An Alternative Framework for Fluorescence Correlation Spectroscopy

Sina Jazani,[1] Ioannis Sgouralis,[1] Omer M. Shafraz,[2] Sanjeevi Sivasankar,[2] and Steve Pressé*[3]
[1)] *Center for Biological Physics, Department of Physics*
Arizona State University, Tempe, AZ 85287
[2)] *Department of Physics and Astronomy, Iowa State University, Ames, IA 50011*
[3)] *Center for Biological Physics, Department of Physics and School of Molecular Sciences*
Arizona State University, Tempe, AZ 85287

Fluorescence correlation spectroscopy (FCS), is a flexible and widely used tool routinely exploited *in vivo*. While FCS provides estimates of dynamical quantities, such as diffusion coefficients, it demands high molecular concentrations, high signal to noise ratio and long time traces typically in the minute range. This is by contrast to data acquisition which, in principle, provides information on the μs timescales that has, thus far, evaded analysis. To overcome these limitations, we adapt novel tools inspired by Bayesian nonparametrics. Using this framework, which starts from the direct analysis of photon counts, we achieve the same accuracy as FCS, but with much shorter traces, even at nearly single molecule concentrations. Our new analysis extends the capability of 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.

Due to its flexibility and limited invasiveness for *in vivo* applications, fluorescence confocal microscopy[1–4] is one of the most widely used experimental techniques of modern Biophysics. In a traditional setup, fluorescent molecules are allowed to freely diffuse into a volume illuminated by a tightly focused laser beam. Molecular motion inside the illuminated volume generates fluctuations in the emitted fluorescence that is recorded and subsequently temporally autocorrelated[1–4] or, most recently, 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 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 used[16]. These traces, capturing multiple molecule traversals of the confocal volume, provide the statistics needed for the pre- and post-processing steps used in traditional FCS analysis[16] (e.g. downsampling, autocorrelation, and nonlinear fitting to theoretical curves). However, processing steps like these downgrade the quality of the available data and demand 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[17–21], that also rely on auto-correlations.

In principle, within milliseconds, for the fluorophore concentrations and confocal volumes used in most experiments[1,2,22] 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 a few ms, thereby minimizing photo-toxic effects which remain a severe limitation of fluorescence microscopy[23–25].

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 optimization is prohibitively computationally expensive. To wit, maximum likelihood approaches[15,26], 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 nonparametrics, a branch of Statistics first suggested[27] in 1973 and only broadly popularized in physical applications over the last few years[15,28–35]. 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 reso-

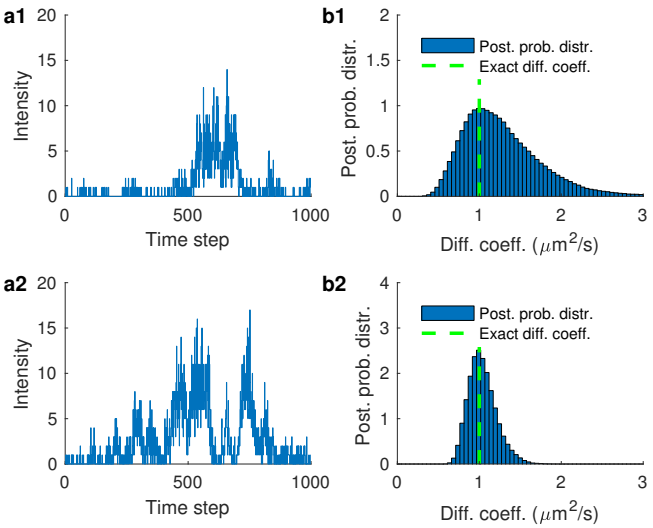

FIG. 1. **Testing the effects of molecular concentration on diffusion coefficient estimates**. **(a1)** Synthetic fluorescent intensity trace produced by a concentration of 20 nM of diffusive molecules. We used a photon emission rate, per labeled molecule, of 10^5 photons/s and a background photon emission 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 a concentration of 80 nM of diffusive molecules 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.

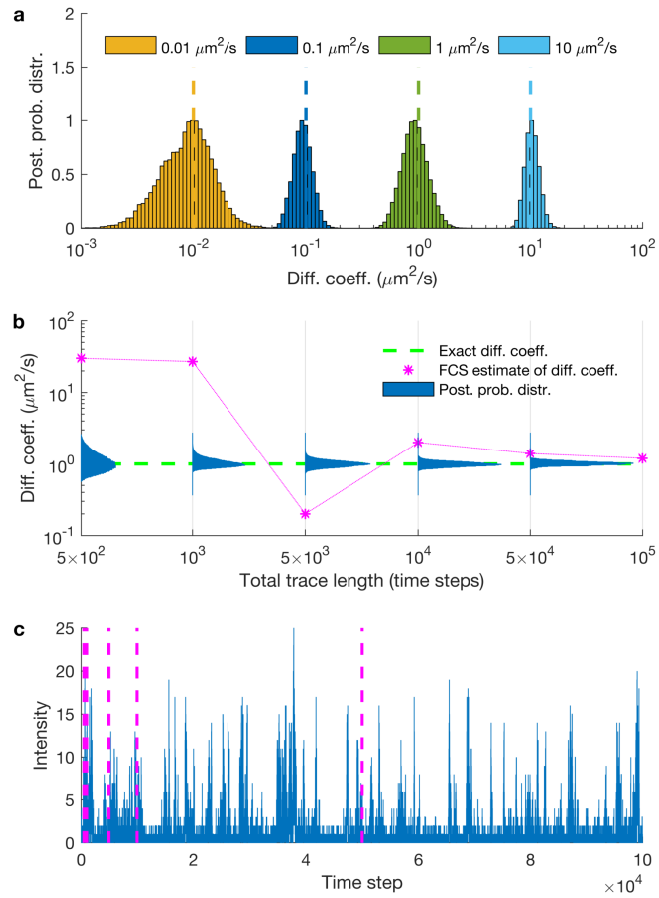

FIG. 2. **Testing the effects of diffusion coefficient and trace length on diffusion coefficient estimates**. **(a)** Posterior probability distributions deduced from traces produced from molecules with diffusion coefficients of 0.01, 0.1, 1 and 10 $\mu m^2/s$. For clarity, posteriors are normalized to maximum 1 and the horizontal axis is shown in logarithmic scale. **(b)** Posterior probability distributions deduced from traces acquired at 100 μs with total trace lengths of 5×10^2, 10^3, 5×10^3, 10^4, 5×10^4 and 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).

lution. Furthermore, our overall framework is principled, generalizable, and can estimate not only diffusion coefficients and molecular populations but also track molecules through time as well as determine their photon emission rates.

RESULTS

The method we propose follows the Bayesian paradigm[15,26,34,36]. Within this paradigm, our goal is to estimate posterior probability distributions over unknown parameters such as diffusion coefficients as well as molecular populations and locations over time.

In this section, we first validate our method by computing posterior distributions using synthetic (simulated) traces mimicking the properties of real 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) labeled molecular concentrations, Fig. 1; (ii) diffusion coefficients, Fig. 2a; (iii) trace lengths, Fig. 2b; and (iv) photon emission rates, Fig. 3. Since, the majority of our 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 trace.

To begin, in Fig. 1 we simulate a 3D Gaussian confocal volume of typical size ($\omega_{xy} = 228$ nm and $\omega_z = 550$ nm)

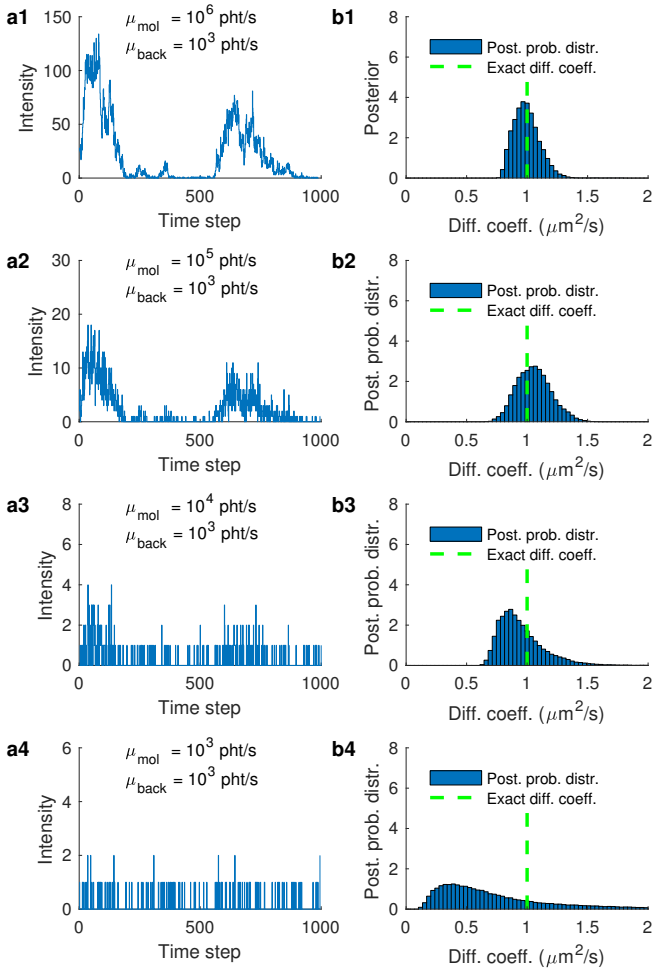

FIG. 3. **Testing the effects of photon emission rate on diffusion coefficient estimates**. **(a)** Intensity traces produced by the same molecular trajectories under molecule photon emission rates of 10^6, 10^5, 10^4 and 10^3 photons/s and background emission rate fixed at 10^3 photons/s (with the most challenging case, (a4), coinciding with identical background and single molecule emission rates). **(b)** Posterior probability distributions and exact values of diffusion coefficients obtained from the corresponding traces.

and molecules, present at concentrations of ≈20 nM (Fig. 1a1) and ≈80 nM (Fig. 1a2) diffusing at 1 μm^2/s for a total period of 0.1 s. Under these conditions, we expect at least one molecule in Fig. 1a1 and multiple molecules in Fig. 1a2 near the confocal center at any given time. It is precisely because we want to analyze photon counts, and avoid computing auto-correlations, that we must learn the number of molecules contributing to those photon counts at any given time over the length of the time trace. Determining the instantaneous number of molecules is a model selection problem that necessitates Bayesian nonparametrics[15,31,34] which allow us to model and account for in our estimates molecule populations of arbitrary size.

As can be seen in Fig. 1, a low number of photons leads to a wide estimate of the diffusion coefficient. However, the higher the number of collected photons, 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 labeled-molecule concentrations 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 a 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 low molecular concentration, Fig. 1b1, than it is for larger concentrations, 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 fluorophores' photon emission rate is, Fig. 3, the sharper the diffusion coefficient estimate becomes.

Specifically, in Fig. 3 we demonstrate the robustness of the diffusion coefficient estimates with varying photon emission rates. 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, weaker traces lead to broader posterior estimates over diffusion coefficients, as one would expect, since these time traces are associated with greater uncertainty.

B. Benchmarking on experimental data

1. Fluorescent dyes

Next, we apply our method on real time traces derived from 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[37], which is parametrized by physical quantities such as temperature and viscosity; and those estimated by FCS. To apply FCS, each time we used the full (6 min) trace available. This is by contrast to our estimates which are obtained on traces ≈1000× shorter (100 ms).

In benchmarking, we obtained and analyzed measurements under different: (i) Cy3 dye concentrations, Fig. 4; (ii) diffusion coefficients, Fig. 5; and (iii) laser powers, Fig. 6. For example, in Fig. 4, 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. Consistent with the synthetic data shown earlier, we obtain a broader posterior for diffusion coefficient when dye

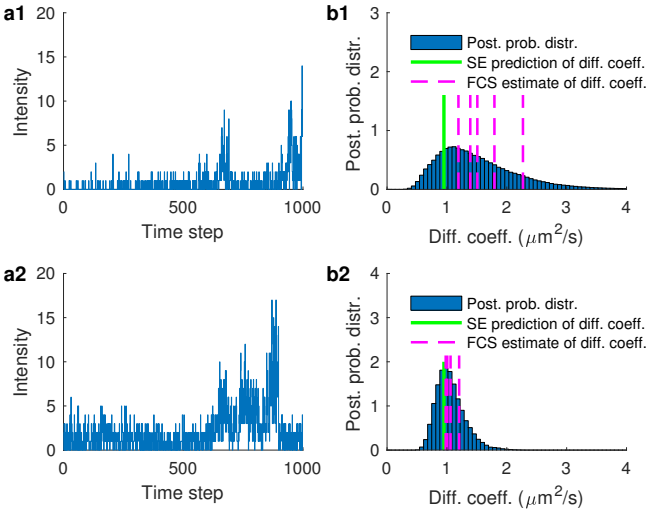

FIG. 4. **Testing the effects of molecular concentration on diffusion coefficient estimates in experimental traces of free Cy3. (a)** Experimental fluorescent intensity traces with a low (a1) and high (a2) concentration (100 pM and 1 nM, respectively) of dye molecules in glycerol/water mixture with 94% glycerol and laser power of 100 μW. **(b)** Posterior probability distribution over the diffusion coefficient estimated from the traces in (a1) and (a2). For comparison, FCS estimates shown by dashed lines are produced by five different traces, each of 6 min, i.e. ≈1000× longer than the segments used in our method. Stokes-Einstein predictions are denoted by a green line. Traces shown in (a1) and (a2) are acquired at 100 μs for a total of 100 ms.

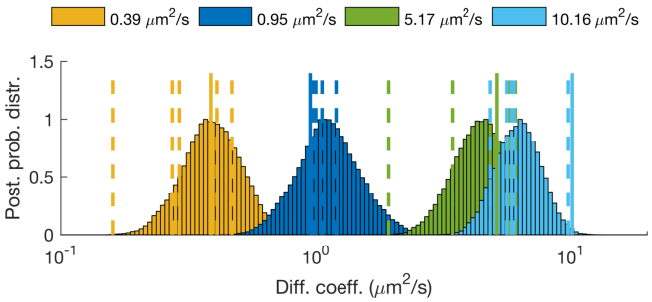

FIG. 5. **Testing the robustness of our method with different diffusion coefficients in experimental traces of free Cy3.** Posterior probability distributions from traces produced from Cy3 molecules of different diffusion coefficient. For comparison, FCS estimates shown by dashed lines are produced by five different traces, each of 6 min, i.e. ≈1000× longer than the segments used in our method. Stokes-Einstein predictions are shown with solid lines. For clarity, posteriors are normalized to maximum 1 and the horizontal axis is shown in logarithmic scale. Time traces used in the analysis are obtained at 1 nM of Cy3 and laser power of 100 μW. Different diffusion coefficients are obtained by varying the amount of glycerol from 99% to 67% in the glycerol/water mixture.

concentration is lower and correspondingly sharper posteriors for the higher concentration, Figs. 4b1 and 4b2. These are in good agreement with the Stokes-Einstein predictions and FCS estimates discussed earlier.

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. 5.

It is worth mentioning that the 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 molecular concentrations or photon emission rates.

2. Labeled proteins

To test our method beyond free dyes, we used labeled proteins, namely freely diffusing Cy3-labeled streptavidin. Similar to the previous cases, we tested a range of concentrations, diffusion coefficients, and laser powers. Figure 7 summarizes characteristic results and compares our analyses against traditional FCS and the Stokes-Einstein predictions.

As can be seen, despite of using ≈1000× more datapoints than our method, the diffusion coefficients estimated by FCS are broadly scattered around our posteriors. This inconsistency of FCS can be resolved only by supplying more datapoints as, for example, in FCS-Bayes[17–21] which analyzes many traces at once. Nevertheless, such inconsistency is absent from our posteriors.

C. 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 confocal center, 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 Supplement.

In addition to cases involving a single diffusion coefficient 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. 8, we compare the analysis of intensities created

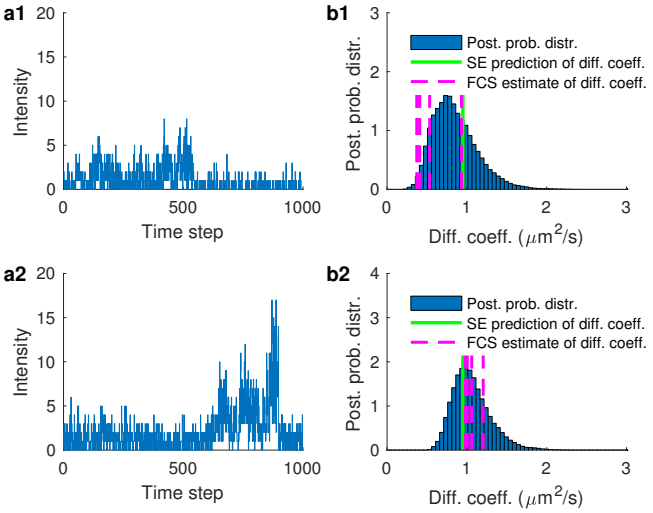

FIG. 6. **Testing the effects of photon emission rate on diffusion coefficient estimates in experimental traces of free Cy3. (a)** Experimental fluorescent intensity traces obtained under low (a1) and high (a2) laser power (25 and 100 μW, respectively) with a Cy3 concentration of 1 nM in a glycerol/water mixture of 94% glycerol. **(b)** Posterior probability distribution over the diffusion coefficient estimated from the traces in (a1) and (a2). For comparison, FCS estimates shown by dashed lines are produced by five different traces, each of 6 min, i.e. ≈1000× longer than the segments used in our method. Stokes-Einstein predictions are also highlighted in (b1) and (b2) with solid lines. Traces shown in (a1) and (a2) are acquired at 100 μs for a total of 100 ms.

