Enumerating High Numbers of Fluorophores from Photobleaching Experiments: a Bayesian Nonparametrics Approach

J. Shepard Bryan IV[1], Ioannis Sgouralis[1,2], & Steve Pressé[1,3]

[1]Center for Biological Physics, Department of Physics, Arizona State University and
[2]School of Mathematics, University of Tennessee
[3]School of Molecular Sciences, Arizona State University.

September 24, 2020

Abstract

Determining the number of biomolecules within a small, diffraction-limited spot, is key to understanding intermolecular interactions that form the basis of all of information processing relevant to life. To enumerate the number of biomolecules within such a small region of interest (ROI), it is possible to illuminate an ROI containing a collection of fluorescently labeled biomolecules and observe the ROI's brightness over time. As most fluorescent labels (fluorophores) are initially in a fluorescently active state, the brightness diminishes in a step-like pattern as fluorophores photobleach one at a time. Naively, by counting steps, one could determine the number of fluorophores present. However, this analysis is complicated by photon shot noise, camera noise that amplifies intrinsic photon shot noise and fluorophore photophysics (i.e., states with variable emission visited by fluorophores). Current state of the art can typically enumerate as many as 20 fluorophores reliably. Yet in many biophysics problems, fluorophore numbers can vastly exceed 20 within an ROI. For this reason, we develop a method to learn the number of fluorophores alongside other parameters required to achieve high counts simultaneously and self-consistently (e.g., fluorophore photophysical rate parameters and camera parameters). As the number of fluorophores is initially unknown and may be very large, and as not all fluorophores may initially be emitting photons, we cannot rely on a traditional (parametric) Bayesian framework. As such, to achieve our goal of exceeding the state of the art by a factor of about 5, we must abandon the parametric Bayesian paradigm and invoke novel tools within Bayesian nonparametrics never previously used in step counting by photobleaching.

Labeling biomolecules such as proteins with fluorophores (i.e., fluorescent molecules) allows us to create contrast between a cell's background and labeled protein species of interest [1, 2, 3]. In particular, bright spots within cells are interpreted as aggregates of labeled biomolecules [4]. Qualitatively, brighter spots suggest more fluorophores and dimmer spots suggest fewer fluorophores; thus, stoichiometry can be deduced from brightness in a region of interest (ROI) [5, 6, 1, 7, 8, 3, 2]. However, quantitatively enumerating the number of fluorophores from brightness measurements is difficult due to shot noise, camera or detector noise and photophysics of the individual fluorophores [9]. What is more, trying to enumerate fluorophores by discriminating between them on the basis of their physical location does not work as protein aggregates are typically at least an order of magnitude smaller than the diffraction limit of light (≈ 10 nm vs ≈ 250 nm) [2, 3].

