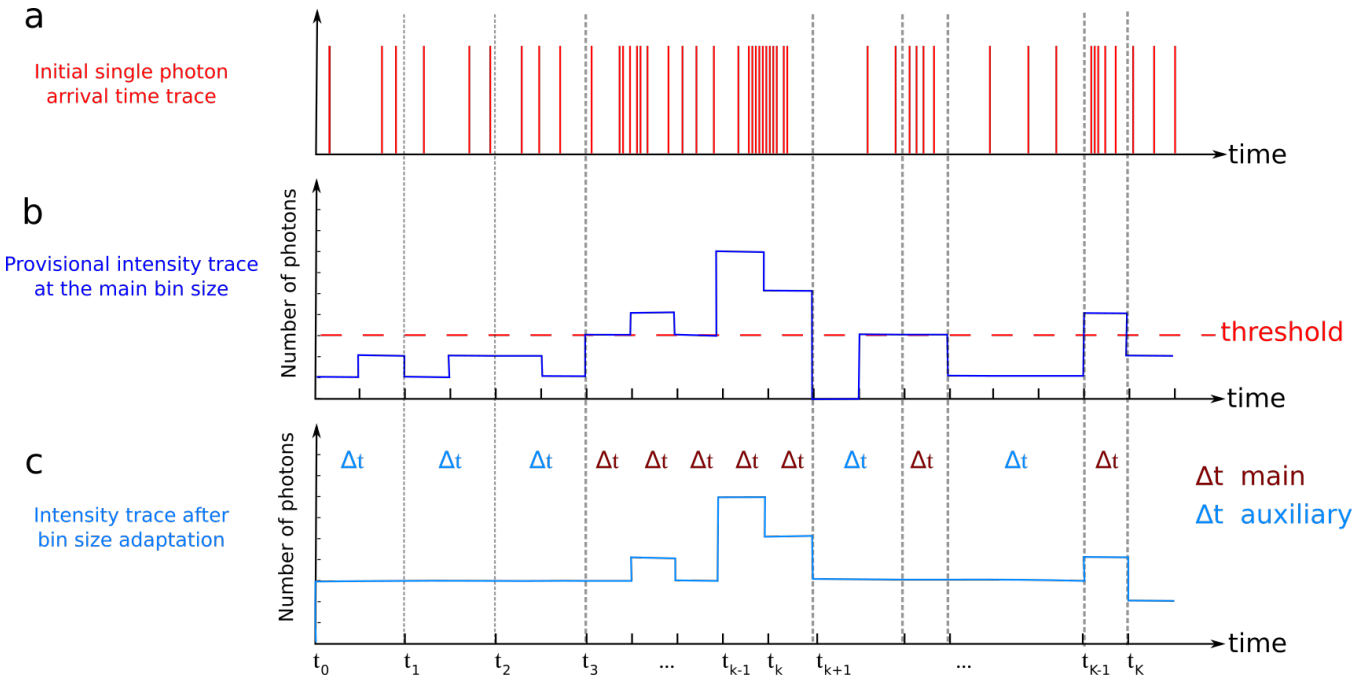

FIG. S13. **Illustration of time trace preparation.** **(a)** Initial trace of single photon arrivals. Here, each vertical line represents the arrival time of a single photon. **(b)** Time trace of photon intensities provisionally binned at the main bin size. The horizontal line denotes the imposed lower threshold on the minimum number of photons in the individual bins. **(c)** Time trace of photon intensities after bin size adaptation. Here, bins, preselected at the main size, with intensities below the imposed threshold are uniformly readjusted to achieve an average intensity similar to the threshold.

Briefly, the user specifies a minimum number of photons per bin as a lower threshold. As illustrated in Fig. S13, those bins, preselected at the main size, containing fewer photons than the specified threshold are enlarged uniformly in order to achieve an average of at least as many photons as specified by the threshold. This occasional adaptation, from the main to the auxiliary bins, becomes important in the analysis of traces from experiments held near single molecule resolution where molecule concentrations are low so that on average only one molecular passage through the confocal volume happens. Consequently, photon intensities are low, and thus the bulk of computational time otherwise would had been spent processing trace portions of poor quality (i.e. with few or no photons).

To carry our the necessary computations, as we detail shortly, we use the Anscombe transformation [28] to approximate the Poissonian likelihoods of photon intensities (see below). This approximation is robust as long as bins contain on average 4 photons or more. Thus, as a minimum requirement, we also use the aforementioned threshold to ensure the validity of the approximations.

S4. Detailed description of the inference framework

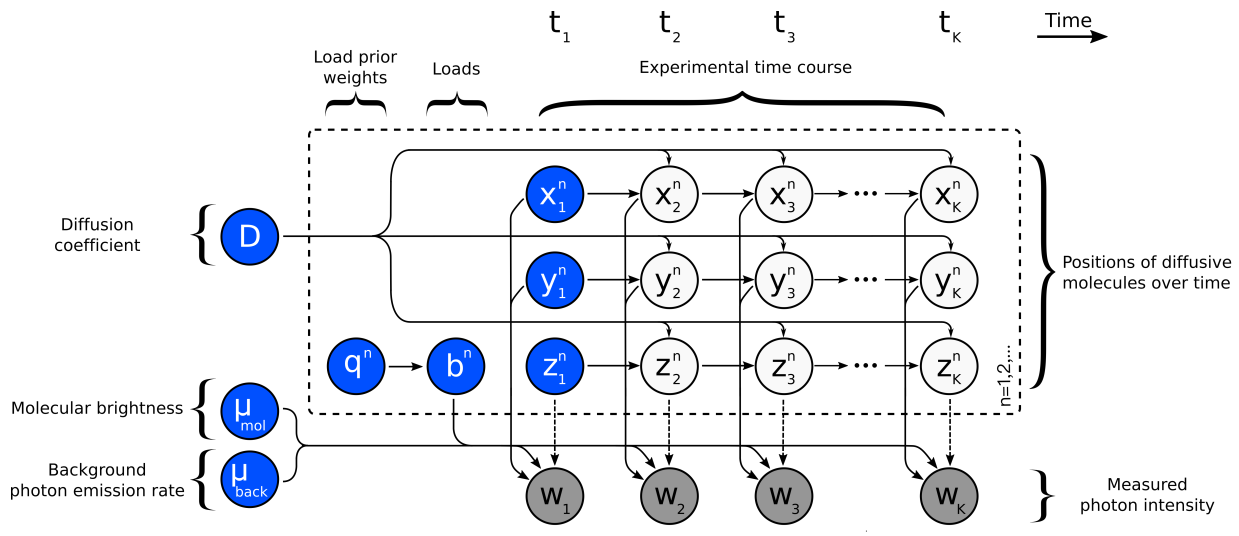

FIG. S14. **Graphical summary of the framework developed.** A population of model molecules, labeled by $n = 1, 2, \ldots,$ evolves during the measurement period which is marked by $k = 1, 2, \ldots, K$. Here, x_k^n, y_k^n and z_k^n denote the position of molecule n at time t_k and μ_{mol}, μ_{back} denote the molecular brightness and background photon emission rates. The diffusion coefficient D determines the evolution of the molecule locations which, in turn, determines the instantaneous photon emission rates and ultimately the recorded photon intensities w_k. Load variables b^n, with prior weights q^n, are introduced to model a molecule population of *a priori* unknown size. Following common machine learning convention, the measurements w_k are dark shaded. Additionally, model variables requiring prior probability distributions are highlighted in blue.

S4.1. Description of prior probability distributions

The model parameters in our framework that require priors are: the diffusion coefficient D; the molecular brightness and background photon emission rates μ_{mol} and μ_{back}; the initial molecule locations x_1^n,y_1^n,z_1^n; and load prior weights q^n. As we already mentioned in the main text, a prior on the population of diffusing molecules is implicitly defined by the prior on both b^n and q^n. Our choices are described below.

S4.1.1. Prior on the diffusion coefficient

To ensure that the D sampled in our formulation attains only positive values, we place an inverse-Gamma prior

$$D \sim \textbf{InvGamma}\left(\alpha_D, \beta_D\right). \tag{S24}$$

Besides ensuring a positive D, this prior is also conjugate to the motion model we use which facilitates the computations (see below).

S4.1.2. Priors on molecular brightness and background photon emission rates

To ensure that μ_{mol} and μ_{back} sampled in our formulation attain only positive values, we place Gamma priors on both

$$\begin{aligned} \mu_{mol} &\sim \textbf{Gamma}\left(\alpha_{mol}, \beta_{mol}\right) \\ \mu_{back} &\sim \textbf{Gamma}\left(\alpha_{back}, \beta_{back}\right). \end{aligned} \tag{S25}$$

Due to the specific dependencies of the likelihood (that we will discuss shortly) on the photon emission rates, conjugate priors cannot be achieved for μ_{mol} and μ_{back}. So, the above choice offers no computational advantage (see below) and

could be readily replaced with more physically motivated choices if additional information on molecular brightness becomes available.

S4.1.3. Priors on initial molecule locations

Due to the symmetries in the confocal PSF, i.e. a molecule at a location (x, y, z) emits the same average number of photons as a molecule at locations $(\pm x, \pm y, \pm z)$, we are unable to gain insight regarding the octant of the 3D Cartesian space in which each molecule is located. To avoid imposing further assumptions on our framework that may determine each molecule's octant uniquely, but may limit the framework's scope to specific experimental setups, we place priors on the initial locations that respect these symmetries. Accordingly, in our framework, at the onset of the measuring period, molecules are equally likely to be located at any of the positions $(\pm x_1^n, \pm y_1^n, \pm z_1^n)$.