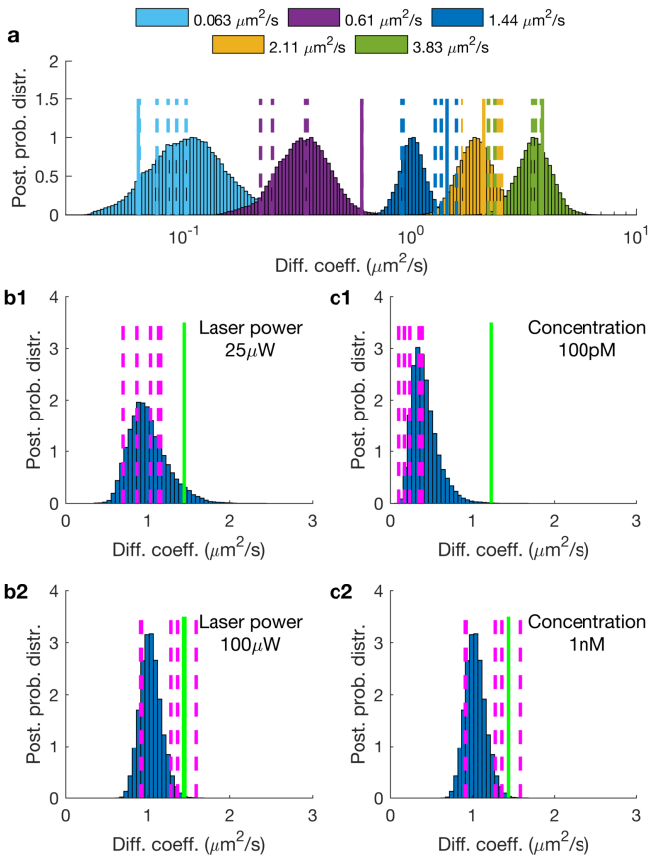

FIG. 7. **Estimating diffusion coefficients of Cy3-labeled streptavidin**. **(a)** Posterior probability distributions of diffusion coefficients of free Cy3-labeled streptavidin in different concentrations of glycerol/water mixture. The legend labels the posteriors according to Stokes-Einstein predictions. For clarity, posteriors are normalized to maximum 1 and the horizontal axis is shown in logarithmic scale. **(b)** Posterior probability distributions over the diffusion coefficients of traces generated by low and high laser powers (25 and 100 μW, respectively). **(c)** Posterior probability distributions over the diffusion coefficients of traces generated by low and high molecule concentrations (100 pM and 1 nM, respectively). For comparison, FCS estimates shown by dashed lines are produced by five different traces, each of 6 min, i.e. ≈1000× longer than the segments used in our method. Stokes-Einstein predictions are highlighted by solid lines. Posteriors are obtained from the analyses of time traces acquired at 100 μs for a total of 100 ms.

by mixing traces containing slow and fast diffusing Cy3 (99% 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

Traditional 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 nonparametrics to overcome the limitations of fluorescent correlative methods in order to accurately deduce molecular properties, such as diffusion coefficients, with significantly less data than the existing correlative approaches. Exploiting novel analysis, to obtain reliable results from short traces or traces obtained under low laser power, is key to minimizing photodamage inherent to all methods relying on illumination and especially critical to gaining insight on rapid or light-sensitive processes[23,25]. Alternatively, 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 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, confocal microscopy yields minimal spatial resolution. However, as our analysis shows, although spatial resolution may be di-

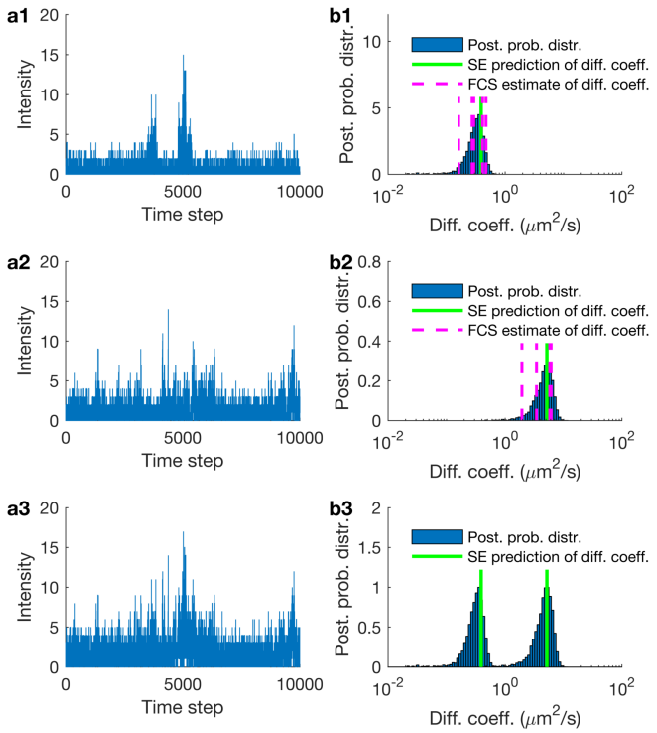

FIG. 8. **Estimating multiple different diffusion coefficients in experimental Cy3 traces. (a1)–(a2)** Experimental traces of free Cy3 in glycerol/water mixtures with 99% and 75% glycerol, respectively. **(a3)** Trace resulting by mixing the traces in (a1) and (a2). **(b1)–(b2)** Posterior probability 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. Stokes-Einstein predictions are highlighted by solid lines. Posteriors are obtained from the analyses of time traces acquired at 100 μs for a total of 1 s.

minished, reduced photo-damage 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 confocal microscopy data[2,11,38–41]. Here, we have developed a principled framework capable of taking advantage of all spatiotemporal information nested within time traces of photon counts and, together with novel Mathematics, we have reformulated the analysis of confocal microscopy data in a way that avoids the strongest limitations of FCS.

Other methods, even those that apply Bayesian techniques such as FCS-Bayes[17–21], still utilize autocorrelation functions and thus demand equally long traces to deduce diffusion coefficients as traditional FCS. By contrast, our method not only operates on exceedingly short traces but also: (i) provides interpretable estimation of errors (i.e. posterior variance) that is determined exclusively from the information content of the trace supplied (i.e. length and noise) as opposed to *ad hoc* fitting metrics (i.e. chi square); (ii) tracks instantaneous molecule photon emissions and locations; and (iii) estimates the molecule 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 demanding illumination profiles, such as those arising in two photon excitation[40,42] and TIRF microscopy[39] or even Airy patterns with[43] or without aberrations[44] by changing the specified point spread function (see Methods section). Additionally, it is possible to extend our framework to treat multiple diffusion coefficients (see Supplement) or photon emission kinetics as would be relevant for FCS-FRET[45,46] and FLIM[47–49]. Also, our method could be extended to handle more complex fluorophore photophysics[22,50–52], 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 confocal center as well as diffusion coefficients.

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

D. 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 ≈ 1 μs, we ignore triplet state and photon anti-bunching effects which occur on vastly different timescales[16,53,54].

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[15,26]

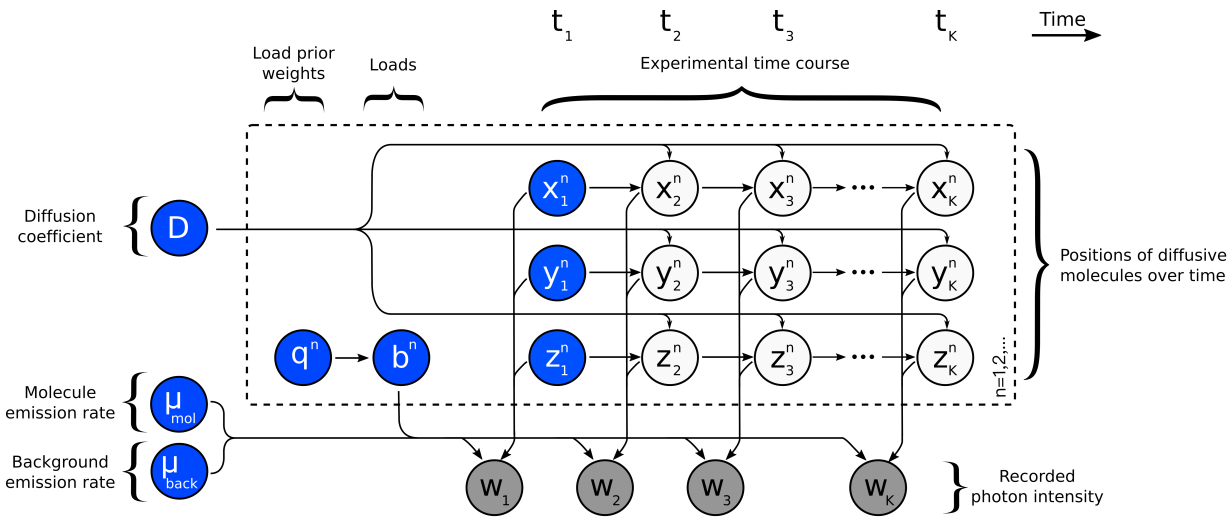

FIG. 9. **Graphical representation of the formulation used in the analysis of time fluorescence traces**. A population of model molecules, labeled by $n = 1, 2, \ldots$, evolves over the course of the experiment which is marked by $k = 1, 2, \ldots, K$. Here, x_k^n,y_k^n,z_k^n denote the location in Cartesian space of molecule n at time t_k; μ_{mol} denotes the maximum photon emission rate achieved by individual molecules; 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.

(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 nonparametrics 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[55,56].

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[57]. Similar to other studies[38,50,58], 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 maximum photon emission rate (of detected photons) of a single molecule, i.e. the emission rate achieved when the molecule is at the center of the confocal volume. Specific choices of PSF models, such as Gaussian or Gaussian-Lorentzian, are detailed in the Supplement.

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 Supplement, these probabilities result directly from the diffusion equation. Additionally, in the Supplement, 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. 9.

E. Model inference

The quantities which we want to estimate, for example the diffusion coefficient D, molecular locations through time (x_k^n, y_k^n, z_k^n), molecular and background photon emission rates μ_{mol}, 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[15,31,36].

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 SUPPLEMENT.

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 nonparametric formulation described below.

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

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

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[35,59–61] which is described in more detail in the SUPPLEMENT.

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 nonparametric 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[36,62] 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 SUPPLEMENT, 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.

F. Data acquisition

1. Synthetic data

We obtain the synthetic data presented in the Results section by standard pseudo-random computer simulations[37,66–69] that mimic the common confocal setup. We provide details and complete parameter choices in the SUPPLEMENT.

2. Experimental data

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 confocal microscope[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). 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.

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. Press, "Unraveling the thousand word picture: An introduction to super-resolution data analysis," Chemical Reviews (2017).

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

[17]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).

[18]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).

[19]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).

[20]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).

[21]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).

[22]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).

[23]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).

[24]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).

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

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

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

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

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

[30]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).

[31]M. Tavakoli, J. N. Taylor, C.-B. Li, T. Komatsuzaki, and S. Pressé, "Single molecule data analysis: an introduction," arXiv preprint arXiv:1606.00403 (2016).

[32]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).

[33]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).

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

[35]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.

[36]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).

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

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

[39]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).

[40]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).

[41]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).

[42]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).

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

[44]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).

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

[46]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).

[47]E. B. van Munster and T. W. Gadella, "Fluorescence lifetime imaging microscopy (flim)," in *Microscopy techniques* (Springer,

2005) pp. 143–175.
[48]R. Duncan, A. Bergmann, M. Cousin, D. Apps, and M. Shipston, "Multi-dimensional time-correlated single photon counting (tcspc) fluorescence lifetime imaging microscopy (flim) to detect fret in cells," Journal of Microscopy **215**, 1–12 (2004).
[49]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).
[50]D. Axelrod, D. Koppel, J. Schlessinger, E. Elson, and W. W. Webb, "Mobility measurement by analysis of fluorescence photobleaching recovery kinetics," Biophysical journal **16**, 1055–1069 (1976).
[51]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).
[52]D. Soumpasis, "Theoretical analysis of fluorescence photobleaching recovery experiments," Biophysical journal **41**, 95–97 (1983).
[53]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).
[54]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).
[55]B. R. Masters and P. So, *Handbook of biomedical nonlinear optical microscopy* (Oxford University Press, 2008).
[56]J. Pawley, *Handbook of biological confocal microscopy* (Springer Science & Business Media, 2010).
[57]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.
[58]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).
[59]T. Broderick, M. I. Jordan, J. Pitman, *et al.*, "Beta processes, stick-breaking and power laws," Bayesian analysis **7**, 439–476 (2012).
[60]L. Al Labadi and M. Zarepour, "On approximations of the beta process in latent feature models: Point processes approach," Sankhya A , 1–21.
[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]J. Haile, I. Johnston, A. J. Mallinckrodt, S. McKay, *et al.*, "Molecular dynamics simulation: elementary methods," Computers in Physics **7**, 625–625 (1993).
[68]D. J. Higham, "An algorithmic introduction to numerical simulation of stochastic differential equations," SIAM review **43**, 525–546 (2001).
[69]R. Erban and S. J. Chapman, "Stochastic modelling of reaction–diffusion processes: algorithms for bimolecular reactions," Physical biology **6**, 046001 (2009).
[70]H. Li, C.-F. Yen, and S. Sivasankar, "Fluorescence axial localization with nanometer accuracy and precision," Nano letters **12**, 3731–3735 (2012).

ACKNOWLEDGMENTS

SP acknowledges support from NSF CAREER grant MCB-1719537. SS acknowledge support from the National Institute of Health grant R01GM121885. We thank Patrick Schmidt, Hussam Ibrahim and John Lajoie for building the FPGA based data acquisition system.