A number of methods exist to enumerate fluorophores within an ROI [10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. In the absence of photobleaching altogether, and by simply comparing background to the brightness of an ROI, one can use ruler methods that approximate the number of fluorophores by relating the maximum brightness of the trace to some reference with known brightness. This approach is complicated
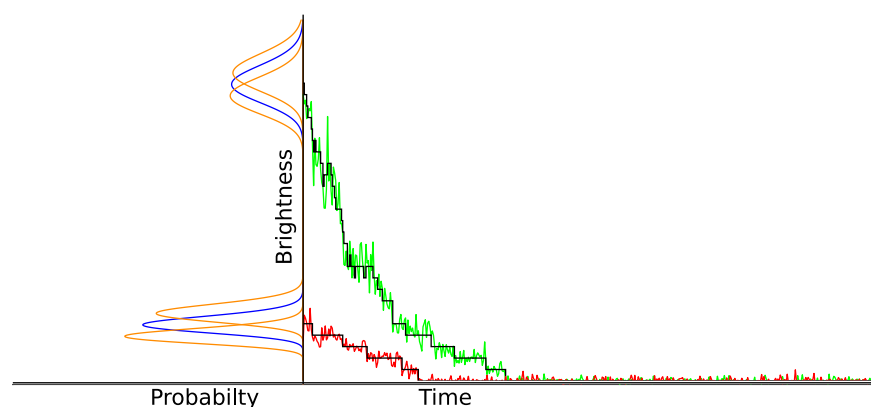
Figure 1: **As the number of active fluorophores increases, photon shot noise amplified by camera noise becomes an essential modeling component.** On the right we see two brightness time traces for an ROI. The ROI giving rise to the red trace contains 5 fluorophores and the ROI giving rise to the green trace contains 26 fluorophores. The black line through each trace represents the expected mean brightness at each time level based on the background brightness, fluorophore brightness, and number of active fluorophores at the time level. On the left we see the probability distribution for the initial brightness of each ROI (blue) as well as the probability distribution for the initial brightness if the ROI contained one more or fewer fluorophores (orange). We notice that: 1) for the red trace, the measured brightness matches closely with the mean, whereas the green trace fluctuates wildly; 2) the probability for the initial brightness is sharply peaked for the red trace which allows us to easily tell how many fluorophores are initially active. By contrast, the probability for the initial brightness is much more spread out for the green trace and overlaps greatly with the distributions for one more and one fewer fluorophores. Together, these two points suggest that while simple theoretical treatments may work for enumerating steps involving few fluorophores, large numbers of fluorophores require us to move beyond the traditional analyses and account for photophysics, camera noise and shot noise.
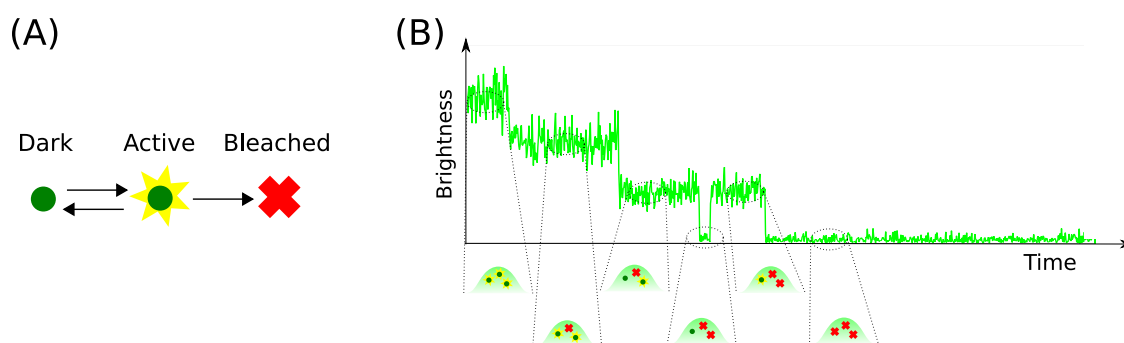

Figure 2: **Summary of the problem.** (A) Each fluorophore can exist in one of three states: dark, active (bright), or photobleached. Only fluorophores in the active state emit photons. There could, in principle, be multiple active and dark states. Each fluorophore transitions between these states according to the arrows (we do not draw self-transitions). (C) The brightness over time of an ROI reflects the states of the fluorophores within this ROI.
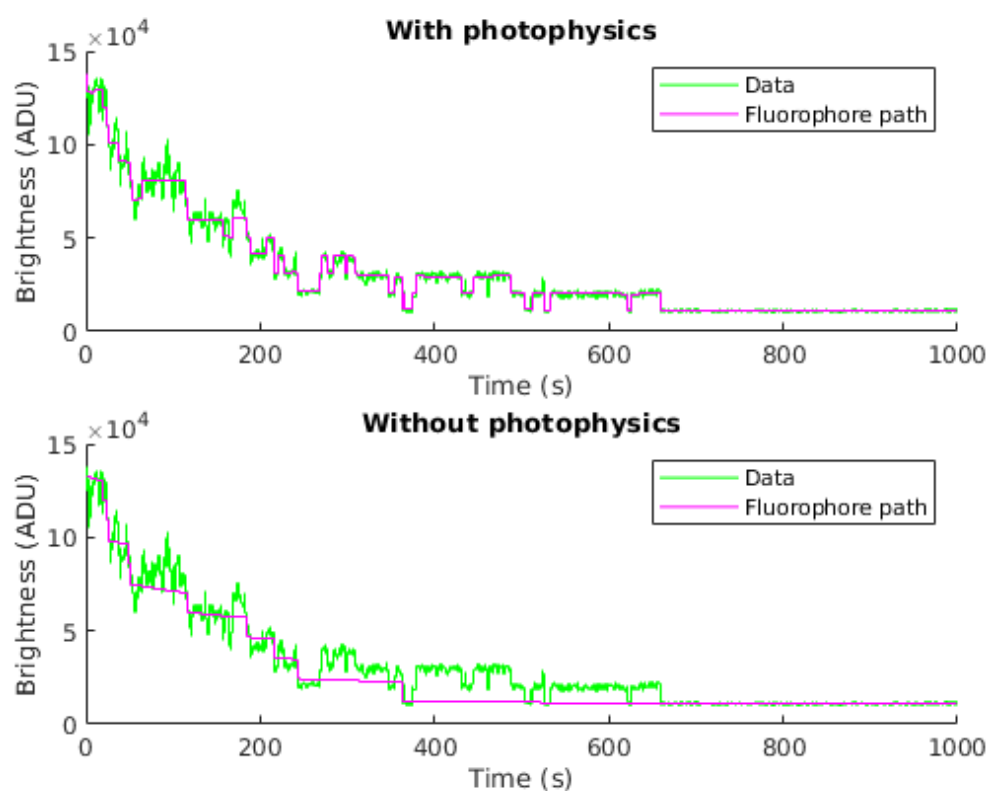
Figure 3: **Photophysics is necessary for accurate fluorophore counting.** Here we show a brightness trace of an ROI (green) overlayed with steps found using two photophysical models (purple). In the top we see a trace in which fluorophores are allowed to transition to a dark state and two bright states. In the bottom we see a trace in which fluorophores can only be bright or photobleached. Both traces assume the same fluorophore brightness and detector gain.