To facilitate the computations (see below), we place independent symmetric normal distributions, see Table S5, on each Cartesian coordinate of the model molecules

$$\begin{aligned} x_1^n &\sim \textbf{SymNormal}\left(\mu_{xy}, \sigma_{xy}^2\right) \\ y_1^n &\sim \textbf{SymNormal}\left(\mu_{xy}, \sigma_{xy}^2\right) \\ z_1^n &\sim \textbf{SymNormal}\left(\mu_z, \sigma_z^2\right). \end{aligned} \tag{S26}$$

We want to emphasize that the symmetric priors above do not affect our estimates. According to the motion model we employ, no matter where molecules are initiated, they may subsequently move freely and eventually switch to a different octant if warranted by the data. Our symmetric priors merely indicate that for each individual molecular trajectory considered, there are another 7 symmetric trajectories that are equally likely to have occurred.

S4.1.4. Priors and hyperpriors for molecule loads

To facilitate the computations (described next), we use a finite, but large, model population consisting of N molecules containing both active and inactive molecules. These model molecules are collectively indexed by $n = 1, 2, \ldots, N$. As explained in the main text, estimating how many molecules are actually warranted by the data under analysis is equivalent to estimating how many of those N molecules are active, i.e. $b^n = 1$, while the remaining inactive ones, i.e. $b^n = 0$, have no impact whatsoever and are instantiated only for computational purposes.

To ensure that each load b^n takes only values 0 or 1, we place a Bernoulli prior of weight q^n. In turn, on each weight q^n, we place a conjugate Beta hyperprior

$$b^n | q^n \sim \textbf{Bernoulli}(q^n) \tag{S27}$$

$$q^n \sim \textbf{Beta}\left(A_q, B_q\right). \tag{S28}$$

To ensure that the resulting formulation avoids overfitting, we make the specific choices $A_q = \alpha_q/N$ and $B_q = \beta_q(N-1)/N$. Under these choices [29–32], and in the limit that $N \to \infty$; that is, when the assumed molecule population is allowed to be large, this prior/hyperprior converge to a Beta-Bernoulli process. Consequently, for $N \gg 1$, the posterior remains well defined and becomes independent of the chosen value of N. In other words, provided N is large enough, its impact on the results is insignificant; while its precise value has only computational implications (see below).

S4.2. Summary of model equations

For concreteness, below we summarize the entire set of equations used in our framework, including a complete list of priors and hyperpriors

$$D \sim \textbf{InvGamma}\left(\alpha_D, \beta_D\right) \tag{S29}$$

$$\mu_{mol} \sim \textbf{Gamma}\left(\alpha_{mol}, \beta_{mol}\right) \tag{S30}$$
$$\mu_{back} \sim \textbf{Gamma}\left(\alpha_{back}, \beta_{back}\right) \tag{S31}$$

$$q^n \sim \textbf{Beta}\left(\frac{\alpha_q}{N}, \beta_q \frac{N-1}{N}\right) \tag{S32}$$
$$b^n | q^n \sim \textbf{Bernoulli}(q^n) \tag{S33}$$

$$x_1^n \sim \textbf{SymNormal}\left(\mu_{xy}, \sigma_{xy}^2\right) \tag{S34}$$
$$y_1^n \sim \textbf{SymNormal}\left(\mu_{xy}, \sigma_{xy}^2\right) \tag{S35}$$
$$z_1^n \sim \textbf{SymNormal}\left(\mu_z, \sigma_z^2\right) \tag{S36}$$

$$x_k^n | x_{k-1}^n, D \sim \textbf{Normal}\left(x_{k-1}^n, 2(t_k - t_{k-1})D\right), \qquad k = 2, \ldots, K \tag{S37}$$
$$y_k^n | y_{k-1}^n, D \sim \textbf{Normal}\left(y_{k-1}^n, 2(t_k - t_{k-1})D\right), \qquad k = 2, \ldots, K \tag{S38}$$
$$z_k^n | z_{k-1}^n, D \sim \textbf{Normal}\left(z_{k-1}^n, 2(t_k - t_{k-1})D\right), \qquad k = 2, \ldots, K \tag{S39}$$

$$w_k | \{x_k^n, y_k^n, z_k^n, b^n\}_n, \mu_{mol}, \mu_{back} \sim \textbf{Poisson}\left(\mu_k\right), \qquad k = 1, \ldots, K \tag{S40}$$
$$\mu_k = (t_k - t_{k-1}) \left(\mu_{back} + \mu_{mol} \sum_n b^n \ \mathrm{PSF}(x_k^n, y_k^n, z_k^n) \right). \tag{S41}$$

For molecules diffusing in a confocal volume that is extremely elongated over the optical axis, the PSF approaches a cylindrical one. In this case, it is safe to eliminate the z_k^n positions from the motion model and simplify Eqs. (S40) and (S41) to

$$w_k | \{x_k^n, y_k^n, b^n\}_n, \mu_{mol}, \mu_{back} \sim \textbf{Poisson}\left(\mu_k\right), \qquad k = 1, \ldots, K \tag{S42}$$
$$\mu_k = (t_k - t_{k-1}) \left(\mu_{back} + \mu_{mol} \sum_n b^n \ \mathrm{PSF}(x_k^n, y_k^n) \right). \tag{S43}$$

S4.3. Description of the computational scheme

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

$$\overline{x}^n = (x_1^n, x_2^n, \ldots, x_K^n) \tag{S44}$$
$$\overline{y}^n = (y_1^n, y_2^n, \ldots, y_K^n) \tag{S45}$$
$$\overline{z}^n = (z_1^n, z_2^n, \ldots, z_K^n) \tag{S46}$$
$$\overline{w} = (w_1, w_2, \ldots, w_K). \tag{S47}$$

Due to the non-linearities in the PSF and the non-parametric prior on q^n and b^n, analytic evaluation or direct sampling of this posterior is impossible. For this reason, we develop a specialized Markov chain Monte Carlo (MCMC) scheme that can be used to generate pseudo-random samples [33–37]. This scheme is explained in detail below.

In order to terminate the MCMC sampler, we need to determine when a representative number of samples has been computed. To do so, we divide the samples already computed into four portions and compare the mean values of the diffusion coefficient of the two last ones

$$\eta_1 = \frac{\sum_{i=2I/4}^{3I/4} D_i}{I/4}, \qquad \eta_2 = \frac{\sum_{i=3I/4}^{I} D_i}{I/4} \tag{S48}$$

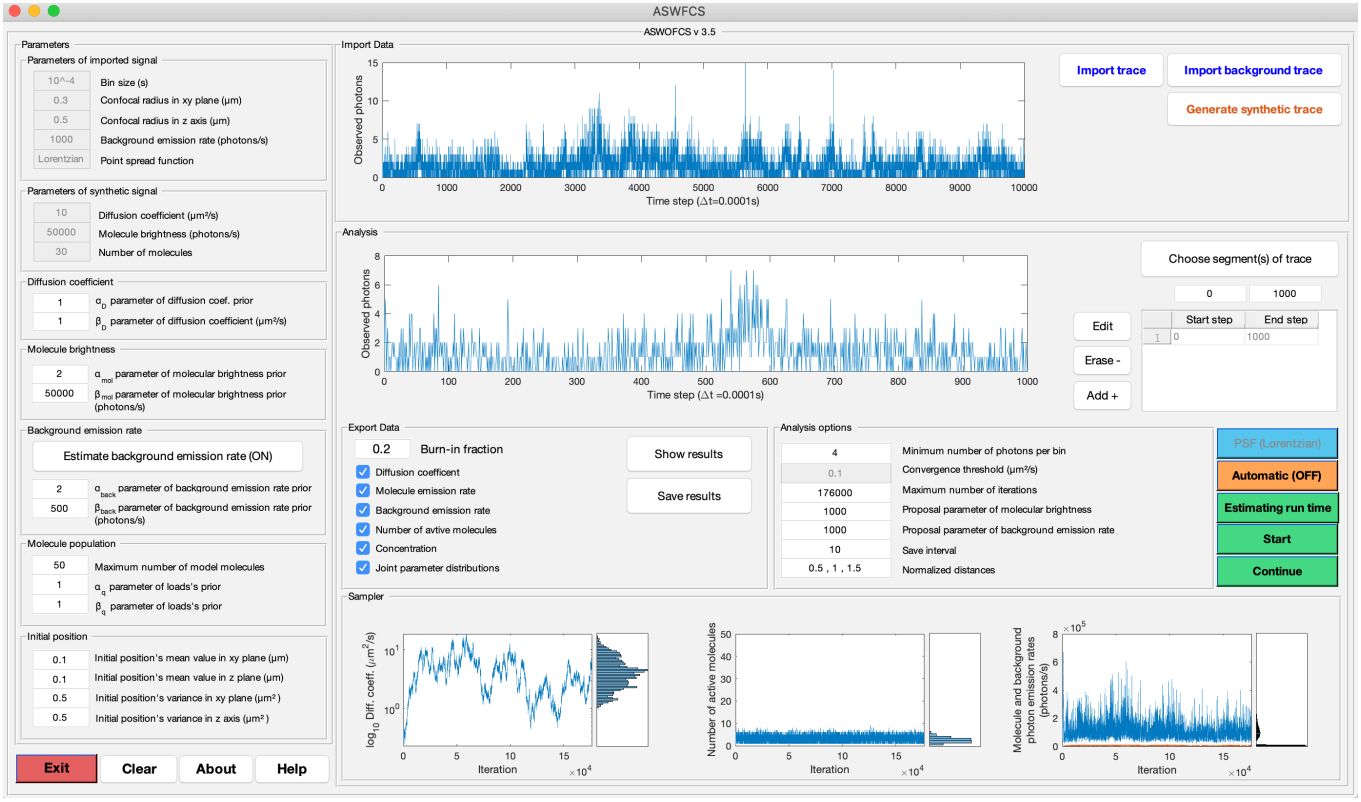

FIG. S15. **A working implementation of the framework described in this study is available through the** SUPPORTING MATERIAL**.** Along with this implementation, we provide a graphical user interface (GUI) that can be used to analyze intensity traces from confocal microscopy.

where, η_1 and η_2 are the mean values of the two last portion of the sampled diffusion coefficients denoted D_i and I is the total number of computed MCMC samples thus far. Following [33, 34], we terminate the sampler when $|\eta_1 - \eta_2| < \epsilon_{thr}$, where ϵ_{thr} is a pre-specified threshold. Also, to avoid incorporating burn-in samples in the calculations, we ensure a minimum number of iterations I of no less than 10^4.