AUTHOR CONTRIBUTIONS

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

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 Physics and Astronomy, Iowa State University, Ames, IA 50011

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

Here we provide supplementary information and technical details that complement the main text. These include: (i) Additional analysis results that demonstrate the estimation of molecule and background photon emission rates, joint posterior probability distributions, molecular locations/concentrations, and additional results for multiple diffusive species. These results are repeated for simulated and experimental data. (ii) Additional details of the methods used including the motion model, the Stokes-Einstein model, point spread functions, 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 3
- S1.1. Analysis of additional simulated data 3
- S1.2. Analysis of additional experimental data 9
- S1.3. Analysis of additional data for multiple diffusive species 14

S2. Summary of point estimates 17

S3. Detailed methods description 18
- S3.1. Representation of molecular diffusive motion 18
- S3.2. Description of Stokes-Einstein model 18
- S3.3. Definition of point spread function models 18
- S3.4. Description of the data simulation 19
- S3.5. Definition of normalized distance and molecular concentration 19
- S3.6. Description of the time trace preparation 20

S4. Detailed description of the inference framework 22
- S4.1. Description of prior probability distributions 22
  - S4.1.1. Prior on the diffusion coefficient 22
  - S4.1.2. Priors on molecular and background photon emission rates 22
  - S4.1.3. Priors on initial molecular locations 23
  - S4.1.4. Priors and hyperpriors for molecular loads 23
- S4.2. Summary of model equations 24
- S4.3. Description of the computational scheme 24
  - S4.3.1. Overview of the sampling updates 25
  - S4.3.2. Sampling of the locations of active molecules 26
  - S4.3.3. Sampling of the locations of inactive molecules 28
  - S4.3.4. Sampling of the diffusion coefficient 29
  - S4.3.5. Sampling of molecule prior weights and loads 29
  - S4.3.6. Joint sampling of molecular and background photon emission rates 29

S5. Extension for multiple diffusive species 31

S6. Summary of notation, abbreviations, parameters and other options 33

S1. Additional results

S1.1. Analysis of additional simulated data

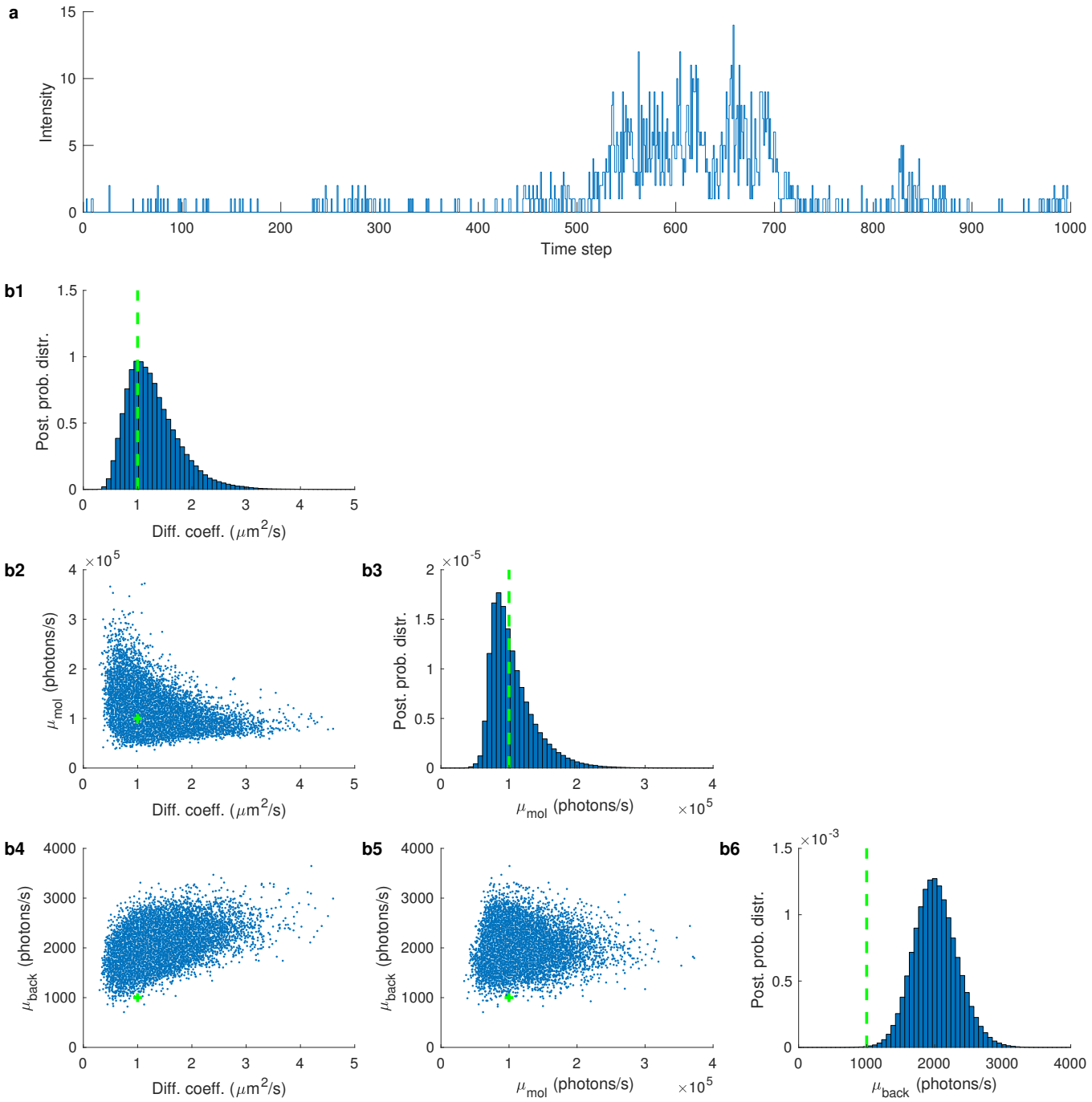

FIG. S1. **Estimated joint posterior probability distribution.** **(a)** Synthetic fluorescent intensity trace used in Fig. 1a1 with a length of 1000 data points and time step 100 μs. The true values of the diffusion coefficient, molecular and background emission rates are, 1 $\mu m^2/s$, 10^5 photons/s 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 photon emission rate. **(b3)** The posterior probability distribution of the molecular photon emission rate. **(b4)** The joint probability distribution of the diffusion coefficient and molecular photon emission rate. **(b5)** The joint probability distribution of the molecular and background photon emission rates. **(b6)** The posterior probability distribution of the background photon emission rate.

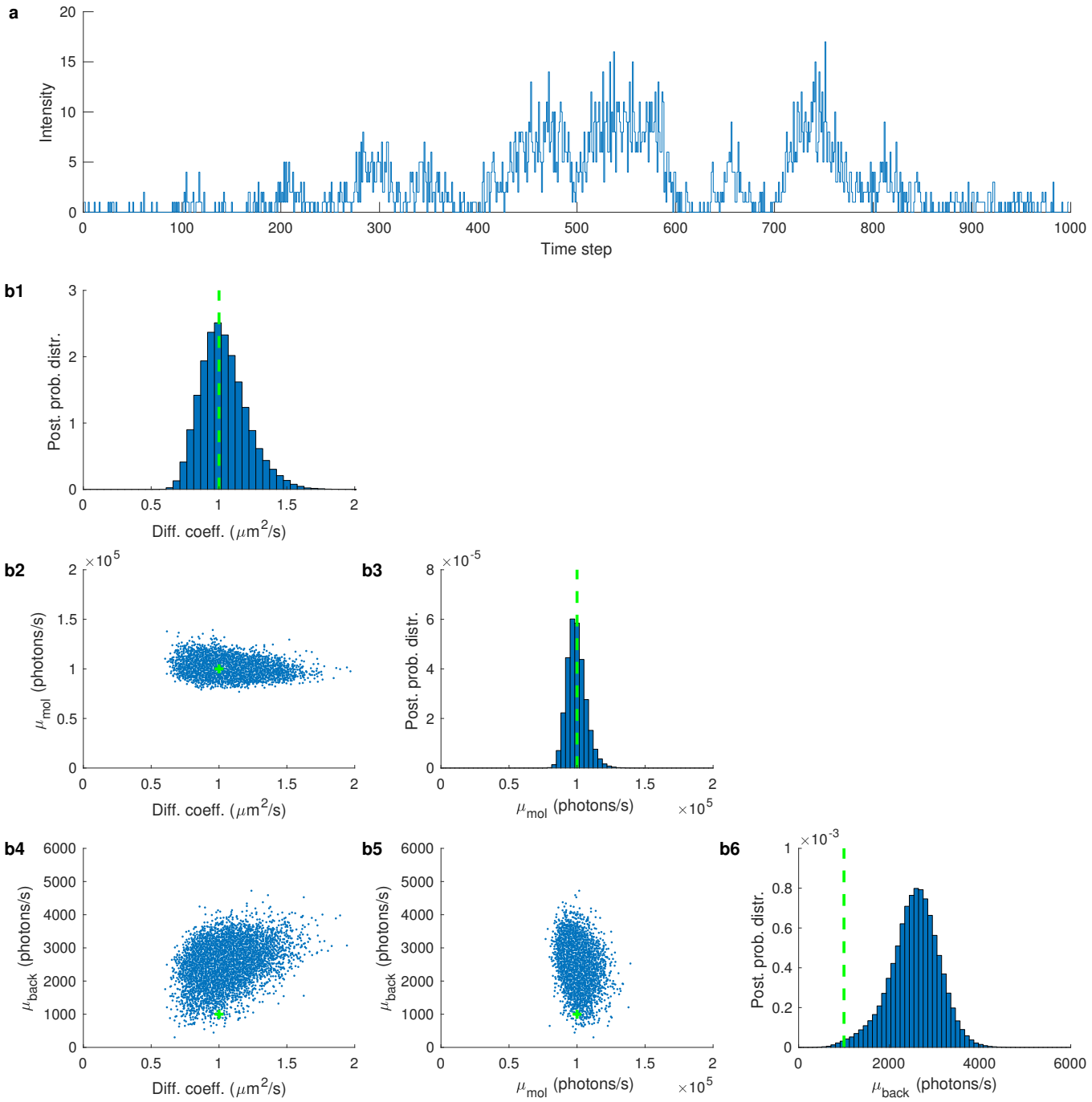

FIG. S2. **Estimated joint posterior probability distribution.** **(a)** Synthetic fluorescent intensity trace used in Fig. 1a2 with a length of 1000 data points and time step 100 μs. The true values of the diffusion coefficient, molecular and background emission rates are, 1 $\mu m^2/s$, 10^5 photons/s 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 photon emission rate. **(b3)** The posterior probability distribution of the molecular photon emission rate. **(b4)** The joint probability distribution of the diffusion coefficient and molecular photon emission rate. **(b5)** The joint probability distribution of the molecular and background photon emission rates. **(b6)** The posterior probability distribution of the background photon emission rate.

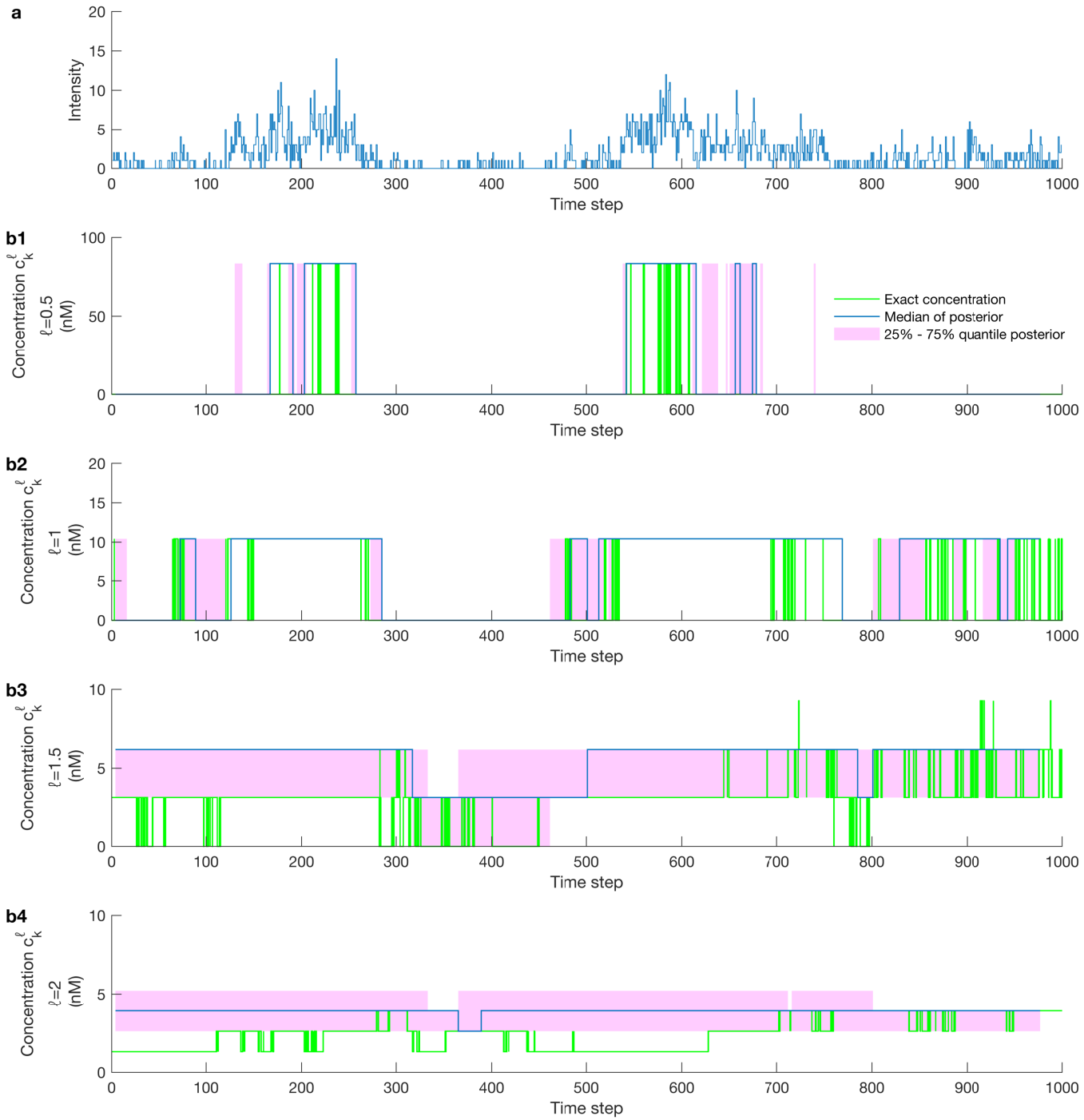

FIG. S3. **Estimated molecular locations/concentrations.** **(a)** Synthetic fluorescent intensity trace produced with a molecular photon emission rate of 10^5 photons/s and a background photon emission rate of 10^3 photons/s, diffusion coefficient of 1 μm^2/s and concentration of 5 molecules/μm^3. **(b1)–(b4)** Concentration 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 concentration is shown by the green lines, the median of the posterior over the concentration is shown by the blue lines, and the area between 25% and 75% quantiles of the posterior over the concentration is highlighted in pink. For details of the definition of concentration c_k^ℓ and the normalized distance ℓ, see Eq. (S13), below.

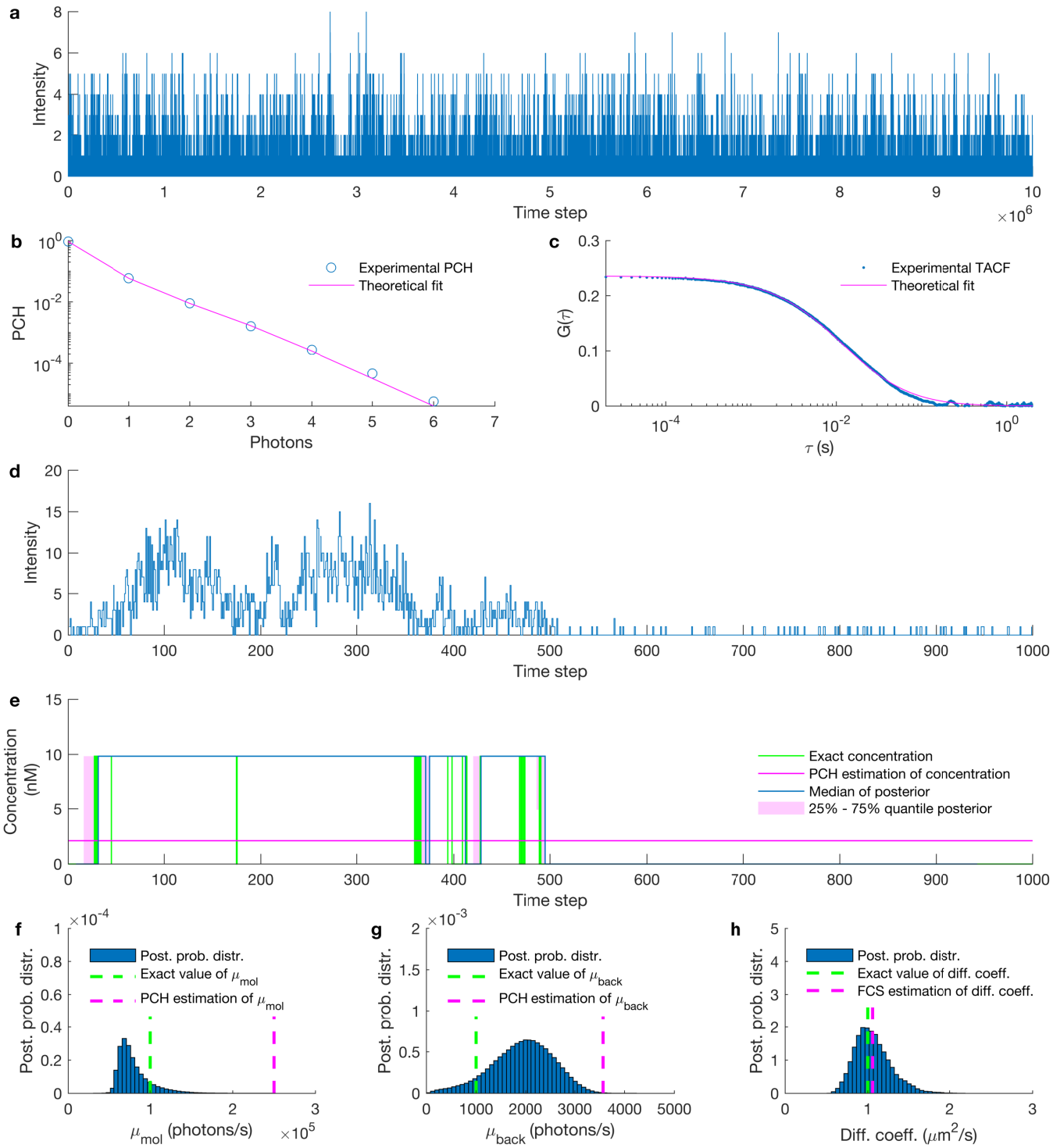

FIG. S4. **Comparison of estimated photon emission rates and concentration with FCS and PCH. (a)** Targeted synthetic fluorescent intensity trace. The time step is 10 μs and the total duration of the trace is 100 s. **(b)** PCH curve and the theoretical fit. **(c)** FCS curve and best theoretical fit. **(d)** The portion of the trace analyzed by our method rebinned at 100 μs. **(e)** The concentration of molecules in the effective volume with ℓ =1, arising from the trace in (d). Exact value of the concentration is shown by the green line and the PCH estimate is shown by the pink line. **(g)** On the posterior probability distribution of the molecular photon emission rate we superpose the PCH estimate of the molecular photon emission rate (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 and background photon emission rates of of 1 μm^2/s, 10^5 photons/s and 10^3 photons/s, respectively.