A working implementation of the resulting scheme in source code and GUI forms, see Fig. S15, are available through the SUPPORTING MATERIAL.

S4.3.1. Overview of the sampling updates

Our MCMC exploits a Gibbs sampling scheme [33–35]. Accordingly, posterior samples are generated by updating each one of the variables involved sequentially by sampling conditioned on all other variables and measurements $\overline{w}$.

Conceptually, the steps involved in the generation of each posterior sample $(D, \mu_{mol}, \mu_{back}, \{q^n, b^n, \overline{x}^n, \overline{y}^n, \overline{z}^n\}_n)$ are:

(1) For each n in the molecule population
  (i) Update trajectory $\overline{x}^n$ of molecule n
  (ii) Update trajectory $\overline{y}^n$ of molecule n
  (iii) Update trajectory $\overline{z}^n$ of molecule n
(2) Update the diffusion coefficient D
(3) Update jointly the prior weights q^n for all molecules
(4) Update jointly the loads b^n for all molecules
(5) Update jointly the molecular brightness and background photon emission rates μ_{mol} and μ_{back}, respectively

Since the locations of the inactive molecules are not associated with the measurements $\overline{w}$, see Fig. S14, and those are

updated independently of the locations of the active ones, to make the algorithm computationally more efficient we carry out the above scheme in the equivalent order

(1) For each n of the *active* molecules

(i) Update trajectory $\overline{x}^n$ of active molecule n

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

(iii) 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

(5) Update jointly the loads b^n for all model molecules

(6) Update jointly the molecular brightness and background photon emission rates μ_{mol} and μ_{back}, respectively

These steps are described in detail below.

S4.3.2. Sampling of active molecule trajectories

For a given active molecule n, we update the trajectory $\overline{x}^n$ by sampling from the corresponding conditional $p(\overline{x}^n|D, \mu_{mol}, \mu_{back}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \overline{w})$, which we achieved through backward sampling [38–40]. To be able to sample a trajectory $\overline{x}^n$ in backward sampling, we factorize the density $p(\overline{x}^n|D, \mu_{mol}, \mu_{back}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \overline{w})$ as

$$\begin{aligned} & p(\overline{x}^n|D, \mu_{mol}, \mu_{back}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \overline{w}) \\ &= p(x_1^n|x_2^n, D, \mu_{mol}, \mu_{back}, \{b^{n'}, y_1^{n'}, z_1^{n'}\}_{n'}, \{x_1^{n'}\}_{n'\neq n}, \overline{w}) \\ &\times p(x_2^n|x_3^n, D, \mu_{mol}, \mu_{back}, \{b^{n'}, y_2^{n'}, z_2^{n'}\}_{n'}, \{x_2^{n'}\}_{n'\neq n}, \overline{w}) \\ &\dots \\ &\times p(x_{K-1}^n|x_K^n, D, \mu_{mol}, \mu_{back}, \{b^{n'}, y_{K-1}^{n'}, z_{K-1}^{n'}\}_{n'}, \{x_{K-1}^{n'}\}_{n'\neq n}, \overline{w}) \\ &\times p(x_K^n|D, \mu_{mol}, \mu_{back}, \{b^{n'}, y_K^{n'}, z_K^{n'}\}_{n'}, \{x_K^{n'}\}_{n'\neq n}, \overline{w}). \end{aligned} \tag{S49}$$

According to this factorization, we sample $\overline{x}^n$, starting from x_K^n and move backward towards x_1^n. To start the sampling steps, we need to determine each one of the individual densities $p(x_K^n|D, \mu_{mol}, \mu_{back}, \{b^{n'}, y_K^{n'}, z_K^{n'}\}_{n'}, \{x_K^{n'}\}_{n'\neq n}, \overline{w})$ and $p(x_k^n|x_{k+1}^n, D, \mu_{mol}, \mu_{back}, \{b^{n'}, y_k^{n'}, z_k^{n'}\}_{n'}, \{x_k^{n'}\}_{n'\neq n}, \overline{w})$. We do this in a forward filtering approach [27, 38–42] which is described in detail below.

S4.3.2.a. Forward filtering

By applying Bayes' rule, each one of the individual densities in Eq. (S49) factorizes as

$$p(x_k^n|x_{k+1}^n, D, \dots, \overline{w}) \propto p(x_{k+1}^n|x_k^n, D)p(x_k^n|D, \dots, w_{1:k}) \tag{S50}$$

where $w_{1:k}$ is an abbreviation for $w_1, \dots, w_k$ and excess parameters are shown by dots. Since the density $p(x_{k+1}^n|x_k^n, D)$ is already known, to sample x_k^n in backward sampling, we only need to determine the filter density $p(x_k^n|D, \dots, w_{1:k})$.

To be able to apply forward filtering and compute $p(x_k^n|D, \dots, w_{1:k})$ efficiently [43], we use an approximate model [44], where Eq. (S40), is replaced with

$$\mathbb{T}_{data}(w_k)|\{x_k^{n'}, y_k^{n'}, z_k^{n'}, b^{n'}\}_{n'}, \mu_{mol} \sim \textbf{Normal}\left(\mathbb{T}_{mean}(\mu_k), 1\right), \qquad k = 1, \dots, K. \tag{S51}$$

Here, μ_k stems from Eq. (S41) for 3D models and Eq. (S43) for 2D models; while, $\mathbb{T}_{data}(w)$ and $\mathbb{T}_{mean}(\mu)$ denote Anscombe transformed [28] variables defined as follows

$$\mathbb{T}_{data}(w) = 2\sqrt{w + \frac{3}{8}} \tag{S52}$$

$$\mathbb{T}_{mean}(\mu) = 2\sqrt{\mu + \frac{3}{8}} - \frac{1}{4\sqrt{\mu}}. \tag{S53}$$

The Anscombe transform exploited here offers a way of transforming Poisson random variables into (approximately) normal ones [28] which facilitates the filtering process described next. The approximation we employ is highly accurate for $\mu \gg 1$, while acceptable accuracy is maintained so long as $\mu > 4$ photons.

Under the Anscombe transform, the densities of both the dynamics, Eq. (S37), and observations, Eq. (S51), are normally distributed. So, we can compute the filter distribution $p(x_k^n|D,\ldots,w_{1:k})$ of the approximate model similar to the standard theory underlying nonlinear Kalman filters [27, 45–51].

More specifically, because the mean of the transformed observation distribution, $T_{mean}(\mu_k)$ is a nonlinear function of the location x_k^n, to apply the Kalman filters we need to approximate the transformed observation distribution in such a way that its mean becomes a linear function of the location x_k^n. To do so, we use two common approaches: (i) extended Kalman filter (EKF) [45, 52–55], which locally approximate the transformed observation distribution around selected values; and (ii) unscented Kalman filter (UKF) [46–48], which globally approximate the transformed observation distribution.

As explained in detail in [27], the linearization alone is not sufficient to properly approximate the filter. This is because both EKF and UKF assume that the filter is a normal density. This assumption is problematic for our particular case which is symmetric across the origin, i.e. observations provide equal probabilities for the molecule to be in negative or positive side of the center of the PSF, i.e. $\pm x_k^n$. Due to this symmetry across the yz-plane, the filtering distribution consists of two modes centered symmetrically across the origin [27]. Therefore, we compute an approximate filter distribution of the form

$$p\left(x_k^n|D,\ldots,w_{1:k}\right) \approx \mathbf{SymNormal}\left(x_k^n; m_k^n, c_k^n\right) \tag{S54}$$

where $\mathbf{SymNormal}\left(m_k^n, c_k^n\right)$ denotes the symmetric normal distribution (see Table S5). The filter's parameters m_k^n and c_k^n can be computed recursively according to

$$\begin{aligned} p\left(x_k^n|D,\ldots,w_{1:k}\right) \propto\; & p\left(w_k|x_k^n, y_k^n, z_k^n, \mu_{mol}, \mu_{back}, \{b^{n'}, x^{n'}, y^{n'}, z^{n'}\}_n'\right) \\ & \times \int_{x_{k-1}^n} p\left(x_{k-1}^n|D,\ldots,w_{1:k-1}\right) p\left(x_k^n|x_{k-1}^n, D\right) dx_{k-1}^n \end{aligned} \tag{S55}$$

which, for our model, reduces to

$$p\left(x_k^n|D,\ldots,w_{1:k}\right) \propto \mathbf{Normal}\left(\mathbb{T}_{data}(w_k); \mathbb{T}_{mean}(\mu_k), 1\right) \mathbf{SymNormal}\left(x_k^n; m_{k-1}^n, c_{k-1}^n + 2D\left(t_k - t_{k-1}\right)\right) \tag{S56}$$

and, in turn, is approximated as

$$p\left(x_k^n|D,\ldots,w_{1:k}\right) \approx \mathbf{SymNormal}\left(x_k^n; m_k^n, c_k^n\right). \tag{S57}$$

To summarize, in the forward pass of the FFBS, we compute m_k^n and c_k^n of the filter of the molecule n, for all time levels $k = 1,\ldots,K$, by linearizing the approximate model around $x_1^n = \mu_{xy}$ for $k = 1$, and around $x_k^n = m_{k-1}^n$ for $k = 2,\ldots,K$. Since our observation is nonlinear, to calculate the filter, we opt between two different methods: **(i)** Extended Kalman filter (EKF) and **(ii)** Unscented Kalman filter (UKF).

In the EKF, we linearize the observations to obtain a closed form for the filter (local approximation) and in the UKF we approximate the joint probability distribution of observations and locations with a multivariate normal distribution (global approximation). The reason to use either of these filters is that the EKF is computationally cheaper but less accurate. According to our analysis it may fail to provide unbiased estimates of the background photon emission rate. On the other hand, the UKF is more robust and provides background emission rate estimates, but these benefits come at an increased computational cost.

In this study, we provide both filters and allow the user to choose between them.

Extended Kalman filter

Within the EKF approximation, the normal probability distribution preceding the symmetric normal of Eq. (S56) is linearized in order for their product to become a symmetric normal one. In this case, we linearize the mean of the observation density $\mathbf{Normal}\left(\mathbb{T}_{data}(w_k); \mathbb{T}_{mean}(\mu_k), 1\right)$, around the modes of the filter in the previous time step

$$\begin{aligned} \mathbb{T}_{mean}\left(\mu_k\left(x_k^n\right)\right) &\approx \mathbb{T}_{mean}\left(\mu_k\left(-m_{k-1}^n\right)\right) + \left.\frac{\partial \mathbb{T}_{mean}\left(\mu_k\left(x_k^n\right)\right)}{\partial x_k^n}\right|_{x_k^n = -m_{k-1}^n} \left(x_k^n + m_{k-1}^n\right) \\ \mathbb{T}_{mean}\left(\mu_k\left(x_k^n\right)\right) &\approx \mathbb{T}_{mean}\left(\mu_k\left(+m_{k-1}^n\right)\right) + \left.\frac{\partial \mathbb{T}_{mean}\left(\mu_k\left(x_k^n\right)\right)}{\partial x_k^n}\right|_{x_k^n = +m_{k-1}^n} \left(x_k^n - m_{k-1}^n\right) \end{aligned} \tag{S58}$$

where the first term linearizes around $x_k^n = -m_{k-1}^n$ and the second term linearizes around $x_k^n = +m_{k-1}^n$. Under these approximations, (S56) attains an analytical solution. In detail

$$\begin{aligned}
&\mathbf{Normal}\left(\mathbb{T}_{data}(w_k);\mathbb{T}_{mean}(\mu_k),1\right)\mathbf{SymNormal}\left(x_k^n;m_{k-1}^n,c_{k-1}^n+2D\left(t_k-t_{k-1}\right)\right)\\
&=\mathbf{Normal}\left(\mathbb{T}_{data}(w_k);\mathbb{T}_{mean}(\mu_k),1\right)\mathbf{Normal}\left(x_k^n;-m_{k-1}^n,c_{k-1}^n+2D\left(t_k-t_{k-1}\right)\right)\\
&+\mathbf{Normal}\left(\mathbb{T}_{data}(w_k);\mathbb{T}_{mean}(\mu_k),1\right)\mathbf{Normal}\left(x_k^n;+m_{k-1}^n,c_{k-1}^n+2D\left(t_k-t_{k-1}\right)\right)\\
&=\mathbf{Normal}\left(x_k^n;-m_{k-1}^n+\frac{e_k^n}{d_k^n},\frac{1}{(d_k^n)^2}\right)\mathbf{Normal}\left(x_k^n;-m_{k-1}^n,c_{k-1}^n+2D\left(t_k-t_{k-1}\right)\right)\\
&+\mathbf{Normal}\left(x_k^n;+m_{k-1}^n-\frac{e_k^n}{d_k^n},\frac{1}{(d_k^n)^2}\right)\mathbf{Normal}\left(x_k^n;+m_{k-1}^n,c_{k-1}^n+2D\left(t_k-t_{k-1}\right)\right)\\
&=\frac{1}{2}\mathbf{Normal}\left(x_k^n;-m_k^n,c_k^n\right)+\frac{1}{2}\mathbf{Normal}\left(x_k^n;+m_k^n,c_k^n\right)\\
&=\mathbf{SymNormal}\left(x_k^n;m_k^n,c_k^n\right).
\end{aligned}\tag{S59}$$

The same calculations apply also for $k=1$, where the starting density is replaced with the prior of (S26). In this case

$$\begin{aligned}
c_1^n &= \frac{\sigma_{xy}^2}{\mathbb{S}(\mu_1)^2+\sigma_{xy}^2(d_1^n)^2}\\
m_1^n &= \mu_{xy}+c_1^n d_1^n e_1^n\\
d_1^n &= \frac{\partial\mathbb{T}_{mean}(\mu_1(x_1^n))}{\partial x_1^n}\Big|_{x_1^n=\mu_{xy}}\\
e_1^n &= \mathbb{T}_{data}(w_1)-\mathbb{T}_{mean}(\mu_1(x_1^n))\Big|_{x_1^n=\mu_{xy}}
\end{aligned}\tag{S60}$$

while for $k=2,\ldots,K$ are

$$\begin{aligned}
c_k^n &= \frac{\left(c_{k-1}^n+2D\left(t_k-t_{k-1}\right)\right)}{1+\left(c_{k-1}^n+2D\left(t_k-t_{k-1}\right)\right)(d_k^n)^2}\\
m_k^n &= m_{k-1}^n+c_k^n d_k^n e_k^n\\
d_k^n &= \frac{\partial\mathbb{T}_{mean}(\mu_k(x_k^n))}{\partial x_k^n}\Big|_{x_k^n=m_{k-1}^n}\\
e_k^n &= \mathbb{T}(w_k)-\mathbb{T}(\mu_k(x_k^n))\Big|_{x_k^n=m_{k-1}^n}.
\end{aligned}\tag{S61}$$

Unscented Kalman filter

The unscented Kalman filter [46, 47] tries to fit the joint probability distribution of the observations and locations globally with a multivariate normal distribution to cope with the nonlinearity in Eq. (S56). Specifically the product of Eq. (S56) is approximated as follows

$$\begin{aligned}
&\mathbf{Normal}\left(\mathbb{T}_{data}(w_k);\mathbb{T}_{mean}(\mu_k),1\right)\mathbf{SymNormal}\left(x_k^n;m_{k-1}^n,c_{k-1}^n+2D\left(t_k-t_{k-1}\right)\right)\\
&\approx\frac{1}{2}\mathbf{BNormal}\left(\begin{bmatrix}x_k^n\\ \mathbb{T}_{data}(w_k)\end{bmatrix};\begin{bmatrix}-X_k^n\\ W_k^n\end{bmatrix},\begin{bmatrix}{}_{xx}\Sigma_k^n, & -{}_{xw}\Sigma_k^n\\ -{}_{wx}\Sigma_k^n, & {}_{ww}\Sigma_k^n\end{bmatrix}\right)\\
&+\frac{1}{2}\mathbf{BNormal}\left(\begin{bmatrix}x_k^n\\ \mathbb{T}_{data}(w_k)\end{bmatrix};\begin{bmatrix}+X_k^n\\ W_k^n\end{bmatrix},\begin{bmatrix}{}_{xx}\Sigma_k^n, & +{}_{xw}\Sigma_k^n\\ +{}_{wx}\Sigma_k^n, & {}_{ww}\Sigma_k^n\end{bmatrix}\right)\\
&\propto\frac{1}{2}\mathbf{Normal}\left(x_k^n;-m_k^n,c_k^n\right)+\frac{1}{2}\mathbf{Normal}\left(x_k^n;+m_k^n,c_k^n\right)\\
&=\mathbf{SymNormal}\left(x_k^n;m_k^n,c_k^n\right)
\end{aligned}\tag{S62}$$

Since we are faced with a filter which has two symmetric modes, we calculate the filter's mean m_k^n and variance c_k^n for one of the modes only, while we recover the other mode's mean and variance by reflection.

TABLE S2. Sigma points and corresponding weights for a standard normal according to [59]

i	1	2	3	4	5	6	7	8	9	10	11
x_i^{sn}	-5.1880	-3.9362	-2.8651	-1.8760	-0.9289	0	0.9289	1.8760	2.8651	3.9362	5.1880
g_i^*	$< 10^{-5}$	0.0002	0.0067	0.0661	0.2422	0.3694	0.2422	0.0661	0.0067	0.0002	$< 10^{-5}$

The means, auto- and cross-covariances in one mode of the (S62) are given by

$$\begin{aligned}
X_k^n &= \int_{-\infty}^{+\infty} xq(x)dx \\
W_k^n &= \int_{-\infty}^{+\infty} \mathbb{T}_{mean}\left(\mu_k\left(x\right)\right) q(x)dx \\
{}_{xx}\Sigma_k^n &= \int_{-\infty}^{+\infty} (x - X_k^n)^T (x - X_k^n) q(x)dx \\
{}_{ww}\Sigma_k^n &= \int_{-\infty}^{+\infty} (\mathbb{T}_{data}(\mu_k) - W_k^n)^T (\mathbb{T}_{data}(\mu_k) - W_k^n) q(x)dx + 1 \\
{}_{xw}\Sigma_k^n &= \int_{-\infty}^{+\infty} (x - X_k^n)^T \left(\mathbb{T}_{data}(\mu_k) - W_k^n\right) q(x)dx \\
{}_{wx}\Sigma_k^n &= \int_{-\infty}^{+\infty} \left(\mathbb{T}_{data}(\mu_k) - W_k^n\right)^T (x - X_k^n) q(x)dx
\end{aligned} \tag{S63}$$

where $q(x) = \mathbf{Normal}\left(x; m_{k-1}^n, c_{k-1}^n + 2D\left(t_k - t_{k-1}\right)\right)$ is the probability density of one mode of the filter. The same formula applies to the other mode too.