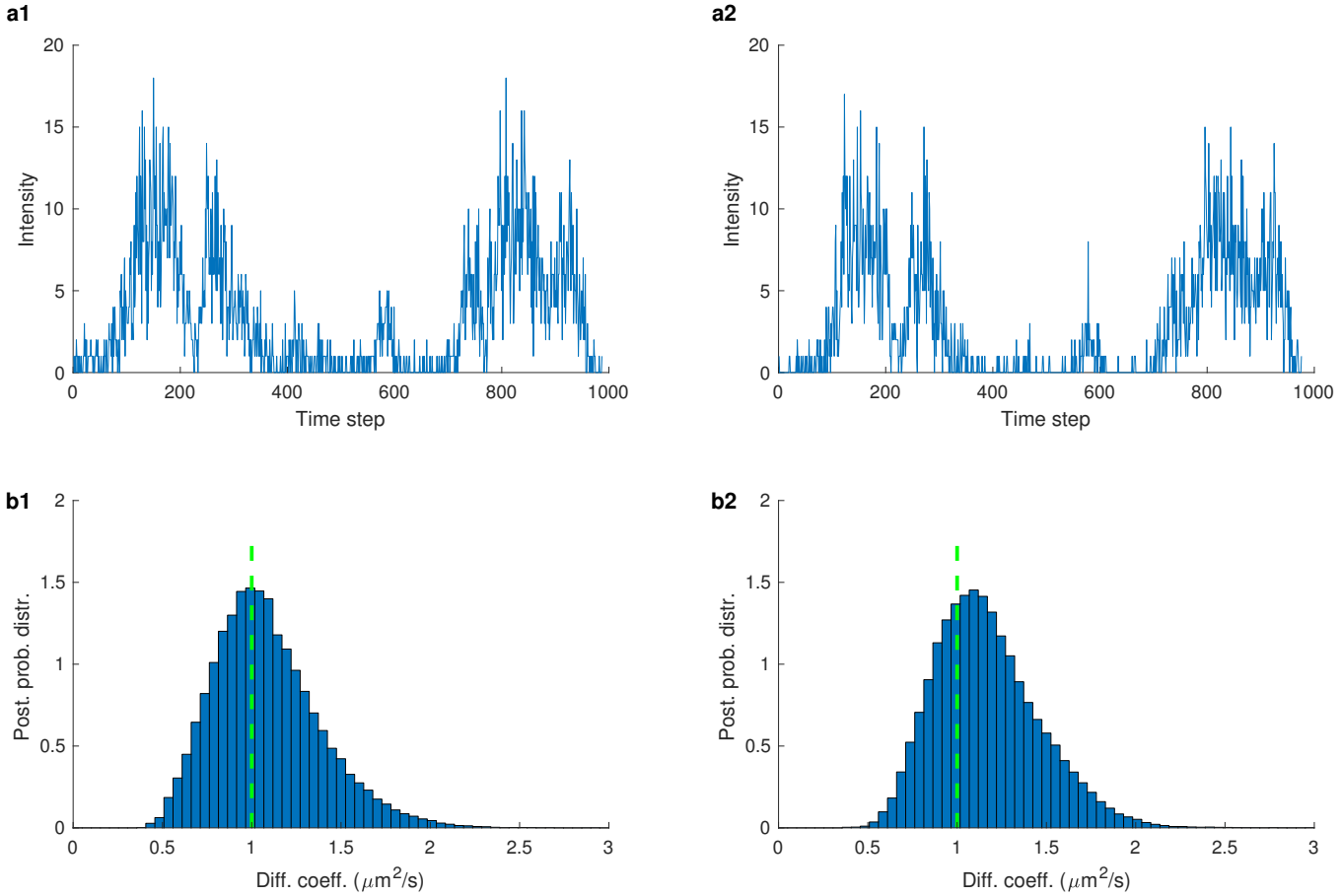

FIG. S5. **Comparison of the posterior on the diffusion coefficient obtained from synthetic fluorescent intensity traces under different PSF models**. **(a1)** Synthetic fluorescent intensity trace produced with a 3DG point spread function, Eq. (S8). **(a2)** Synthetic fluorescent intensity trace produced with a 2DG-L point spread function, Eq. (S9). **(b1)** Posterior of the diffusion coefficient using a 3DG point spread function model on the trace in (a1). **(b2)** Posterior of the diffusion coefficient using a 2DG-L point spread function model on the trace in (a2). To facilitate the comparison both traces analyzed are generated using the same underlying molecule trajectories with molecular and background photon emission rates of 10^5 photons/s and 10^3 photons/s, diffusion coefficient of 1 μm^2/s (shown by green dashed lines) and concentration of 5 molecules/μm^3.

S1.2. Analysis of additional experimental data

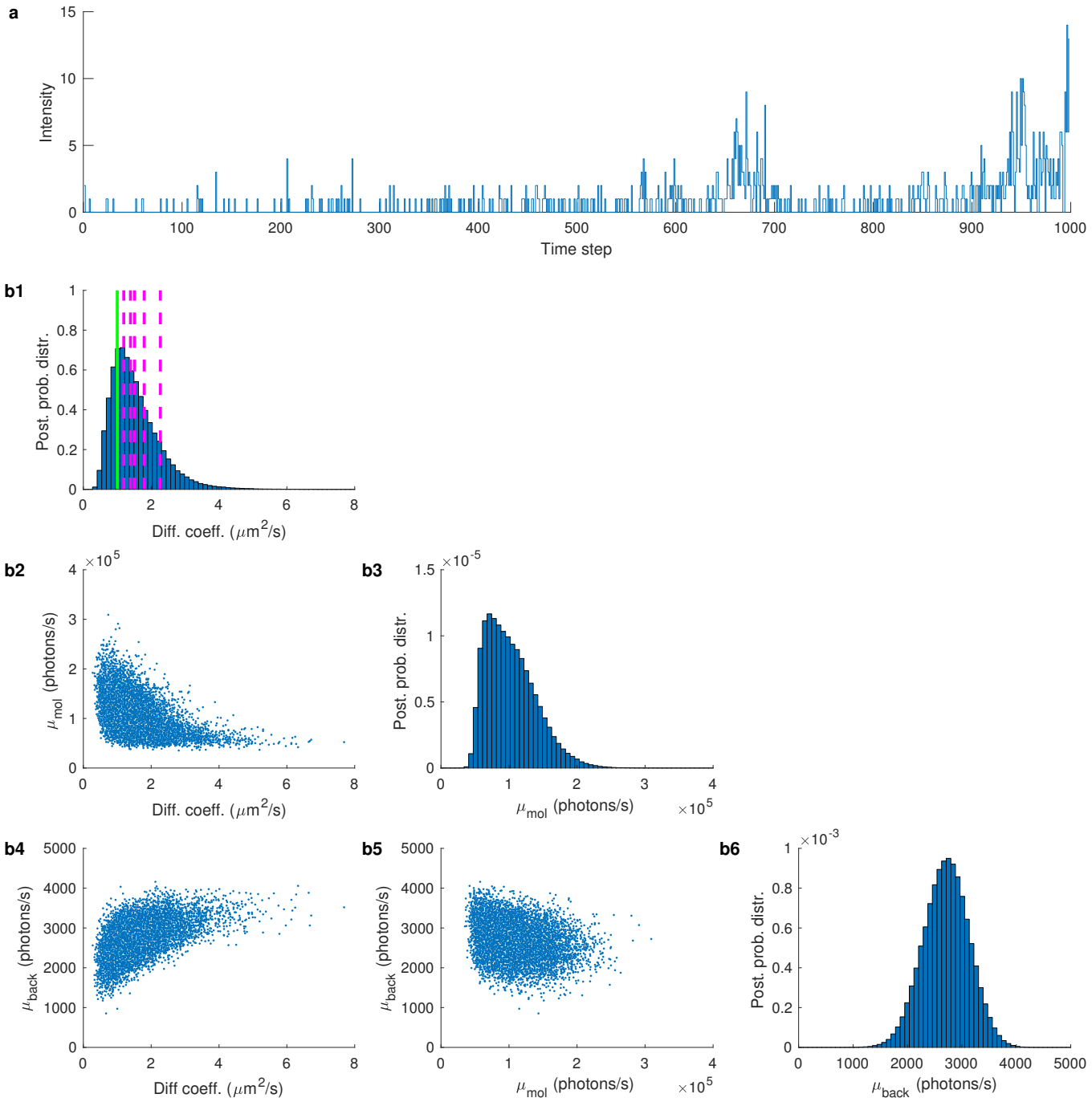

FIG. S6. **Estimated joint posterior probability distribution.** **(a)** Experimental fluorescent intensity trace used in Fig. 4a1 with a length of 1000 data points and time step 100 μs. **(b1)** The posterior of the diffusion coefficient. The Stokes-Einstein prediction of the diffusion coefficient is shown by a green solid line and the FCS estimate is shown by a magenta dashed line. **(b2)** The joint probability distribution of the diffusion coefficient and the molecular photon emission rate. **(b3)** The posterior probability distribution of the molecular photon emission rate. **(b4)** The joint probability distribution of the diffusion coefficient and molecular photon emission rate. **(b5)** The joint probability distribution of the molecular and background photon emission rates. **(b5)** The posterior probability distribution of the background photon emission rate.

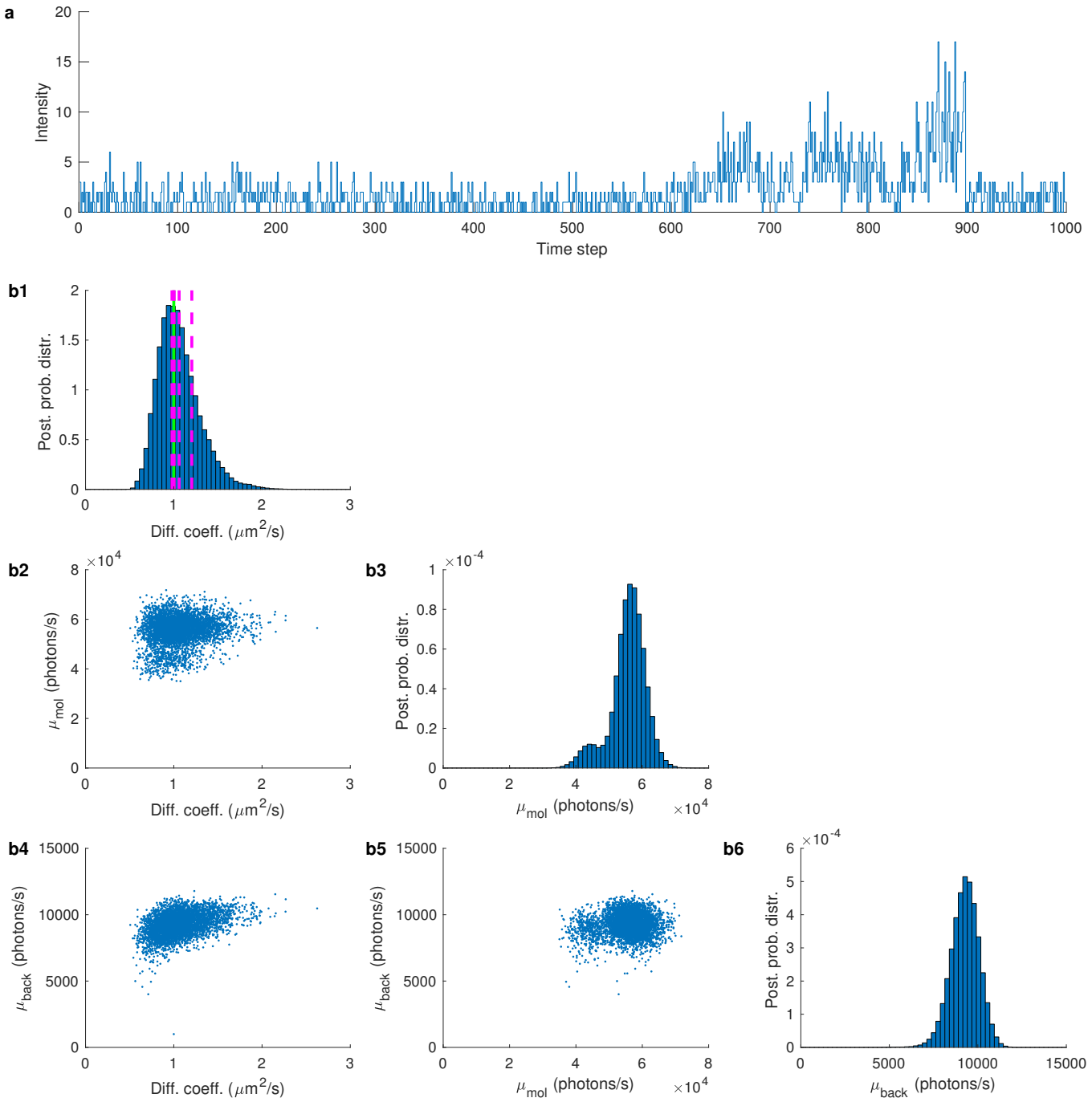

FIG. S7. **Estimated joint posterior probability distribution.** **(a)** Experimental fluorescent intensity trace used in Fig. 4a2 with a length of 1000 data points and time step 100 μs. **(b1)** The posterior of the diffusion coefficient. The Stokes-Einstein prediction of the diffusion coefficient is shown by a green solid line and the FCS estimate is shown by a magenta dashed line. **(b2)** The joint probability distribution of the diffusion coefficient and the molecular photon emission rate. **(b3)** The posterior probability distribution of the molecular photon emission rate. **(b4)** The joint probability distribution of the diffusion coefficient and the molecular photon emission rate. **(b5)** The joint probability distribution of the molecular and background photon emission rates. **(b5)** The posterior probability distribution of the background photon emission rate.

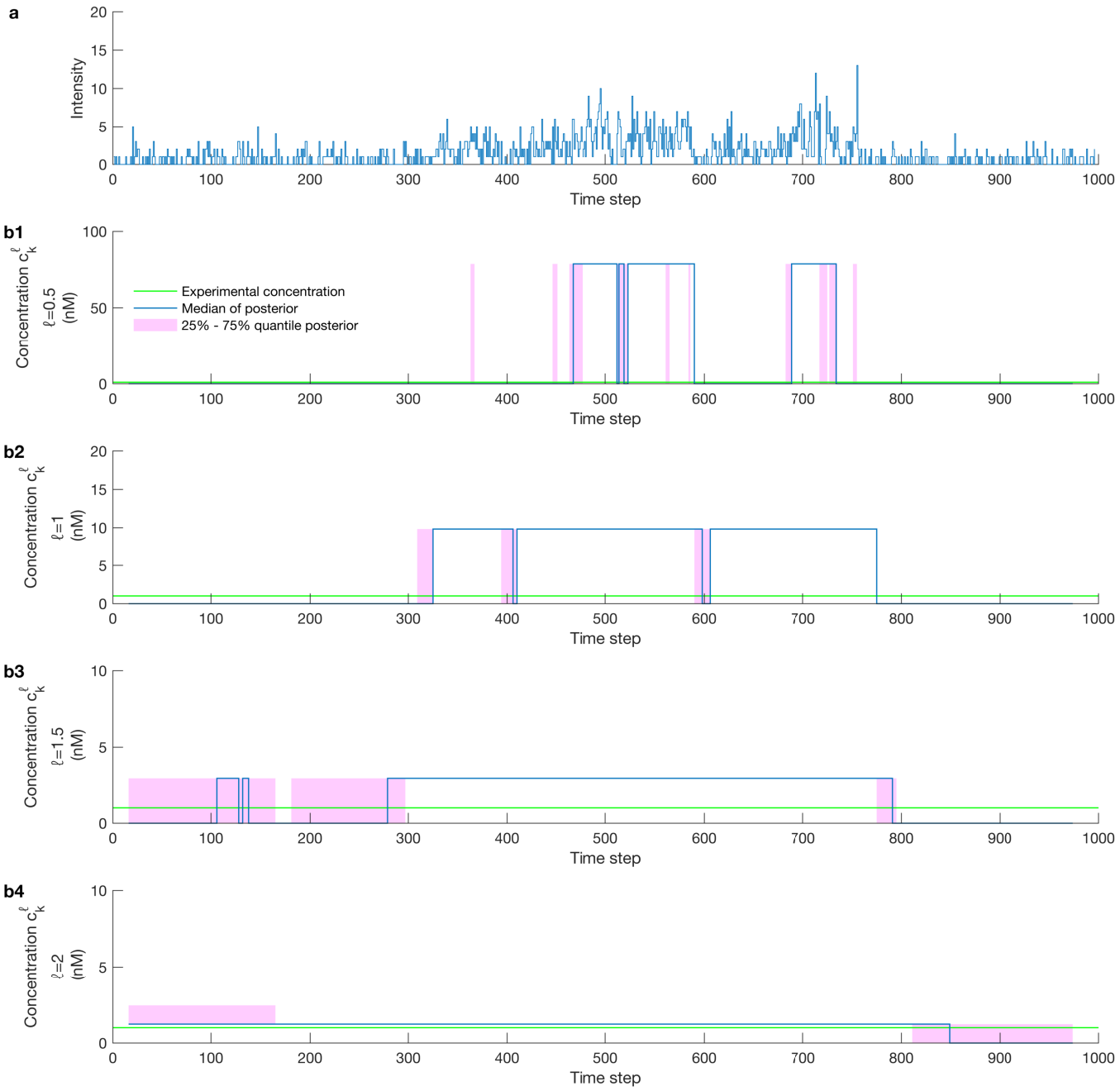

FIG. S8. **Estimated molecule locations/concentrations**. **(a)** Experimental fluorescent intensity trace produced with a concentration of 1 nM of Cy3 in a mixture of 94% glycerol. **(b1)–(b4)** Concentration of molecules estimated from the trace in (a) with $\ell = 0.5, 1, 1.5, 2$, respectively. The median of the posterior over the concentration is shown by a blue line. The area between the 25% and 75% quantile of the posterior over the concentration is highlighted in pink. For details of the definition of concentration c_k^ℓ and the normalized distance ℓ, see Eq. (S13), below.