To calculate the mean value m_k^n and variance c_k^n of each normal contributing to the symmetric normal shown above, we need to specify a set of sample points, termed *sigma points* in the UKF literature [46–48, 56–58], to estimate the mean values and covariance matrix of the bivariate normal on which m_k^n and c_k^n depend. To specify sigma points, we first calculate sigma points x_i^{sn} and their weights g_i^* for a standard normal $\mathbf{Normal}(x; 0, 1)$ in Table. S2 according to [59]. We then transform x_i^{sn} that will be used in this $\mathbf{Normal}(x; m_{k-1}, c_{k-1} + 2D(t_k - t_{k-1}))$. The transformed sigma points are

$$x_i^* = m_{k-1} + x_i^{sn}\sqrt{c_{k-1} + 2D\left(t_k - t_{k-1}\right)}. \tag{S64}$$

Finally, given g_i^*, x_i^*, we calculate the mean and covariance of the bivariate normal previously introduced by

$$\begin{aligned}
X_k^n &= \int_{-\infty}^{\infty} xq(x)dx \approx \sum_i g_i^* x_i^* \\
W_k^n &= \int_{-\infty}^{\infty} \mathbb{T}_x\left(x\right) q(x)dx \approx \sum_i g_i^* \mathbb{T}_x\left(x_i^*\right) \\
{}_{xx}\Sigma_k^n &= \int_{-\infty}^{\infty} (x -_x M_k)^T (x -_x M_k) q(x)dx \approx \sum_i g_i^* \left({}_xM_k - x_i^*\right)^T \left({}_xM_k - x_i^*\right) \\
{}_{ww}\Sigma_k^n &= \int_{-\infty}^{\infty} \left(\mathbb{T}_x\left(x\right) -_w M_k\right)^T \left(\mathbb{T}_x\left(x\right) -_w M_k\right) q(x)dx \approx \sum_i g_i^* \left({}_wM_k - \mathbb{T}_x\left(x_i^*\right)\right)^T \left({}_wM_k - \mathbb{T}_x\left(x_i^*\right)\right) \\
{}_{xw}\Sigma_k^n &= \int_{-\infty}^{\infty} (x -_x M_k)^T (\mathbb{T}_x\left(x\right) -_w M_k) q(x)dx \approx \sum_i g_i^* \left({}_wM_k - \mathbb{T}_x\left(x_i^*\right)\right)^T \left({}_xM_k - x_i^*\right) \\
{}_{wx}\Sigma_k^n &= \int_{-\infty}^{\infty} \left(\mathbb{T}_x\left(x\right) -_w M_k\right)^T (x -_x M_k) q(x)dx \approx \sum_i g_i^* \left({}_xM_k - x_i^*\right)^T \left({}_wM_k - \mathbb{T}_x\left(x_i^*\right)\right).
\end{aligned} \tag{S65}$$

After computing the means X_k^n and W_k^n and auto-covariances and cross-covariances ${}_{xx}\Sigma_k$, ${}_{ww}\Sigma_k$, ${}_{xw}\Sigma_k$, ${}_{wx}\Sigma_k$, the

mean and variance of each mode of the filter are given by

$$m_k^n = X_k^n + K_k^n \left(\mathbb{T}_{data}(w_k) - W_k^n\right) \tag{S66}$$

$$c_k^n =_{xx} \Sigma_k^n - K_k^n({}_{ww}\Sigma_k^n)K_k^{nT} \tag{S67}$$

$$K_k^n = \frac{{}_{xw}\Sigma_k^n}{{}_{ww}\Sigma_k^n}. \tag{S68}$$

S4.3.2.b. Backward sampling

After we compute the filter densities $p\left(x_k^n|D,\ldots,w1:k\right)$ in the forward filtering step, through the EKF or UKF, we are able to sample the location x_k^n by using backward sampling as in Eq. (S50). Specifically, given a computed filter, we sample sequentially x_k^n according to

$$x_K^n \sim p\left(x_K^n|\{x_{k'}^n\}_{k'<K}, D, \mu_{mol}, \mu_{back}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \overline{w}\right) \tag{S69}$$

$$x_k^n \sim p\left(x_k^n|x_{k+1}^n, \{x_{k'}^n\}_{k'<k}, D, \mu_{mol}, \mu_{back}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \overline{w}\right), \qquad k=1,\ldots,K-1. \tag{S70}$$

Due to the specific choices of our problem these reduce to

$$x_K^n \sim \textbf{SymNormal}\left(m_K^n, c_K^n\right) \tag{S71}$$

$$x_k^n \sim \textbf{SymNormal}\left(m_k^n, c_k^n\right) \times \textbf{Normal}\left(x_{k+1}^n, 2D(t_{k+1}-t_k)\right), \qquad k=1,\ldots,K-1 \tag{S72}$$

where m_k^n and c_k^n are the parameters of the filter which are calculated in the forward filtering step.

S4.3.3. Sampling of inactive molecule trajectories

After updating the trajectories of the active molecules, we update the trajectories of the inactive ones. For this, we sample from the corresponding conditionals $p(\{\overline{x}^n, \overline{y}^n, \overline{z}^n\}_{n:b^n=0}|D, \mu_{mol}, \mu_{back}, \{q^n, b^n\}_n, \overline{w})$. Since the locations of inactive molecules are not associated with the observations in $\overline{w}$ and hyper-priors $\{q^n\}_n$, these conditionals simplify to $p(\{\overline{x}^n, \overline{y}^n, \overline{z}^n\}_{n:b^n=0}|D, \{b^n\}_n)$ which can be readily simulated jointly in the same manner as standard 3D Brownian motion.

S4.3.4. Sampling of the diffusion coefficient

Now that we have updated the locations of active and inactive molecules, we update the diffusion coefficient D by sampling from the corresponding conditional $p(D|\mu_{mol}, \mu_{back}, \{q^n, b^n, \overline{x}^n, \overline{y}^n, \overline{z}^n\}_n, \overline{w})$, which, due to the specific dependencies of the variables in our formulation, e.g. Eqs. (S24), (S37), (S38) and (S39), simplifies to $p(D|\{\overline{x}^n, \overline{y}^n, \overline{z}^n\}_n)$. Because of conjugacy, the latter reduces to

$$D|\{\overline{x}^n, \overline{y}^n, \overline{z}^n\}_n \sim \textbf{InvGamma}\left(\alpha', \beta'\right) \tag{S73}$$

where α' and β' are given by

$$\alpha' = \alpha_D + \frac{3N(K-1)}{2}, \quad \beta' = \beta_D + \frac{1}{4}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\frac{\left(x_{k+1}^n - x_k^n\right)^2 + \left(y_{k+1}^n - y_k^n\right)^2 + \left(z_{k+1}^n - z_k^n\right)^2}{t_{k+1}-t_k} \tag{S74}$$

S4.3.5. Sampling of the molecule prior weights and loads

We update the load prior weights q^n by sampling from the corresponding conditional $p(\{q^n\}_n|D, \mu_{mol}, \mu_{back}, \{b^n, \overline{x}^n, \overline{y}^n, \overline{z}^n\}_n, \overline{w})$, which simplifies to $p(\{q^n\}_n|\{b^n\}_n)$. For this, we use Eqs. (S33) and (S32), and because of conjugacy, the latter distribution is sampled by sampling each q^n separately according to

$$p(q^n|b^n) \propto p(b^n|q^n)p(q^n) = \textbf{Beta}\left(q^n; \frac{\alpha_q}{N} + b^n, \frac{\beta_q(N-1)}{N} + 1 - b^n\right). \tag{S75}$$

Once the weights q^n are updated, we update the loads b^n by sampling from the corresponding conditional $p(\{b^n\}_n|D, \mu_{mol}, \mu_{back}, \{q^n, \overline{x}^n, \overline{y}^n, \overline{z}^n\}_n, \overline{w})$. We perform this sampling using a Metropolis-Hastings update with proposals of the form

$$(b^n)^{prop} \sim \textbf{Bernoulli}(q^n). \tag{S76}$$

In this case, by choosing the proposal distribution similar to the prior distribution, the acceptance ratio becomes

$$\begin{aligned} r_b = \prod_{k=1}^{K} \Bigg[& \left(\frac{\mu_{back} + \mu_{mol} \sum_{n=1}^{N} (b^n)^{prop} \textbf{PSF}\left(x_k^n, y_k^n, z_k^n\right)}{\mu_{back} + \mu_{mol} \sum_{n=1}^{N} (b^n)^{old} \textbf{PSF}\left(x_k^n, y_k^n, z_k^n\right)} \right)^{w_k} \\ & \times \exp\left(-(t_k - t_{k-1}) \mu_{mol} \sum_{n=1}^{N} \left((b^n)^{old} - (b^n)^{prop} \right) \textbf{PSF}\left(x_k^n, y_k^n, z_k^n\right) \right) \Bigg] \end{aligned} \tag{S77}$$

where $(b^n)^{old}$ denotes the existing sample.

S4.3.6. Joint sampling of the molecular brightness and background photon emission rates

Finally, after we updated the locations of molecules, and loads, we update the molecular brightness and background photon emission rates μ_{mol} and μ_{back} by sampling from the corresponding conditional $p(\mu_{mol}, \mu_{back}|D, \{q^n, b^n, \overline{x}^n, \overline{y}^n, \overline{z}^n\}_n, \overline{w})$, which simplifies to $p(\mu_{mol}, \mu_{back}|\{b^n, \overline{x}^n, \overline{y}^n, \overline{z}^n\}_n, \overline{w})$. We carry over this sampling using a Metropolis-Hastings update where proposals for μ_{mol} and μ_{back} are computed according to

$$\begin{aligned} \mu_{mol}^{prop} &\sim \textbf{Gamma}\left(\alpha_{mol}^{prop}, \frac{\mu_{mol}^{old}}{\alpha_{mol}^{prop}} \right) \\ \mu_{back}^{prop} &\sim \textbf{Gamma}\left(\alpha_{back}^{prop}, \frac{\mu_{back}^{old}}{\alpha_{back}^{prop}} \right) \end{aligned} \tag{S78}$$

where μ_{mol}^{old} and μ_{back}^{old} denote the existing samples. The acceptance ratio is

$$\begin{aligned} r_\mu = \prod_{k=1}^{K} \Bigg[& \left(\frac{\mu_{back}^{prop} + \mu_{mol}^{prop} \sum_{n=1}^{N} b^n \textbf{PSF}\left(x_k^n, y_k^n, z_k^n\right)}{\mu_{back}^{old} + \mu_{mol}^{old} \sum_{n=1}^{N} b^n \textbf{PSF}\left(x_k^n, y_k^n, z_k^n\right)} \right)^{w_k} \\ & \times \exp\left((t_k - t_{k-1}) \left(\left(\mu_{back}^{old} - \mu_{back}^{prop} \right) + \left(\mu_{mol}^{old} - \mu_{mol}^{prop} \right) \sum_{n=1}^{N} b^n \textbf{PSF}\left(x_k^n, y_k^n, z_k^n\right) \right) \right) \Bigg] \\ & \times \left(\frac{\mu_{mol}^{old}}{\mu_{mol}^{prop}} \right)^{2\alpha_{mol}^{prop} - \alpha_{mol}} \exp\left(\frac{\mu_{mol}^{old} - \mu_{mol}^{prop}}{\beta_{mol}} + \alpha_{mol}^{prop} \left(\frac{\mu_{mol}^{prop}}{\mu_{mol}^{old}} - \frac{\mu_{mol}^{old}}{\mu_{mol}^{prop}} \right) \right) \\ & \times \left(\frac{\mu_{back}^{old}}{\mu_{back}^{prop}} \right)^{2\alpha_{back}^{prop} - \alpha_{back}} \exp\left(\frac{\mu_{back}^{old} - \mu_{back}^{prop}}{\beta_{back}} + \alpha_{back}^{prop} \left(\frac{\mu_{back}^{prop}}{\mu_{back}^{old}} - \frac{\mu_{back}^{old}}{\mu_{back}^{prop}} \right) \right). \end{aligned} \tag{S79}$$

We should emphasize, due to the weakness of the extended Kalman filter as compared to the unscented Kalman filter, we consider the background photon emission rate is fixed for EKF. So, in this case we update the molecular brightness μ_{mol} only by sampling from the corresponding conditional $p(\mu_{mol}|D, \{q^n, b^n, \overline{x}^n, \overline{y}^n, \overline{z}^n\}_n, \overline{w})$, which simplifies to $p(\mu_{mol}|\{b^n, \overline{x}^n, \overline{y}^n, \overline{z}^n\}_n, \overline{w})$. Again, we carry over this sampling using a Metropolis-Hastings update where proposals for μ_{mol} are computed according to

$$\mu_{mol}^{prop} \sim \textbf{Gamma}\left(\alpha_{mol}^{prop}, \frac{\mu_{mol}^{old}}{\alpha_{mol}^{prop}} \right) \tag{S80}$$

and the acceptance ration will reduces to

$$
\begin{aligned}
r_\mu = \prod_{k=1}^{K} \Bigg[& \left(\frac{\mu_{back} + \mu_{mol}^{prop} \sum_{n=1}^{N} b^n \mathbf{PSF}\left(x_k^n, y_k^n, z_k^n\right)}{\mu_{back} + \mu_{mol}^{old} \sum_{n=1}^{N} b^n \mathbf{PSF}\left(x_k^n, y_k^n, z_k^n\right)} \right)^{w_k} \\
& \times \exp\left(\left(t_k - t_{k-1}\right) \left(\left(\mu_{mol}^{old} - \mu_{mol}^{prop}\right) \sum_{n=1}^{N} b^n \mathbf{PSF}\left(x_k^n, y_k^n, z_k^n\right) \right) \right) \Bigg] \\
& \times \left(\frac{\mu_{mol}^{old}}{\mu_{mol}^{prop}} \right)^{2\alpha_{mol}^{prop} - \alpha_{mol}} \exp\left(\frac{\mu_{mol}^{old} - \mu_{mol}^{prop}}{\beta_{mol}} + \alpha_{mol}^{prop} \left(\frac{\mu_{mol}^{prop}}{\mu_{mol}^{old}} - \frac{\mu_{mol}^{old}}{\mu_{mol}^{prop}} \right) \right).
\end{aligned}
\tag{S81}
$$

S5. Extension for multiple diffusive species

In the case of more than one diffusive species, we can readily modify the model to capture multiple diffusion coefficients. To show that our method can be extended, we consider two diffusive species. Namely, the extended formulation is

$$ {}_1D \sim \textbf{InvGamma}\left(\alpha_D, \beta_D\right) \tag{S82} $$
$$ {}_2D \sim \textbf{InvGamma}\left(\alpha_D, \beta_D\right) \tag{S83} $$

$$ \mu_{mol} \sim \textbf{Gamma}\left(\alpha_{mol}, \beta_{mol}\right) \tag{S84} $$
$$ \mu_{back} \sim \textbf{Gamma}\left(\alpha_{back}, \beta_{back}\right) \tag{S85} $$

$$ {}_1q^n \sim \textbf{Beta}\left(\frac{\alpha_q}{{}_1N}, \beta_q \frac{{}_1N - 1}{{}_1N}\right) \tag{S86} $$
$$ {}_1b^n | {}_1q^n \sim \textbf{Bernoulli}({}_1q^n) \tag{S87} $$

$$ {}_2q^n \sim \textbf{Beta}\left(\frac{\alpha_q}{{}_2N}, \beta_q \frac{{}_2N - 1}{{}_2N}\right) \tag{S88} $$
$$ {}_2b^n | {}_2q^n \sim \textbf{Bernoulli}({}_2q^n) \tag{S89} $$

$$ {}_1x_1^n \sim \textbf{SymNormal}\left(\mu_{xy}, \sigma_{xy}^2\right) \tag{S90} $$
$$ {}_1y_1^n \sim \textbf{SymNormal}\left(\mu_{xy}, \sigma_{xy}^2\right) \tag{S91} $$
$$ {}_1z_1^n \sim \textbf{SymNormal}\left(\mu_z, \sigma_z^2\right) \tag{S92} $$

$$ {}_1x_k^n | {}_1x_{k-1}^n, {}_1D \sim \textbf{Normal}\left({}_1x_{k-1}^n, 2(t_k - t_{k-1}) {}_1D\right), \qquad k = 2, \ldots, K \tag{S93} $$
$$ {}_1y_k^n | {}_1y_{k-1}^n, {}_1D \sim \textbf{Normal}\left({}_1y_{k-1}^n, 2(t_k - t_{k-1}) {}_1D\right), \qquad k = 2, \ldots, K \tag{S94} $$
$$ {}_1z_k^n | {}_1z_{k-1}^n, {}_1D \sim \textbf{Normal}\left({}_1z_{k-1}^n, 2(t_k - t_{k-1}) {}_1D\right), \qquad k = 2, \ldots, K \tag{S95} $$

$$ {}_2x_1^n \sim \textbf{SymNormal}\left(\mu_{xy}, \sigma_{xy}^2\right) \tag{S96} $$
$$ {}_2y_1^n \sim \textbf{SymNormal}\left(\mu_{xy}, \sigma_{xy}^2\right) \tag{S97} $$
$$ {}_2z_1^n \sim \textbf{SymNormal}\left(\mu_z, \sigma_z^2\right) \tag{S98} $$

$$ {}_2x_k^n | {}_2x_{k-1}^n, {}_2D \sim \textbf{Normal}\left({}_2x_{k-1}^n, 2(t_k - t_{k-1}) {}_2D\right), \qquad k = 2, \ldots, K \tag{S99} $$
$$ {}_2y_k^n | {}_2y_{k-1}^n, {}_2D \sim \textbf{Normal}\left({}_2y_{k-1}^n, 2(t_k - t_{k-1}) {}_2D\right), \qquad k = 2, \ldots, K \tag{S100} $$
$$ {}_2z_k^n | {}_2z_{k-1}^n, {}_2D \sim \textbf{Normal}\left({}_2z_{k-1}^n, 2(t_k - t_{k-1}) {}_2D\right), \qquad k = 2, \ldots, K \tag{S101} $$

$$ w_k | \{{}_1x_k^n, {}_1y_k^n, {}_1z_k^n, {}_2x_k^n, {}_2y_k^n, {}_2z_k^n, {}_1b^n, {}_2b^n\}_n, \mu_{mol}, \mu_{back} \sim \textbf{Poisson}\left(\mu_k\right), \qquad k = 1, \ldots, K \tag{S102} $$

$$ \mu_k = (t_k - t_{k-1}) \left(\mu_{back} + \mu_{mol} \sum_n {}_1b^n \mathrm{PSF}({}_1x_k^n, {}_1y_k^n, {}_1z_k^n) + \mu_{mol} \sum_n {}_2b^n \mathrm{PSF}({}_2x_k^n, {}_2y_k^n, {}_2z_k^n) \right) \tag{S103} $$

where pre-scripts 1 and 2 are used to distinguish the two species. A graphical summary is show on Fig. S16. We use this formulation for the estimates shown on Figs. S11 and S12.