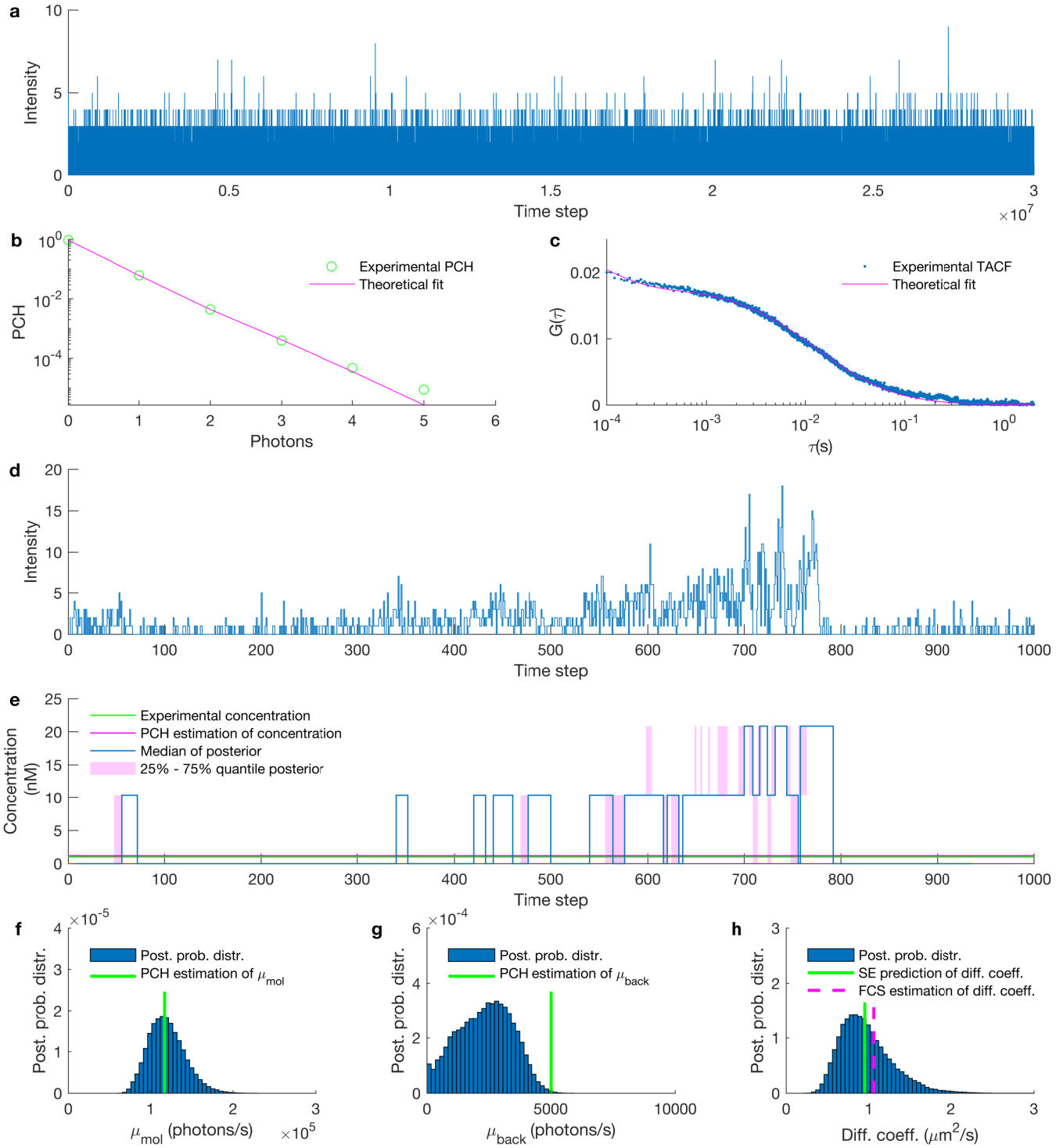

FIG. S9. **Comparison of estimated photon emission rates and concentration with FCS and PCH. (a)** Targeted experimental fluorescent intensity trace. The time step is 10 μs with a total time of 5 min. **(b)** PCH curve and the theoretical fit. **(c)** FCS curve and the best theoretical fit. **(d)** The 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 photon emission rate with the PCH estimated of the molecular photon emission rate 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 dashed line, with the Stokes-Einstein prediction shown by solid green line. The targeted experimental trace is generated by free diffusive Cy3 in a mixture of water and glycerol with 94% 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.

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

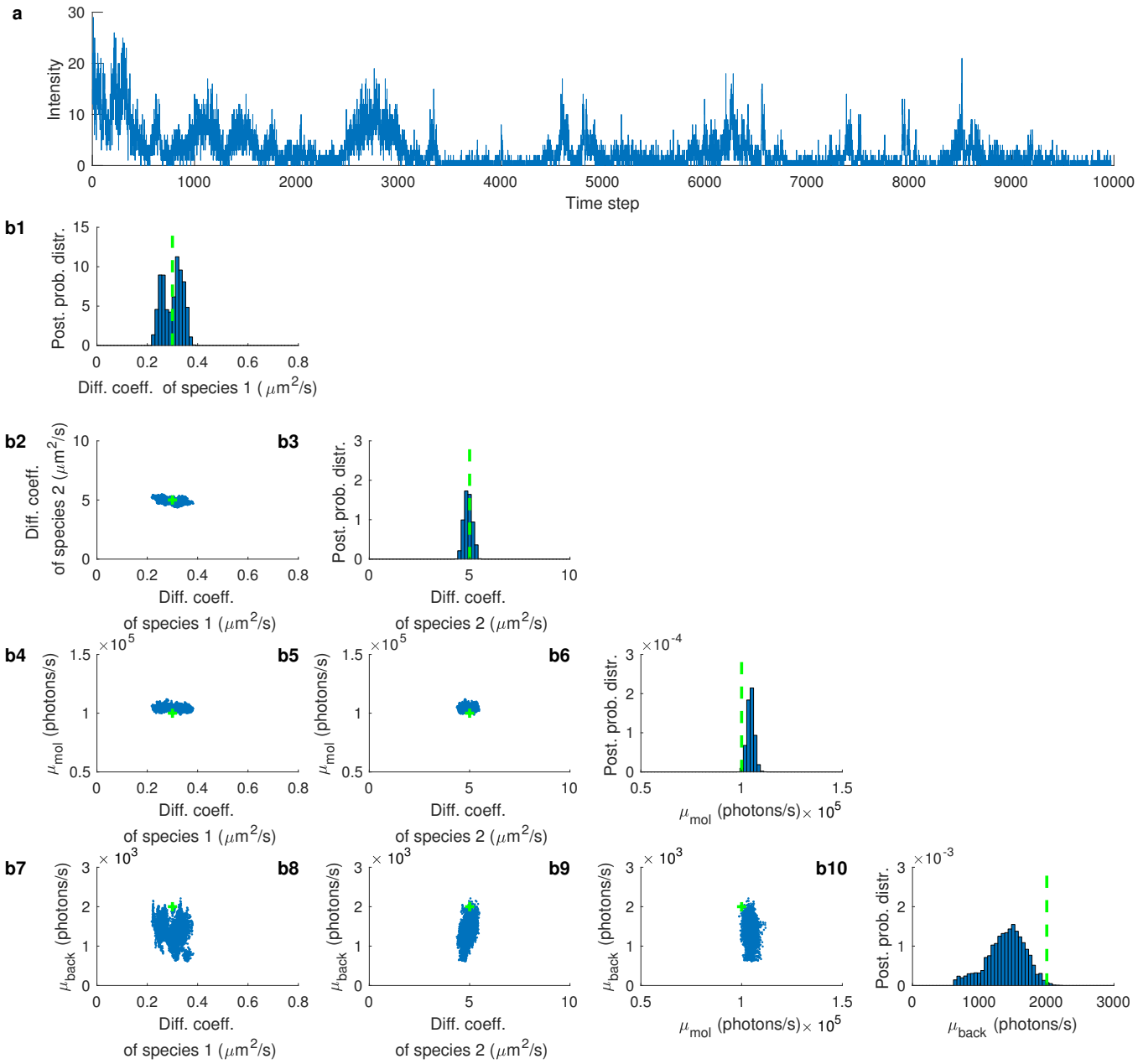

FIG. S10. **Estimated joint posterior probability distribution of multiple diffusive species**. **(a)** A mixed fluorescent intensity trace was obtained by combining the traces from two different synthetic signals with molecular and background emission rates of 10^5 and 10^3 photons/s, respectively, and diffusion coefficients of 0.3 and 5 $\mu m^2/s$. **(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 photon emission rate. **(b5)** The joint probability distribution of diffusion coefficient of diffusive species 2 along with the molecular photon emission rate. **(b6)** The posterior probability distribution of the molecular photon emission rate. **(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 photon emission rate 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 diffusion coefficients are shown by green dashed lines.

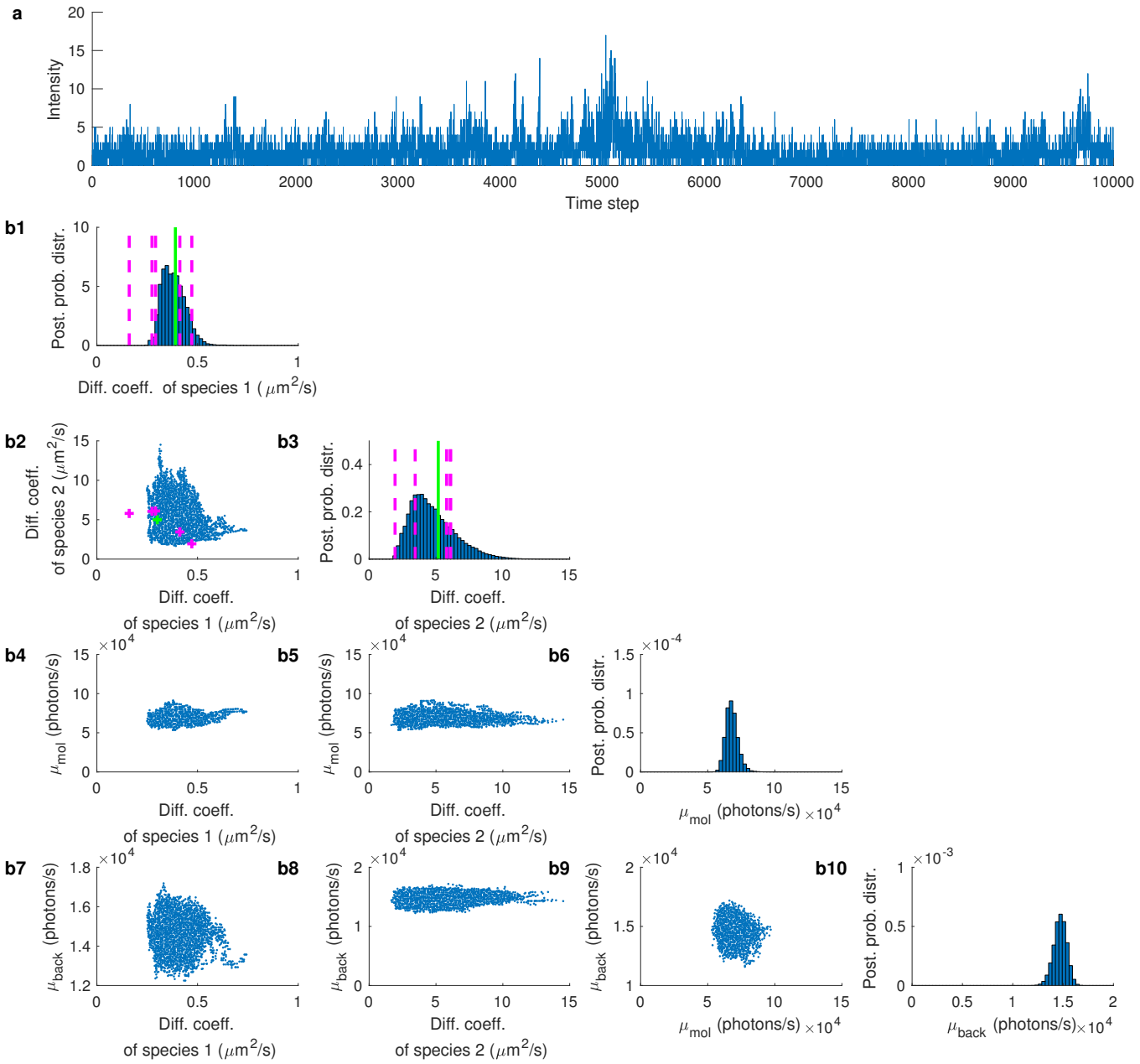

FIG. S11. **Estimated joint posterior probability distribution of multiple diffusive species.** **(a)** Experimental fluorescent intensity trace used in Fig. 8a3 with length 10^4 data points and step $100\mu s$. **(b1)** The posterior probability distribution of the diffusion coefficient of diffusive 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 photon emission rate. **(b5)** The joint probability distribution of diffusion coefficient of diffusive species 2 along with the molecular photon emission rate. **(b6)** The posterior probability distribution of the molecular photon emission rate. **(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 photon emission rate 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 99% and 75% glycerol each. The laser power, wavelength, NA and refractive index are 100 μW, 532 nm, 1.42 and 1.4, respectively. The Stokes-Einstein predictions of the diffusion coefficients is shown by green solid lines and the FCS estimates are shown by a magenta dashed lines.