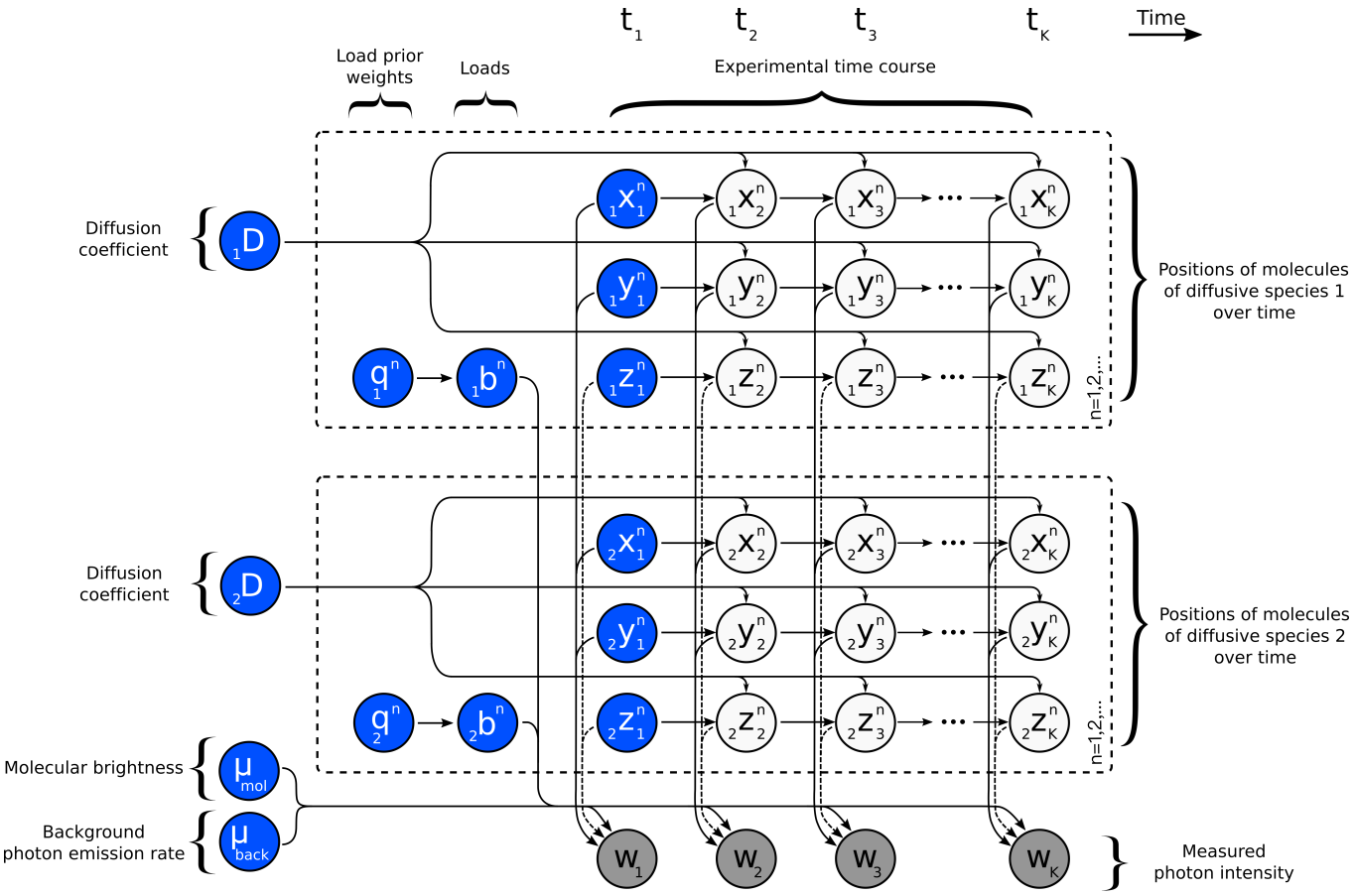

FIG. S16. **Graphical summary of the framework capturing two independent diffusion coefficients.** A multi-species population of model molecules, labeled by $n = 1, 2, \ldots$, evolves during the measurement period which is marked by $k = 1, 2, \ldots, K$. Here, ${}_1x_k^n$, ${}_1y_k^n$ and ${}_1z_k^n$ denote the location of molecule n at time t_k of species 1, ${}_2x_k^n$, ${}_2y_k^n$ and ${}_2z_k^n$ denote the location of molecule n at time t_k of species 2, μ_{mol} and μ_{back} denote molecular brightness and background photon emission rates. The diffusion coefficient ${}_1D$ and ${}_2D$ determine the evolution of the molecule locations of species 1 and 2 which, in turn, determine the instantaneous photon emission rates and ultimately the recorded photon intensities w_k. Load variables ${}_1b^n$ and ${}_2b^n$, with prior weights ${}_1q^n$ and ${}_2q^n$, respectively, are introduced to model molecule populations of the two species of *a priori* unknown sizes. Following common machine learning convention, the measurements w_k are dark shaded. Additionally, model variables requiring prior probability distributions are highlighted in blue.

S6. Summary of notation, abbreviations, parameters and other options

TABLE S3. Summary of notation.

Description	Variable	Units
Diffusion coefficient	D	$\mu m^2/s$
α parameter of the diffusion coefficient prior	α_D	-
β parameter of the diffusion coefficient prior	β_D	$\mu m^2/s$
Total time trace duration	T_{total}	s
Molecular brightness at the center of the confocal volume	μ_{mol}	photons/s
α parameter of the molecular brightness's prior	α_{mol}	-
β parameter of the molecular brightness's prior	β_{mol}	photons/s
Proposal parameter of the molecule photon emission rate	α_{mol}^{prop}	-
Emission rate of molecule n at time t_k	μ_k^n	photons/s
Combined photon emission rates of all molecules at time t_k	μ_k	photons/s
Background photon emission rate	μ_{back}	photons/s
α parameter of the background photon emission rate's prior	α_{back}	-
β parameter of the background photon emission rate's prior	β_{back}	photons/s
Proposal parameter of the background photon emission rate	α_{back}^{prop}	-
Minor semi-axis of confocal PSF (focal plane)	ω_{xy}	μm
Major semi-axis of confocal 3DG PSF (optical axis)	ω_z	μm
Major semi-axis of confocal 2DG-L PSF (optical axis)	z_R	μm
Laser wavelength	λ_{exc}	μm
Numerical aperture	NA	-
Solution refractive index	n_{sol}	-
Location of sigma points	x^{sn}	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 intensity at time t_k	w_k	photons
Load 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}	μm
Variance of the initial molecule position's prior on the z-axis	σ_z	μm
Normalized distance of molecule n at time t_k	d_k^n	-
Normalized distance for the definition of effective volume	ℓ	-
Effective volume	V_ℓ	μm^3
Heaviside function	H	-
Periodic boundary in the xy-plane (focal plane)	L_{xy}	μm
Periodic boundary on the z-axis (optical axis)	L_z	μm
Convergence threshold	ϵ_{thr}	$\mu m^2/s$
Bin threshold	ϵ_{bin}	photons

TABLE S4. List of abbreviations.

Phrase	Abbreviation
Fluorescence correlation spectroscopy	FCS
Point spread function	PSF
Three dimensional Gaussian	3DG
Two dimensional Gaussian-Lorentzian	2DGL
Two dimensional Gaussian-Cylindrical	2DGC
Forward filtering backward sampling	FFBS
Markov chain Monte Carlo	MCMC
Graphical user interface	GUI
Excitation profile	EXC
Collection efficiency function	CEF

TABLE S5. Probability distributions used and their densities. Here, the corresponding random variables are denoted by x.

Distribution	Notation	Probability density function	Mean value	Variance/Covariance
Normal	$\mathbf{Normal}(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Symmetric Normal	$\mathbf{SymNormal}(\mu, \sigma^2)$	$\frac{1}{2}\frac{e^{-\frac{(x+\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} + \frac{1}{2}\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$	0	$\mu^2 + \sigma^2$
Bivariate Normal	$\mathbf{BNormal}(\mu, \Sigma)$	$\frac{1}{2\pi\sqrt{\lvert\Sigma\rvert}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$	μ	Σ
Poisson	$\mathbf{Poisson}(\lambda)$	$\frac{\lambda^x e^{-\lambda}}{x!}$	λ	λ
Gamma	$\mathbf{Gamma}(\alpha, \beta)$	$\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}$	$\alpha\beta$	$\alpha\beta^2$
Inverse Gamma	$\mathbf{InvGamma}(\alpha, \beta)$	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\frac{\beta}{x}}$	$\frac{\beta}{\alpha-1}$	$\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$
Beta	$\mathbf{Beta}(\alpha, \beta)$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Bernoulli	$\mathbf{Bernoulli}(q)$	$(q-1)\delta_0(x) + q\delta_1(x)$	q	$q(1-q)$

TABLE S6. Parameter values used in the generation of the synthetic traces. Choices are listed according to figures.

	PSF	L_{xy}	L_z	ω_{xy}	ω_z,z_R	N	D	μ_{mol}	μ_{back}	T_{total}	δt
Units		μm	μm	μm	μm	-	$\mu m^2/s$	photons/s	photons/s	s	s
Fig. 1(a1)	3DG	2	3	0.30	1.50	10	10	5×10^4	10^3	0.1	10^{-4}
Fig. 1(a2)	3DG	2	3	0.30	1.50	100	10	5×10^4	10^3	0.1	10^{-4}
Fig. 2(a)	3DG	2,2,2,2,4	3,3,3,3,7	0.30	1.50	$10^2, 10^2, 10^2, 10^2, 10^3$	$10^{-2}, 10^{-1}, 1, 10, 10^2$	5×10^4	10^3	0.1	10^{-4}
Fig. 2(c)	3DG	2	3	0.30	1.50	150	10	5×10^4	10^3	10	10^{-4}
Fig. 3(a)	3DG	2	3	0.30	1.50	150	1	$10^5, 5{\times}10^4, 10^4$	10^3	0.1	10^{-4}
Fig. S3	3DG	2	3	0.30	1.50	50	1	5×10^4	10^3	0.1	10^{-4}
Fig. S4(a)	3DG	2	3	0.30	1.50	50	10	5×10^4	10^3	100	10^{-4}
Fig. S5(a1)	3DG	2	3	0.30	1.50	50	10	5×10^4	10^3	0.1	10^{-4}
Fig. S5(a2)	2DGL	2	3	0.30	1.50	50	10	5×10^4	10^3	0.1	10^{-4}
Fig. S5(a3)	2DGC	2	3	0.30	-	50	10	5×10^4	10^3	0.1	10^{-4}
Fig. S11(a)	3DG	2	3	0.30	1.50	20, 20	1,10	5×10^4	10^3	1	10^{-4}

TABLE S7. Parameter values used in the analyses of the traces. Choices are listed according to figures.

	PSF	ω_{xy}	ω_z,z_R	N	α_D	β_D	α_{mol}	β_{mol}	α_{mol}^{prop}	α_{back}	β_{back}	α_{back}^{prop}	α_q	β_q	μ_{xy}	μ_z	σ_{xy}^2	σ_z^2	ϵ_{thr}	ϵ_{bin}
Units		μm	μm	-	-	$\mu m^2/s$	-	pht/s	-	-	pht/s	-	-	-	μm	μm	μm^2	μm^2	$\mu m^2/s$	pht
Fig. 1(a1)	3DG	0.30	1.50	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 1(a2)	3DG	0.30	1.50	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 2(a)	3DG	0.30	1.50	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 2(b)	3DG	0.30	1.50	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 3	3DG	0.30	1.50	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 4	2DGL	0.40	-	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 5	3DG	0.27	4.51	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 6	3DG	0.27	4.51	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 7(a1)	3DG	0.27	4.51	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 7(a2)	3DG	0.27	4.51	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 7(a3)	3DG	0.27	4.51	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. S3	3DG	0.30	1.50	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. S4	3DG	0.23	0.55	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. S5(b1)	3DG	0.27	4.51	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. S5(b2)	2DGL	0.27	4.51	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. S5(b3)	2DGC	0.27	-	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. S8	3DG	0.27	4.51	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. S9	3DG	0.27	4.51	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. S10	2DGL	0.42	-	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. S11	3DG	0.3	1.5	50	1	1	2	10^4	10^3	2	500	10^3	1	1	0.2	0.2	2	2	0.1	4

[1] H. C. Berg, *Random walks in biology* (Princeton University Press, 1993).
[2] T. Hida, in *Brownian Motion* (Springer, 1980) pp. 44–113.
[3] A. Einstein, Annalen der physik **322**, 549 (1905).
[4] H. S. Muddana, S. Sengupta, A. Sen, and P. J. Butler, arXiv preprint arXiv:1410.0844 (2014).
[5] N.-S. Cheng, Industrial & engineering chemistry research **47**, 3285 (2008).
[6] L. Cohen, in *Advanced Signal Processing Algorithms, Architectures, and Implementations III*, Vol. 1770 (International Society for Optics and Photonics, 1992) pp. 378–394.
[7] J. Widengren, U. Mets, and R. Rigler, The Journal of Physical Chemistry **99**, 13368 (1995).
[8] J. Enderlein, I. Gregor, D. Patra, and J. Fitter, Journal of fluorescence **15**, 415 (2005).
[9] R. Rigler and E. S. Elson, *Fluorescence correlation spectroscopy: theory and applications*, Vol. 65 (Springer Science & Business Media, 2012).
[10] S.-M. Guo, J. He, N. Monnier, G. Sun, T. Wohland, and M. Bathe, Analytical chemistry **84**, 3880 (2012).
[11] J. Enderlein and W. P. Ambrose, Applied optics **36**, 5298 (1997).
[12] Y. Chen, J. D. Müller, Q. Ruan, and E. Gratton, Biophysical journal **82**, 133 (2002).
[13] T. Wilson *et al.*, *Confocal microscopy*, Vol. 426 (Academic press London, 1990).
[14] J. Pawley, *Handbook of biological confocal microscopy* (Springer Science & Business Media, 2010).
[15] P. Schwille, Cell biochemistry and biophysics **34**, 383 (2001).
[16] E. Wolf, in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, Vol. 253 (The Royal Society, 1959) pp. 349–357.
[17] B. Richards and E. Wolf, in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, Vol. 253 (The Royal Society, 1959) pp. 358–379.
[18] R. Rigler, Ü. Mets, J. Widengren, and P. Kask, European Biophysics Journal **22**, 169 (1993).
[19] B. Zhang, J. Zerubia, and J.-C. Olivo-Marin, Applied Optics **46**, 1819 (2007).
[20] K. M. Berland, P. So, and E. Gratton, Biophysical Journal **68**, 694 (1995).
[21] T. Dertinger, V. Pacheco, I. von der Hocht, R. Hartmann, I. Gregor, and J. Enderlein, ChemPhysChem **8**, 433 (2007).
[22] A. E. Siegman, "Lasers (mill valley, ca," (1986).
[23] H. Blom and G. Björk, Applied optics **48**, 6050 (2009).

[24] M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light* (Elsevier, 2013).
[25] S. F. Gibson and F. Lanni, JOSA A **9**, 154 (1992).
[26] V. Buschmann, B. Krämer, F. Koberling, R. Macdonald, and S. Rättinger, Application Note PicoQuant GmbH, Berlin (2009).
[27] S. Jazani, I. Sgouralis, and S. Pressé, The Journal of Chemical Physics (2019 (to appear)).
[28] F. J. Anscombe, Biometrika **35**, 246 (1948).
[29] J. Paisley and M. I. Jordan, arXiv preprint arXiv:1604.00685 (2016).
[30] T. Broderick, M. I. Jordan, J. Pitman, *et al.*, Bayesian analysis **7**, 439 (2012).
[31] J. Paisley and L. Carin, in *Proceedings of the 26th Annual International Conference on Machine Learning* (ACM, 2009) pp. 777–784.
[32] L. Al Labadi and M. Zarepour, Sankhya A , 1.
[33] C. Robert and G. Casella, *Introducing Monte Carlo Methods with R* (Springer Science & Business Media, 2009).
[34] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*, Vol. 2 (CRC press Boca Raton, FL, 2014).
[35] U. Von Toussaint, Reviews of Modern Physics **83**, 943 (2011).
[36] H. Liu and H. Motoda, *Computational methods of feature selection* (CRC Press, 2007).
[37] M. Tavakoli, J. N. Taylor, C.-B. Li, T. Komatsuzaki, and S. Pressé, "Single molecule data analysis: An introduction," in *Advances in Chemical Physics* (John Wiley & Sons, 2017) Chap. 4, pp. 205–305.
[38] I. Sgouralis and S. Pressé, Biophysical Journal **112**, 2021 (2017).
[39] O. Cappé, E. Moulines, and T. Rydén, in *Proceedings of EUSFLAT Conference* (Springer, 2009) pp. 14–16.
[40] T. Rydén *et al.*, Bayesian Analysis **3**, 659 (2008).
[41] L. Rabiner and B. Juang, ieee assp magazine **3**, 4 (1986).
[42] C. M. Bishop, *Pattern recognition and machine learning* (springer, 2006).
[43] L. R. Rabiner, Proceedings of the IEEE **77**, 257 (1989).
[44] S. L. Scott, Journal of the American Statistical Association **97**, 337 (2002).
[45] M. Y. Byron, K. V. Shenoy, and M. Sahani, Technical report, Department of Electrical Engineering, Stanford University **19**, 25 (2004).
[46] E. A. Wan and R. Van Der Merwe, in *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000* (Ieee, 2000) pp. 153–158.
[47] H. M. Menegaz, J. Y. Ishihara, G. A. Borges, and A. N. Vargas, IEEE Transactions on automatic control **60**, 2583 (2015).
[48] J. Stoer and R. Bulirsch, *Introduction to numerical analysis*, Vol. 12 (Springer Science & Business Media, 2013).
[49] R. E. Kalman, Journal of Basic Engineering **82**, 35 (1960).
[50] R. E. Kalman and R. S. Bucy, Journal of Basic Engineering **83**, 95 (1961).
[51] H. W. Sorenson, in *Advances in Control Systems*, Vol. 3 (Elsevier, 1966) pp. 219–292.
[52] R. Frühwirth, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **262**, 444 (1987).
[53] L. Ljung, IEEE Transactions on Automatic Control **24**, 36 (1979).
[54] T. Song and J. Speyer, IEEE Transactions on Automatic Control **30**, 940 (1985).
[55] M. Hoshiya and E. Saito, Journal of engineering mechanics **110**, 1757 (1984).
[56] J. Štecha and V. Havlena, in *Information Fusion (FUSION), 2012 15th International Conference on* (IEEE, 2012) pp. 495–502.
[57] S. Sarkka, IEEE Transactions on automatic control **52**, 1631 (2007).
[58] S. J. Julier and J. K. Uhlmann, in *Signal Processing, sensor fusion, and target recognition VI*, Vol. 3068 (International Society for Optics and Photonics, 1997) pp. 182–194.
[59] F. Heiss and V. Winschel, journal of Econometrics **144**, 62 (2008).