S2. Summary of point estimates

TABLE S1. Here we list characteristic values (point estimates) summarizing the posterior probability distributions of this study. Mean and SD 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	SD	mean	SD	mean	SD
	$\mu m^2/s$	$\mu m^2/s$	photons/s	photons/s	photons/s	photons/s
Fig. 1(b1)	1.28	0.49	1.05×10^5	3.32×10^4	2.01×10^3	0.32×10^3
Fig. 1(b2)	1.04	0.17	9.96×10^4	6.89×10^3	2.55×10^3	0.55×10^3
Fig. 2(a)	1.28×10^{-2}	0.72×10^{-2}	1.07×10^4	5.87×10^3	1.15×10^3	0.98×10^3
	0.99×10^{-1}	0.19×10^{-1}	2.44×10^4	3.38×10^3	1.06×10^3	1.08×10^3
	1.03	0.29	3.84×10^4	4.69×10^3	2.23×10^3	1.05×10^3
	10.68	1.60	4.87×10^4	5.70×10^3	3.29×10^3	0.62×10^3
Fig. 2(b)	1.12	0.38	1.94×10^4	8.42×10^3	0.58×10^3	0.35×10^3
	1.04	0.16	5.69×10^4	1.59×10^4	0.68×10^3	0.32×10^3
	1.01	0.11	8.02×10^4	8.16×10^3	1.12×10^3	0.12×10^3
	0.99	0.08	9.02×10^4	7.77×10^3	1.10×10^3	0.9×10^2
	1.00	0.06	1.04×10^4	6.63×10^3	1.02×10^3	0.54×10^2
Fig. 3(a1)	1.03	0.13	6.05×10^5	5.55×10^4	0.79×10^3	0.25×10^3
Fig. 3(a2)	1.07	0.25	1.21×10^5	3.07×10^4	0.81×10^3	0.24×10^3
Fig. 3(a3)	0.95	0.21	1.01×10^4	3.67×10^3	0.73×10^3	0.28×10^3
Fig. 3(a4)	0.91	1.03	8.63×10^3	9.89×10^3	1.31×10^3	0.13×10^3
Fig. 4(a1)	1.53	0.72	1.03×10^5	3.63×10^4	2.71×10^3	0.42×10^3
Fig. 4(a2)	1.06	0.24	5.56×10^4	5.43×10^3	9.26×10^3	0.81×10^3
Fig. 5	0.44	0.12	6.03×10^4	1.14×10^4	3.22×10^3	0.18×10^3
	1.06	0.24	5.56×10^4	5.43×10^3	9.26×10^3	0.81×10^3
	4.86	1.23	1.17×10^5	2.44×10^4	1.13×10^4	1.07×10^3
	6.62	1.32	2.82×10^5	1.92×10^4	1.1×10^4	0.96×10^3
Fig. 6(a1)	0.44	0.14	6.32×10^4	1.86×10^4	5.77×10^3	0.57×10^3
Fig. 6(a2)	1.06	0.24	5.56×10^4	5.43×10^3	9.26×10^3	0.81×10^3
Fig. 7(a)	0.13	0.05	1.89×10^5	4.61×10^4	9.97×10^3	2.27×10^3
	0.37	0.09	5.57×10^5	8.14×10^4	3.12×10^4	2.18×10^3
	1.03	0.13	1.28×10^5	1.25×10^4	1.21×10^3	1.25×10^3
	2.03	0.37	3.94×10^5	2.12×10^4	3.18×10^4	1.35×10^3
	3.68	0.64	4.36×10^5	4.04×10^4	2.94×10^4	1.77×10^3
Fig. 7(b1)	1.01	0.23	2.09×10^5	2.10×10^4	1.06×10^4	1.64×10^3
Fig. 7(c1)	0.41	0.16	4.22×10^4	1.54×10^4	2.57×10^3	0.49×10^3
Fig. 7(b2)	1.03	0.13	1.28×10^5	1.25×10^4	1.21×10^3	1.25×10^3
Fig. 7(c2)	1.03	0.13	1.28×10^5	1.25×10^4	1.21×10^3	1.25×10^3
Fig. 8(a1)	0.41	0.11	2.09×10^5	2.10×10^4	1.06×10^4	1.64×10^3
Fig. 8(a2)	5.12	1.43	4.22×10^4	1.54×10^4	2.57×10^3	0.49×10^3
Fig. 8(a3)	0.48	0.09	6.79×10^4	4.58×10^3	1.46×10^4	0.71×10^3
	5.19	1.39				
Fig. S4	1.06	0.22	8.01×10^4	2.04×10^4	1.94×10^3	0.62×10^3
Fig. S5(b1)	1.08	0.29	5.29×10^4	1.22×10^4	1.18×10^3	0.47×10^3
Fig. S5(b2)	1.17	0.30	3.65×10^4	3.96×10^3	0.11×10^3	0.10×10^3
Fig. S9	1.11	0.32	1.19×10^5	2.17×10^4	2.29×10^3	1.11×10^3
Fig. S10	0.29	0.04	1.05×10^5	2.17×10^3	1.45×10^3	0.31×10^3
	5.01	0.43				

S3. Detailed methods description

S3.1. Representation of molecular diffusive motion

Consider a particle satisfying *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 S4. 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

$$x_k|x_{k-1} \sim \textbf{Normal}\left(x_{k-1}, 2(t_k - t_{k-1})D\right) \tag{S4}$$
$$y_k|y_{k-1} \sim \textbf{Normal}\left(y_{k-1}, 2(t_k - t_{k-1})D\right) \tag{S5}$$
$$z_k|z_{k-1} \sim \textbf{Normal}\left(z_{k-1}, 2(t_k - t_{k-1})D\right) \tag{S6}$$

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

$$D = \frac{kT}{6\pi r\eta} \tag{S7}$$

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. Definition of point spread function models

In this study we use two different point spread functions (PSFs) as approximations to the more realistic Airy function [6–8], namely a 3D-Gaussian (3DG) [9] and a 2D Gaussian-Lorentzian (2DG-L) [10–13]. By convention, the maximum value of both PSFs is 1. Further, both PSFs are centered at the center of the excitation PSF which is considered the point of origin.

The definition of the PSF for the 3DG case is

$$\mathrm{PSF}_{\mathrm{3DG}}(x,y,z) = \exp\left(-2\frac{x^2+y^2}{\omega_{xy}^2} - 2\frac{z^2}{\omega_z^2}\right) \tag{S8}$$

where 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 $\omega_z = 1.5 n_{\mathrm{sol}}\lambda_{\mathrm{exc}}/\mathrm{NA}^2$, for example see [14, 15].

The definition of the PSF for the 2DG-L case is

$$\mathrm{PSF}_{\mathrm{2DG-L}} = \frac{1}{1+\left(\frac{z}{z_R}\right)^2} \exp\left(\frac{-2\frac{x^2+y^2}{\omega_{xy}^2}}{1+\left(\frac{z}{z_R}\right)^2}\right), \qquad z_R = \frac{n_{\mathrm{sol}}\pi\omega_{xy}^2}{\lambda_{\mathrm{exc}}} \tag{S9}$$

where ω_{xy}, λ_{exc}, and n_{sol} are similar to the 3DG case.

S3.4. Description of the data simulation

To generate fluorescence intensity time traces that mimic a realistic confocal setup, we simulate molecules moving 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.

Molecular locations at the first assessment x_1^n, y_1^n, z_1^n are sampled randomly from the 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

$$\mu_k = \mu_{back} + \mu_{mol} \sum_{n=1}^{N} \mathrm{PSF}(x_k^n, y,_k^n z_k^n) \tag{S10}$$

where both background and (maximum) molecular photon emission rates, 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. (S8)–(S9), above.

Detailed parameter choices for all simulations performed are listed in Table S5.

S3.5. Definition of normalized distance and molecular concentration

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 molecular concentrations, 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. For this, for a molecule at (x_k^n, y_k^n, z_k^n), when the 3DG PSF is used, Eq. (S8), we rely on

$$d_k^n = \sqrt{\left(\frac{x_k^n}{\omega_{xy}}\right)^2 + \left(\frac{y_k^n}{\omega_{xy}}\right)^2 + \left(\frac{z_k^n}{\omega_z}\right)^2} \tag{S11}$$

while, when the 2DG-L PSF is used, Eq. (S9), we rely on

$$d_k^n = \sqrt{\frac{1}{2}\log\left(1+\left(\frac{z_k^n}{z_R}\right)^2\right) + \frac{\left(\frac{x_k^n}{\omega_{xy}}\right)^2 + \left(\frac{y_k^n}{\omega_{xy}}\right)^2}{1+\left(\frac{z_k^n}{z_R}\right)^2}} \tag{S12}$$

These distances are obtained by setting the 3DG and 2DG-L 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 concentration of molecules c_k^ℓ as the number of estimated (active) molecules within the corresponding volume

$$c_k^\ell = \frac{1}{V_\ell} \sum_n b^n H\left(1 - \frac{d_k^n}{\ell}\right) \tag{S13}$$

where H is the Heaviside step function, b^n is the load of molecule n, and V_ℓ is the volume of the designated effective region chosen to agree with FCS [16]. For the 3DG case, this is given by

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

while, for the 2DG-L case, it is given by

$$V_\ell = \pi^{\frac{3}{2}} \omega_{xy}^2 z_R \ell^3. \tag{S15}$$

S3.6. 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).

Briefly, the user specifies a minimum number of photons per bin as a lower threshold. As illustrated in Fig. S12, 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,

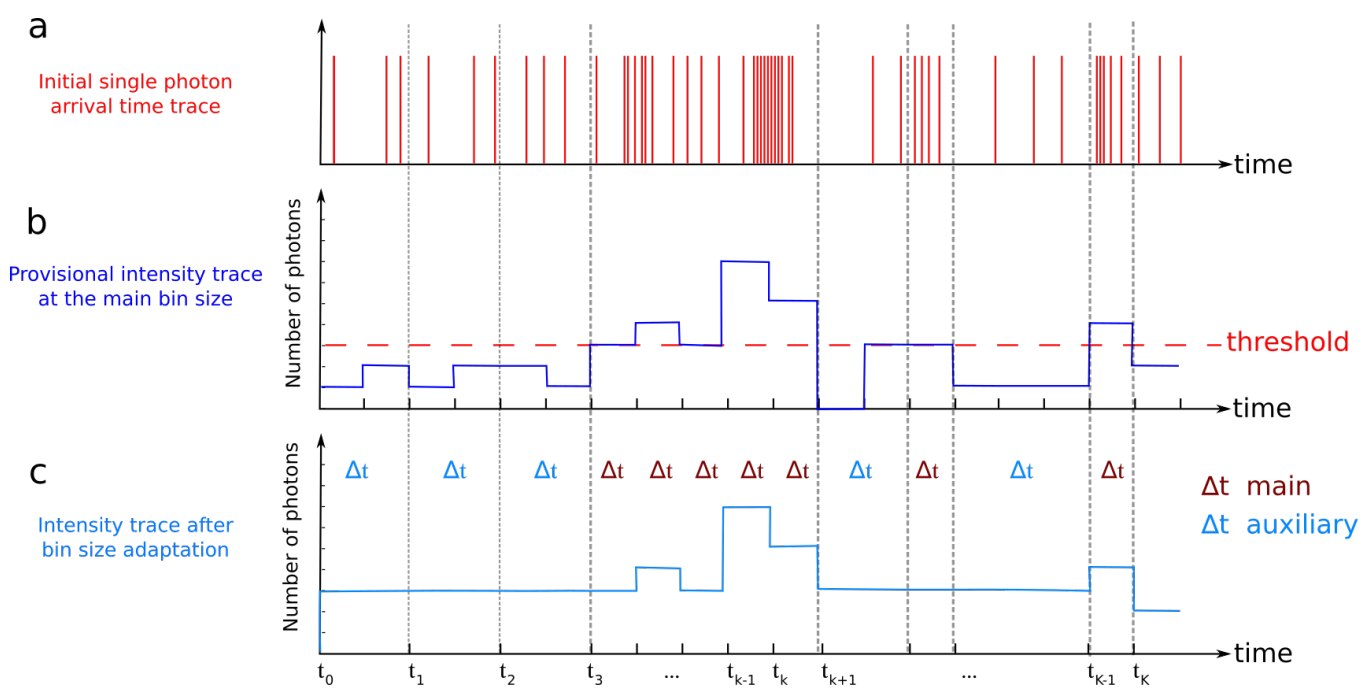

FIG. S12. **Illustration of time trace preparation.** **(a)** Initial trace of single photon arrivals. Here, each vertical line represents the arrival time of a single photon. **(b)** Time trace of photon intensities provisionally binned at the main bin size. The horizontal line denotes the imposed lower 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.

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, 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. 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

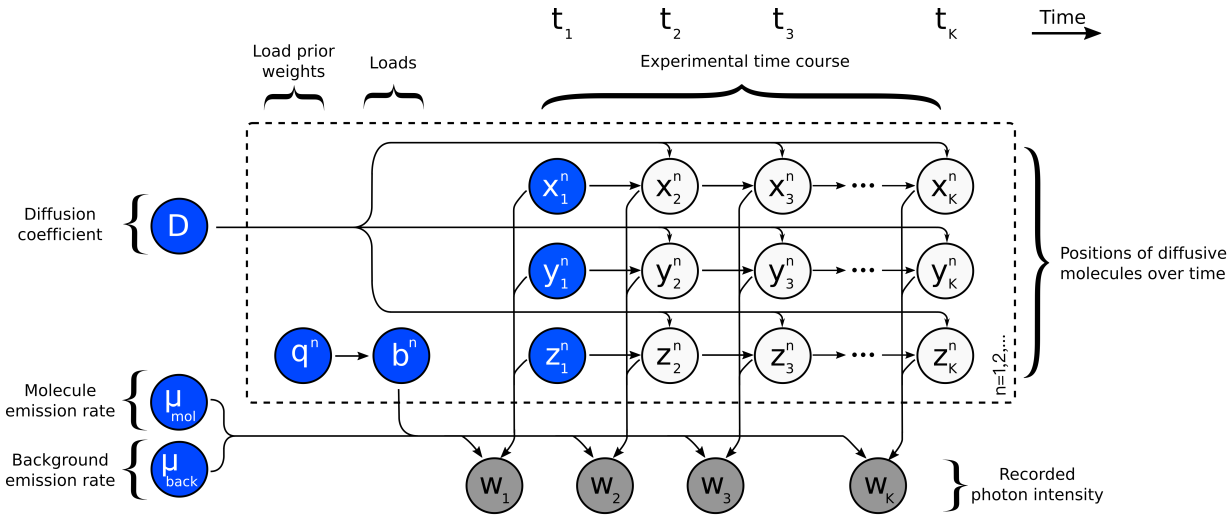

FIG. S13. **Graphical summary of the framework developed.** A population of model molecules, labeled by $n = 1, 2, \ldots,$ evolves during the measurement period which is marked by $k = 1, 2, \ldots, K$. Here, x_k^n, y_k^n and z_k^n denote the location of molecule n at time t_k and μ_{mol}, μ_{back} denote molecular and background photon emission rates. The diffusion coefficient D determines the evolution of the molecular 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 molecular 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 and background photon emission rates μ_{mol} and μ_{back}; the initial molecular 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

$$D \sim \mathbf{InvGamma}\left(\alpha_D, \beta_D\right). \tag{S16}$$

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 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

$$\mu_{mol} \sim \mathbf{Gamma}\left(\alpha_{mol}, \beta_{mol}\right) \tag{S17}$$

$$\mu_{back} \sim \mathbf{Gamma}\left(\alpha_{back}, \beta_{back}\right). \tag{S18}$$

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 photon emission rates becomes available.

S4.1.3. Priors on initial molecular 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 S4, on each Cartesian coordinate of the model molecules

$$x_1^n \sim \textbf{SymNormal}\left(\mu_{xy}, \sigma_{xy}^2\right) \tag{S19}$$

$$y_1^n \sim \textbf{SymNormal}\left(\mu_{xy}, \sigma_{xy}^2\right) \tag{S20}$$

$$z_1^n \sim \textbf{SymNormal}\left(\mu_z, \sigma_z^2\right). \tag{S21}$$

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 molecular 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

$$b^n | q^n \sim \textbf{Bernoulli}(q^n) \tag{S22}$$

$$q^n \sim \textbf{Beta}\left(A_q, B_q\right). \tag{S23}$$

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 [17–20], 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

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

$$\mu_{mol} \sim \textbf{Gamma}\left(\alpha_{mol}, \beta_{mol}\right) \tag{S25}$$
$$\mu_{back} \sim \textbf{Gamma}\left(\alpha_{back}, \beta_{back}\right) \tag{S26}$$

$$q^n \sim \textbf{Beta}\left(\frac{\alpha_q}{N}, \beta_q \frac{N-1}{N}\right) \tag{S27}$$
$$b^n | q^n \sim \textbf{Bernoulli}(q^n) \tag{S28}$$

$$x_1^n \sim \textbf{SymNormal}\left(\mu_{xy}, \sigma_{xy}^2\right) \tag{S29}$$
$$y_1^n \sim \textbf{SymNormal}\left(\mu_{xy}, \sigma_{xy}^2\right) \tag{S30}$$
$$z_1^n \sim \textbf{SymNormal}\left(\mu_z, \sigma_z^2\right) \tag{S31}$$

$$x_k^n | x_{k-1}^n, D \sim \textbf{Normal}\left(x_{k-1}^n, 2(t_k - t_{k-1})D\right), \qquad k = 2, \ldots, K \tag{S32}$$
$$y_k^n | y_{k-1}^n, D \sim \textbf{Normal}\left(y_{k-1}^n, 2(t_k - t_{k-1})D\right), \qquad k = 2, \ldots, K \tag{S33}$$
$$z_k^n | z_{k-1}^n, D \sim \textbf{Normal}\left(z_{k-1}^n, 2(t_k - t_{k-1})D\right), \qquad k = 2, \ldots, K \tag{S34}$$

$$w_k | \{x_k^n, y_k^n, z_k^n, b^n\}_n, \mu_{mol}, \mu_{back} \sim \textbf{Poisson}\left(\mu_k\right), \qquad k = 1, \ldots, K \tag{S35}$$
$$\mu_k = (t_k - t_{k-1})\left(\mu_{back} + \mu_{mol} \sum_n b^n \ \mathrm{PSF}(x_k^n, y_k^n, z_k^n)\right). \tag{S36}$$

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

$$\overline{x}^n = (x_1^n, x_2^n, \ldots, x_K^n) \tag{S37}$$
$$\overline{y}^n = (y_1^n, y_2^n, \ldots, y_K^n) \tag{S38}$$
$$\overline{z}^n = (z_1^n, z_2^n, \ldots, z_K^n) \tag{S39}$$
$$\overline{w} = (w_1, w_2, \ldots, w_K). \tag{S40}$$

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 [21–24]. 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

$$\eta_1 = \frac{\sum_{i=2I/4}^{3I/4} D_i}{I/4}, \qquad \eta_2 = \frac{\sum_{i=3I/4}^{I} D_i}{I/4} \tag{S41}$$

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 [21, 22], 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. S14, are available through the Supporting Material.

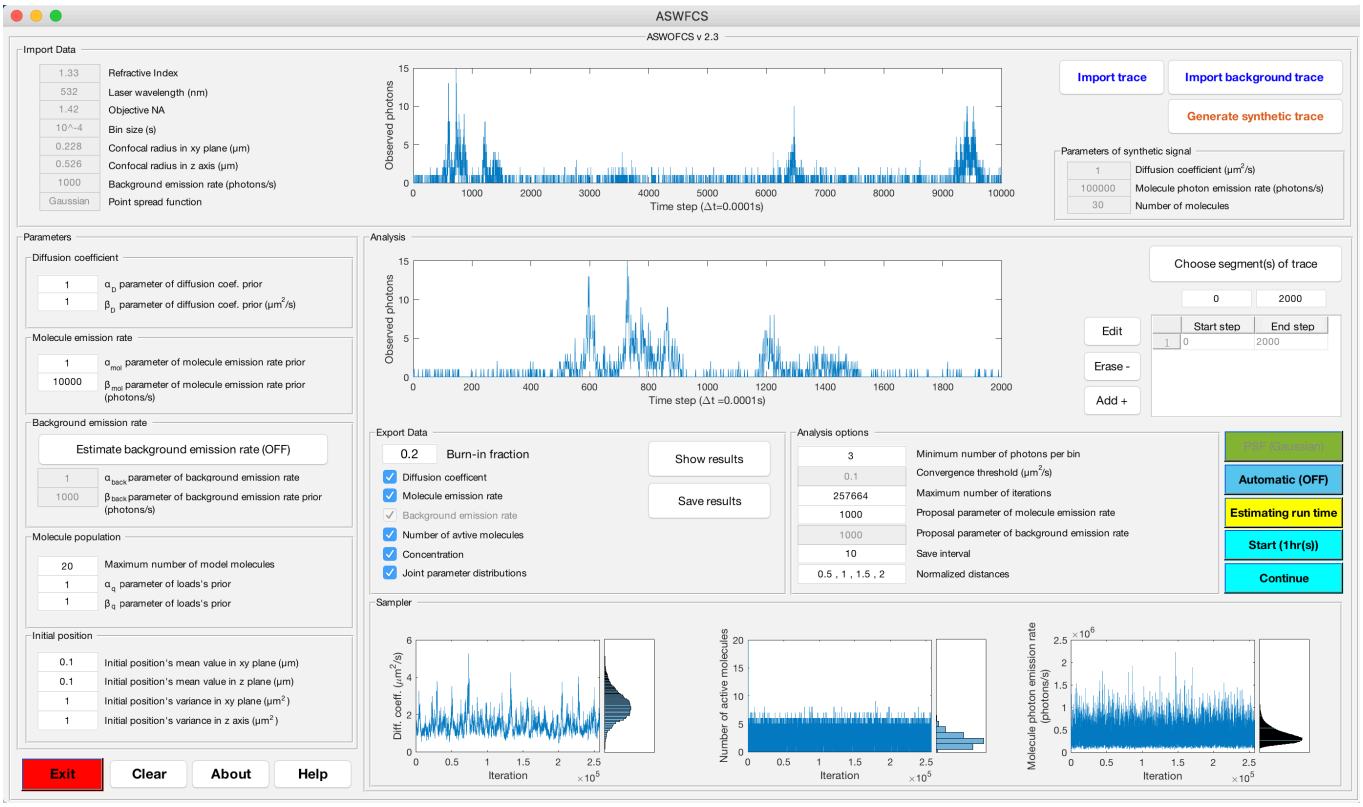

FIG. S14. **A working implementation of the framework described in this study is available through the** SUPPORTING MATERIAL**.** Along with this implementation, we provide a graphical user interface (GUI) that can be used to analyze intensity traces from confocal microscopy.

S4.3.1. Overview of the sampling updates

Our MCMC exploits a Gibbs sampling scheme [21–23]. 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 molecular 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 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. S13, 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 and background photon emission rates μ_{mol} and μ_{back}, respectively

These steps are described in detail below.

S4.3.2. Sampling of the locations of active molecules

For a given active molecule n, we update the trajectory $\overline{x}^n$ by sampling from the corresponding conditional $p(\overline{x}^n|D, \mu_{mol}, \mu_{back}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \overline{w})$. Once $\overline{x}^n$ is updated, we update $\overline{y}^n$ and $\overline{z}^n$ similarly.

To sample the trajectory [25, 26], we perform forward filtering-backward sampling (FFBS) on an approximate model [27], where Eq. (S35), is replaced with

$$\mathbb{T}_{data}(w_k)|\{x_k^{n'}, y_k^{n'}, z_k^{n'}, b^{n'}\}_{n'}, \mu_{mol} \sim \textbf{Normal}\left(\mathbb{T}_{mean}(\mu_k), \mathbb{S}(\mu_k)^2\right), \qquad k = 1, \ldots, K. \tag{S42}$$

Here, μ comes from Eq. (S36) and $\mathbb{T}_{data}(w)$, $\mathbb{T}_{mean}(\mu)$ and $\mathbb{S}(\mu)^2$ denote Anscombe transformed [28] variables defined as follows

$$\mathbb{T}_{data}(w) = 2\sqrt{w + \frac{3}{8}} \tag{S43}$$

$$\mathbb{T}_{mean}(\mu) = 2\sqrt{\mu + \frac{3}{8}} - \frac{1}{4\sqrt{\mu}} + \frac{1}{32\sqrt{\mu^3}} \tag{S44}$$

$$\mathbb{S}(\mu)^2 = 1 + \frac{1}{16\mu^2}. \tag{S45}$$

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.

We carry out the forward filtering [26, 29] by computing a filter distribution $p(x_k^n|\{x_{k'}^n\}_{k'<k}, ...)$ on the approximate model similar to the standard theory underlying nonlinear Kalman filters [30–33]. More specifically, due to the symmetry of the PSF across the yz-plane, the filtering distribution consists of two modes centered symmetrically across the origin. As a result, we compute a filter distribution of the form

$$p\left(x_k^n|\{x_{k'}^n\}_{k'<k}, D, \mu_{mol}, \mu_{back}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \overline{w}\right) = \textbf{SymNormal}\left(x_k^n; m_k^n, c_k^n\right) \tag{S46}$$

where $\textbf{SymNormal}\left(m_k^n, c_k^n\right)$ denotes the symmetric normal distribution (see Table S4). The filter is updated recursively according to

$$\begin{aligned} &p\left(x_k^n|\{x_{k'}^n\}_{k'<k}, D, \mu_{mol}, \mu_{back}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \overline{w}\right) \propto \\ &p\left(\overline{w}|x_k^n, y_k^n, z_k^n, \mu_{mol}, \mu_{back}, \{b^{n'}, x^{n'}, y^{n'}, z^{n'}\}_n'\right) \times \\ &\int_{x_{k-1}^n} p\left(x_{k-1}^n|\{x_{k'}^n\}_{k'<k}, D, \mu_{mol}, \mu_{back}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \overline{w}\right) p\left(x_k^n|x_{k-1}^n, D\right) dx_{k-1}^n. \end{aligned} \tag{S47}$$

Due to the specific choices of our problem (i.e. diffusive kinetics, symmetric normal filter at the preseeding time, and normal likelihood), the previous expression reduces to

$$\begin{aligned} &p\left(x_k^n|\{x_{k'}^n\}_{k'<k}, D, \mu_{mol}, \mu_{back}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \overline{w}\right) \\ &\quad = \textbf{Normal}\left(\mathbb{T}_{data}(w_k); \mathbb{T}_{mean}(\mu_k), \mathbb{S}(\mu_k)^2\right) \textbf{SymNormal}\left(x_k^n; m_{k-1}^n, c_{k-1}^n + 2D\left(t_k - t_{k-1}\right)\right) \end{aligned}. \tag{S48}$$

which, in turn, is approximated as

$$p\left(x_k^n|\{x_{k'}^n\}_{k'<k}, D, \mu_{mol}, \mu_{back}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \overline{w}\right) \approx \textbf{SymNormal}\left(x_k^n; m_k^n, c_k^n\right) \tag{S49}$$

Shortly, we discuss these approximations in detail.

S4.3.2.a. Backward sampling

Given a computed filter, we sample sequentially x_k^n according to

$$x_K^n \sim p\left(x_K^n | \{x_{k'}^n\}_{k'<K}, D, \mu_{mol}, \mu_{back}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \overline{w}\right) \tag{S50}$$

$$x_k^n \sim p\left(x_k^n | x_{k+1}^n, \{x_{k'}^n\}_{k'<k}, D, \mu_{mol}, \mu_{back}, \{b^{n'}, \overline{y}^{n'}, \overline{z}^{n'}\}_{n'}, \{\overline{x}^{n'}\}_{n'\neq n}, \overline{w}\right), \qquad k = 1, \ldots, K-1. \tag{S51}$$

Due to the specific choices of our problem these reduce to

$$x_K^n \sim \mathbf{SymNormal}\left(m_K^n, c_K^n\right) \tag{S52}$$

$$x_k^n \sim \mathbf{SymNormal}\left(m_k^n, c_k^n\right) \times \mathbf{Normal}\left(x_{k+1}^n, 2D(t_{k+1} - t_k)\right), \qquad k = 1, \ldots, K-1 \tag{S53}$$

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

S4.3.2.b. Forward filtering

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) [30] and **(ii)** Unscented Kalman filter (UKF) [31–33].

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. (S48) is linearized in order for their product to become a symmetric normal one. The resulting computations for $k = 1$ are

$$\begin{aligned}
c_1^n &= \frac{\mathbb{S}(\mu_1)^2 \sigma_{xy}^2}{\mathbb{S}(\mu_1)^2 + \sigma_{xy}^2 (d_1^n)^2} \\
m_1^n &= \mu_{xy} + \frac{c_1^n d_1^n e_1^n}{\mathbb{S}(\mu_1)^2} \\
d_1^n &= \frac{\partial \mathbb{T}_{mean}(\mu_1)}{\partial x_1^n}\Big|_{x_1^n = \mu_{xy}} \\
e_1^n &= \mathbb{T}_{data}(w_1) - \mathbb{T}_{mean}(\mu_1)\Big|_{x_1^n = \mu_{xy}}
\end{aligned} \tag{S54}$$

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

$$\begin{aligned}
c_k^n &= \frac{\mathbb{S}(\mu_k)^2 \left(c_{k-1}^n + 2D\left(t_k - t_{k-1}\right)\right)}{\mathbb{S}(\mu_k)^2 + \left(c_{k-1}^n + 2D\left(t_k - t_{k-1}\right)\right)(d_k^n)^2} \\
m_k^n &= m_{k-1}^n + \frac{c_k^n d_k^n e_k^n}{\mathbb{S}(\mu_k)^2}. \\
d_k^n &= \frac{\partial \mathbb{T}_{mean}(\mu_k)}{\partial x_k^n}\Big|_{x_k^n = m_{k-1}^n} \\
e_k^n &= \mathbb{T}(w_k) - \mathbb{T}(\mu_k)\Big|_{x_k^n = m_{k-1}^n}.
\end{aligned} \tag{S55}$$

Unscented Kalman filter

The unscented Kalman filter [31, 32] 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. (S48). Specifically the product of Eq. (S48) is approximated as follows

$$\begin{aligned}
&\mathbf{Normal}\left(\mathbb{T}_{data}(w_k);\mathbb{T}_{mean}(\mu_k),\mathbb{S}(\mu_k)^2\right)\mathbf{SymNormal}\left(x_k^n;m_{k-1}^n,c_{k-1}^n+2D\left(t_k-t_{k-1}\right)\right)\\
&\approx\frac{1}{2}\mathbf{BNormal}\left(\begin{bmatrix}X_k^n\\W_k^n\end{bmatrix},\begin{bmatrix}{}_{xx}\Sigma_k^n & , & {}_{xw}\Sigma_k^n\\{}_{wx}\Sigma_k^n & , & {}_{ww}\Sigma_k^n\end{bmatrix}\right)+\frac{1}{2}\mathbf{BNormal}\left(\begin{bmatrix}-X_k^n\\W_k^n\end{bmatrix},\begin{bmatrix}{}_{xx}\Sigma_k^n & , & {}_{xw}\Sigma_k^n\\{}_{wx}\Sigma_k^n & , & {}_{ww}\Sigma_k^n\end{bmatrix}\right)\\
&\propto\frac{1}{2}\mathbf{Normal}\left(x_k^n;m_k^n,c_k^n\right)+\frac{1}{2}\mathbf{Normal}\left(x_k^n;-m_k^n,c_k^n\right)\\
&=\mathbf{SymNormal}\left(x_k^n;m_k^n,c_k^n\right)
\end{aligned}\tag{S56}$$

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. The mean and variance of one mode of the filter are given by

$$m_k^n = X_k^n + K_k^n(\mathbb{T}(w_k) - \mathbb{T}\left(\mu_k\left(X_k^n\right)\right) \tag{S57}$$

$$c_k^n = {}_{xx}\Sigma_k^n - K_k^n({}_{ww}\Sigma_k^n)(K_k^n)^T \tag{S58}$$

$$K_k^n = \frac{{}_{xw}\Sigma_k^n}{{}_{ww}\Sigma_k^n}. \tag{S59}$$

To calculate the mean value and variance 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, 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 and weights for a standard normal as x^{sn} and g^* [34] and then we transform them according to the previous filter's mean value and variance

$$x_i^* = m_{k-1}^n + x^{sn}\sqrt{c_{k-1}^n + 2D\left(t_k - t_{k-1}\right)}. \tag{S60}$$

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

$$\begin{aligned}
X_k^n &= \sum_{i=1} g_i^* x_i^*\\
W_k^n &= \sum_{i=1} g_i^* \mathbb{T}_{mean}\left(\mu_k\left(x_i^*\right)\right)
\end{aligned}\tag{S61}$$

$$\begin{aligned}
{}_{xx}\Sigma_k^n &= \sum_{i=1} g_i^*\left(X_k^n - x_i^*\right)^T\left(X_k^n - x_i^*\right)\\
{}_{ww}\Sigma_k^n &= \sum_{i=1} g_i^*\left(W_k^n - \mathbb{T}_{mean}\left(\mu_k\left(x_i^*\right)\right)\right)^T\left(W_k^n - \mathbb{T}_{mean}\left(\mu_k\left(x_i^*\right)\right)\right) + \mathbb{S}\left(\mu_k\right)^2\\
{}_{wx}\Sigma_k^n = {}_{xw}\Sigma_k^n &= \sum_{i=1} g_i\left(W_k^n - \mathbb{T}_{mean}\left(\mu_k\left(x_i^*\right)\right)\right)^T\left(X_k^n - x_i^*\right).
\end{aligned}\tag{S62}$$

S4.3.3. Sampling of the locations of inactive molecules

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. (S16), (S32), (S33) and (S34), simplifies to $p(D|\{\overline{x}^n, \overline{y}^n, \overline{z}^n\}_n)$. Because of conjugacy, the latter reduces to

$$D|\{\overline{x}^n, \overline{y}^n, \overline{z}^n\}_n \sim \mathbf{InvGamma}\left(\alpha', \beta'\right) \tag{S63}$$

where α' and β' are given by

$$\begin{aligned} \alpha' =& \alpha_D + \frac{3N(K-1)}{2} \\ \beta' =& \beta_D + \frac{1}{4}\sum_{n=1}^{N}\sum_{k=1}^{K-1} \frac{\left(x_{k+1}^n - x_k^n\right)^2 + \left(y_{k+1}^n - y_k^n\right)^2 + \left(z_{k+1}^n - z_k^n\right)^2}{t_{k+1} - t_k} \end{aligned} \tag{S64}$$

S4.3.5. Sampling of 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. (S28) and (S27), and because of conjugacy, the latter distribution is sampled by sampling each q^n separately according to

$$p(q^n|b^n) \propto p(b^n|q^n)p(q^n) = \mathbf{Beta}\left(q^n; \frac{\alpha_q}{N} + b^n, \frac{\beta_q(N-1)}{N} + 1 - b^n\right). \tag{S65}$$

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

$$(b^n)^{prop} \sim \mathbf{Bernoulli}(q^n). \tag{S66}$$

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

$$\begin{aligned} r_b = \prod_{k=1}^{K} \Bigg[& \left(\frac{\mu_{back} + \mu_{mol}\sum_{n=1}^{N}(b^n)^{prop}\,\mathbf{PSF}\left(x_k^n, y_k^n, z_k^n\right)}{\mu_{back} + \mu_{mol}\sum_{n=1}^{N}(b^n)^{old}\,\mathbf{PSF}\left(x_k^n, y_k^n, z_k^n\right)} \right)^{w_k} \\ & \times \exp\left(-(t_k - t_{k-1})\mu_{mol}\sum_{n=1}^{N}\left((b^n)^{old} - (b^n)^{prop}\right)\mathbf{PSF}\left(x_k^n, y_k^n, z_k^n\right) \right) \Bigg] \end{aligned} \tag{S67}$$

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

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

Finally, after we updated the locations of molecules, and loads, we update the molecular 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

$$\begin{aligned} \mu_{mol}^{prop} &\sim \mathbf{Gamma}\left(\alpha_{mol}^{prop}, \frac{\mu_{mol}^{old}}{\alpha_{mol}^{prop}}\right) \\ \mu_{back}^{prop} &\sim \mathbf{Gamma}\left(\alpha_{back}^{prop}, \frac{\mu_{back}^{old}}{\alpha_{back}^{prop}}\right) \end{aligned} \tag{S68}$$

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

$$
\begin{aligned}
r_\mu = \prod_{k=1}^{K} \Bigg[& \left(\frac{\mu_{back}^{prop} + \mu_{mol}^{prop} \sum_{n=1}^{N} b^n \mathbf{PSF}\left(x_k^n, y_k^n, z_k^n\right)}{\mu_{back}^{old} + \mu_{mol}^{old} \sum_{n=1}^{N} b^n \mathbf{PSF}\left(x_k^n, y_k^n, z_k^n\right)} \right)^{w_k} \\
& \times \exp\left(\left(t_k - t_{k-1}\right) \left(\left(\mu_{back}^{old} - \mu_{back}^{prop}\right) + \left(\mu_{mol}^{old} - \mu_{mol}^{prop}\right) \sum_{n=1}^{N} b^n \mathbf{PSF}\left(x_k^n, y_k^n, z_k^n\right) \right) \right) \Bigg] \\
& \times \left(\frac{\mu_{mol}^{old}}{\mu_{mol}^{prop}} \right)^{2\alpha_{mol}^{prop} - \alpha_{mol}} \exp\left(\frac{\mu_{mol}^{old} - \mu_{mol}^{prop}}{\beta_{mol}} + \alpha_{mol}^{prop} \left(\frac{\mu_{mol}^{prop}}{\mu_{mol}^{old}} - \frac{\mu_{mol}^{old}}{\mu_{mol}^{prop}} \right) \right) \\
& \times \left(\frac{\mu_{back}^{old}}{\mu_{back}^{prop}} \right)^{2\alpha_{back}^{prop} - \alpha_{back}} \exp\left(\frac{\mu_{back}^{old} - \mu_{back}^{prop}}{\beta_{back}} + \alpha_{back}^{prop} \left(\frac{\mu_{back}^{prop}}{\mu_{back}^{old}} - \frac{\mu_{back}^{old}}{\mu_{back}^{prop}} \right) \right).
\end{aligned}
\tag{S69}
$$

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

$$
\begin{aligned}
{}_1D &\sim \textbf{InvGamma}\left(\alpha_D, \beta_D\right) && \text{(S70)}\\
{}_2D &\sim \textbf{InvGamma}\left(\alpha_D, \beta_D\right) && \text{(S71)}\\
\mu_{mol} &\sim \textbf{Gamma}\left(\alpha_{mol}, \beta_{mol}\right) && \text{(S72)}\\
\mu_{back} &\sim \textbf{Gamma}\left(\alpha_{back}, \beta_{back}\right) && \text{(S73)}\\
{}_1q^n &\sim \textbf{Beta}\left(\frac{\alpha_q}{{}_1N}, \beta_q \frac{{}_1N-1}{{}_1N}\right) && \text{(S74)}\\
{}_1b^n|{}_1q^n &\sim \textbf{Bernoulli}({}_1q^n) && \text{(S75)}\\
{}_2q^n &\sim \textbf{Beta}\left(\frac{\alpha_q}{{}_2N}, \beta_q \frac{{}_2N-1}{{}_2N}\right) && \text{(S76)}\\
{}_2b^n|{}_2q^n &\sim \textbf{Bernoulli}({}_2q^n) && \text{(S77)}\\
{}_1x_1^n &\sim \textbf{SymNormal}\left(\mu_{xy}, \sigma_{xy}^2\right) && \text{(S78)}\\
{}_1y_1^n &\sim \textbf{SymNormal}\left(\mu_{xy}, \sigma_{xy}^2\right) && \text{(S79)}\\
{}_1z_1^n &\sim \textbf{SymNormal}\left(\mu_z, \sigma_z^2\right) && \text{(S80)}\\
{}_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 && \text{(S81)}\\
{}_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 && \text{(S82)}\\
{}_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 && \text{(S83)}\\
{}_2x_1^n &\sim \textbf{SymNormal}\left(\mu_{xy}, \sigma_{xy}^2\right) && \text{(S84)}\\
{}_2y_1^n &\sim \textbf{SymNormal}\left(\mu_{xy}, \sigma_{xy}^2\right) && \text{(S85)}\\
{}_2z_1^n &\sim \textbf{SymNormal}\left(\mu_z, \sigma_z^2\right) && \text{(S86)}\\
{}_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 && \text{(S87)}\\
{}_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 && \text{(S88)}\\
{}_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 && \text{(S89)}\\
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 && \text{(S90)}
\end{aligned}
$$

$$
\mu_k = (t_k - t_{k-1})\left(\mu_{back} + \mu_{mol}\sum_n {}_1b^n \mathrm{PSF}({}_1x_k^n, {}_1y_k^n, {}_1z_k^n) + \mu_{mol}\sum_n {}_2b^n \mathrm{PSF}({}_2x_k^n, {}_2y_k^n, {}_2z_k^n)\right) \qquad \text{(S91)}
$$

where pre-scripts 1 and 2 are used to distinguish the two species. A graphical summary is show on Fig. S15. We use this formulation for the estimates shown on Figs. S10 and S11.

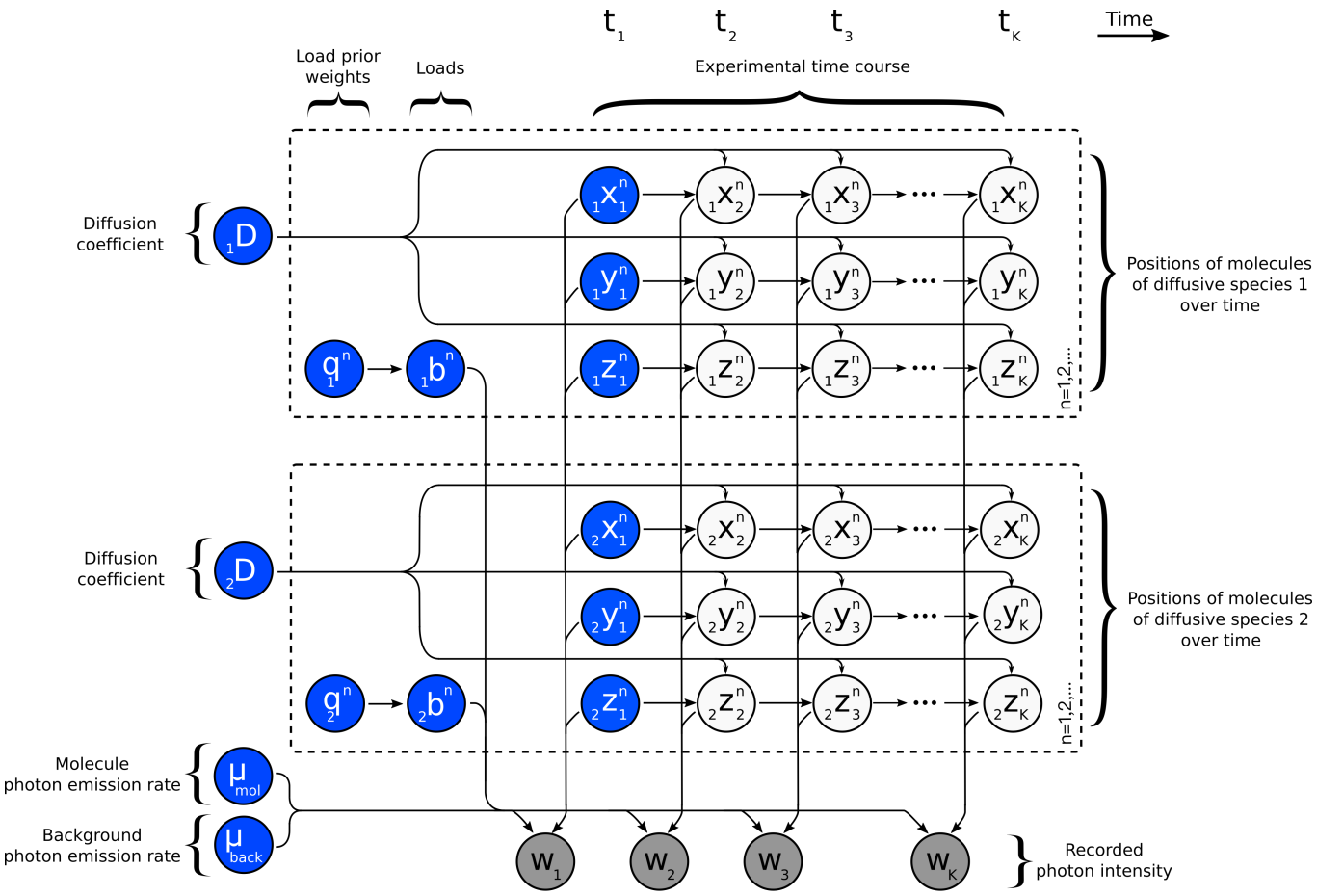

FIG. S15. **Graphical summary of the framework capturing two independent diffusion coefficients.** A multi-species population of model molecules, labeled by $n = 1, 2, \ldots$, evolves during the measurement period which is marked by $k = 1, 2, \ldots, K$. Here, ${}_1x_k^n$, ${}_1y_k^n$ and ${}_1z_k^n$ 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 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 S2. Summary of notation.

Description	Variable	Units
Diffusion coefficient	D	μm^2/s
α parameter of the diffusion coefficient prior	α_D	-
β parameter of the diffusion coefficient prior	β_D	μm^2/s
Total time trace duration	T_{total}	s
Molecular photon emission rate (maximum)	μ_{mol}	photons/s
α parameter of the molecular photon emission rate's prior	α_{mol}	-
β parameter of the molecular photon emission rate's prior	β_{mol}	photons/s
Proposal parameter of the molecular 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 molecular position's prior in the xy-plane	μ_{xy}	μm
Mean value of initial molecular position's prior on the z-axis	μ_z	μm
Variance of the initial molecular position's prior in the xy-plane	σ_{xy}	μm
Variance of the initial molecular 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 S3. List of abbreviations.

Phrase	Abbreviation
Fuorescence correlation spectroscopy	FCS
Fluorescence cross-correlation spectroscopy	FCCS
Spot variation fluorescence correlation spectroscopy	svFCS
Fluorescence resonance energy transfer-fluorescence correlation spectroscopy	FCS-FRET
Point spread function	PSF
Three dimensional Gaussian	3DG
Two dimensional Gaussian-Lorentzian	2DG-L
Forward filtering backward sampling	FFBS
Markov chain Monte Carlo	MCMC
Graphical user interface	GUI

TABLE S4. 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 S5. 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	0.23	0.55	150	1	10^5	10^3	0.1	10^{-4}
Fig. 1(a2)	3DG	2	3	0.23	0.55	600	1	10^5	10^3	0.1	10^{-4}
Fig. 2(a)	3DG	2	3	0.23	0.55	150	$10^{-2}, 10^{-1}, 1, 10$	10^5	10^3	0.1	10^{-4}
Fig. 2(c)	3DG	2	3	0.23	0.55	150	1	10^5	10^3	10	10^{-4}
Fig. 3(a)	3DG	2	3	0.23	0.55	150	1	$10^6, 10^5, 10^4, 10^3$	10^3	0.1	10^{-4}
Fig. S3	3DG	2	3	0.23	0.55	150	1	10^5	10^3	0.1	10^{-4}
Fig. S4(a)	3DG	2	3	0.23	0.55	150	1	10^5	10^3	0.1	10^{-4}
Fig. S5(a1)	3DG	2	3	0.23	0.55	150	1	10^5	10^3	0.1	10^{-4}
Fig. S5(a2)	2DG-L	2	3	0.23	0.55	150	1	10^5	10^3	0.1	10^{-4}
Fig. S10(a)	3DG	2	3	0.23	0.55	150	0.3, 5	10^5	10^3	1	10^{-4}

TABLE S6. 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	0.23	0.55	50	1	1	1	10^4	10^3	1	10^3	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 1(a2)	3DG	0.23	0.55	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 2(a)	3DG	0.23	0.55	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 2(b)	3DG	0.23	0.55	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 3	3DG	0.23	0.55	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 4(a1)	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 4(a2)	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 5	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
	3DG	0.23	0.57	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
	3DG	0.23	0.57	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 6(a1)	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 6(a2)	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 7(a)	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
	3DG	0.23	0.57	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
	3DG	0.23	0.56	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
	3DG	0.23	0.55	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 7(b1)	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 7(b2)	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 7(c1)	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 7(c2)	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 8(a1)	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 8(a2)	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. 8(a3)	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. S3	3DG	0.23	0.55	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. S4	3DG	0.23	0.55	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. S5(b1)	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. S5(b2)	2DG-L	0.23	0.43	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. S8	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. S9	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	10^3	1	1	0.2	0.2	2	2	0.1	4
Fig. S10	3DG	0.23	0.58	50	1	1	1	10^4	10^3	1	10^4	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] 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.
[7] 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.
[8] R. Rigler, Ü. Mets, J. Widengren, and P. Kask, European Biophysics Journal **22**, 169 (1993).
[9] B. Zhang, J. Zerubia, and J.-C. Olivo-Marin, Applied Optics **46**, 1819 (2007).
[10] K. M. Berland, P. So, and E. Gratton, Biophysical Journal **68**, 694 (1995).
[11] T. Dertinger, V. Pacheco, I. von der Hocht, R. Hartmann, I. Gregor, and J. Enderlein, ChemPhysChem **8**, 433 (2007).
[12] A. E. Siegman, "Lasers (mill valley, ca," (1986).
[13] H. Blom and G. Björk, Applied optics **48**, 6050 (2009).

[14] M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light* (Elsevier, 2013).
[15] S. F. Gibson and F. Lanni, JOSA A **9**, 154 (1992).
[16] R. Rigler and E. S. Elson, *Fluorescence correlation spectroscopy: theory and applications*, Vol. 65 (Springer Science & Business Media, 2012).
[17] J. Paisley and M. I. Jordan, arXiv preprint arXiv:1604.00685 (2016).
[18] T. Broderick, M. I. Jordan, J. Pitman, *et al.*, Bayesian analysis **7**, 439 (2012).
[19] J. Paisley and L. Carin, in *Proceedings of the 26th Annual International Conference on Machine Learning* (ACM, 2009) pp. 777–784.
[20] L. Al Labadi and M. Zarepour, Sankhya A , 1.
[21] C. Robert and G. Casella, *Introducing Monte Carlo Methods with R* (Springer Science & Business Media, 2009).
[22] 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).
[23] U. Von Toussaint, Reviews of Modern Physics **83**, 943 (2011).
[24] H. Liu and H. Motoda, *Computational methods of feature selection* (CRC Press, 2007).
[25] C. M. Bishop, *Pattern recognition and machine learning* (springer, 2006).
[26] O. Cappé, E. Moulines, and T. Rydén, in *Proceedings of EUSFLAT Conference* (2009) pp. 14–16.
[27] S. L. Scott, Journal of the American Statistical Association **97**, 337 (2002).
[28] F. J. Anscombe, Biometrika **35**, 246 (1948).
[29] M. Briers, A. Doucet, and S. Maskell, Annals of the Institute of Statistical Mathematics **62**, 61 (2010).
[30] M. Y. Byron, K. V. Shenoy, and M. Sahani, UR L: http://www-npl. stanford. edu/˜ byronyu/papers/derive_eks. pdf **19**, 25 (2004).
[31] 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.
[32] H. M. Menegaz, J. Y. Ishihara, G. A. Borges, and A. N. Vargas, IEEE Transactions on automatic control **60**, 2583 (2015).
[33] J. Stoer and R. Bulirsch, *Introduction to numerical analysis*, Vol. 12 (Springer Science & Business Media, 2013).
[34] F. Heiss and V. Winschel, journal of Econometrics **144**, 62 (2008).