when the number of fluorophores is large because the uncertainty in the brightness arising from photon shot noise increases with the number of fluorophores [10, 11] (See figure 1).

Another method by which fluorophores are enumerated is by "Photobleaching step analysis" (PBSA). Briefly, PBSA works by continuously illuminating an ROI with a laser source. Upon illumination, most fluorophores will start in their bright state, giving rise to an initially high brightness. Subsequently, one-by-one, fluorophores will blink and photobleach, causing step-like transitions between brightness levels until all fluorophores are photobleached [5, 12, 2, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34]. A cartoon depicting this process is laid out in figure 2. Algorithms for PBSA have been developed by many groups [10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. These methods exploit, for example, hidden Markov models [20, 22], data filtering in order to see steps more clearly [14], statistical measures used to identify violations of statistics where steps are located [18, 15, 17, 16], or neural nets to count fluorophores [19].

In a recently submitted manuscript additional ways of automating and improving upon PBSA methods listed above have been explored [35]. Such methods are computationally cheap and effective at counting fluorophores however, just as in the limited to roughly 20 fluorophores. In this manuscript our goal is to exceed this limit by a factor of 5. This is motivated by work in which the number of fluorophores in a diffraction limited spot is known to vastly exceed 20 [36, 37].

In order to reach such a high limit any method will have to meet the following criteria: 1) account for accurate photophysics including photophysical artifacts such as missed steps (when two or more fluorophores photobleach simultaneously, thus making two steps appear to be one) and blinking (see figure 3); 2) utilize an accurate model for shot noise and detector noise; 3) provide full credible intervals (error bars) about the estimates as error will grow when we have a larger number of fluorophores; and 4) exploit every data point available without down-sampling, binning, or low-pass filtering which naturally reduce data quality. Existing methods can satisfy these criteria, however, they cannot do so simultaneously. Therefore they are typically limited to about 20 fluorophores depending on experimental conditions.

Traditional Bayesian methods would try to tackle the problem by treating the number of fluorophores as a random variable and relating the probability of the data given the number of fluorophores to the probability of the number of fluorophores given the data [38, 22]. When blinking is ignored this strategy is straightforward because the number of steps directly translates to the number of fluorophores. However, when blinking and multiple bright states are taken into account, the number of steps is not directly related to the number of fluorophores (a step could be a photobleaching event, a blink, two photobleaching events while one fluorophore that was previously dark becomes bright again, etc.). Thus in order to count the number of fluorophores we must be able to explain each brightness step. To explain each brightness step we must keep track of the photostates of each fluorophore at each time level. But in order to consider the photostates of each fluorophore, the model needs to already know how many fluorophores are in an ROI, which defeats the purpose of the problem. Thus the traditional Bayesian approach cannot count the number of fluorophores and learn photophysics simultaneously because it treats the number of fluorophores as a model selection problem. We must therefore use techniques from Bayesian nonparametrics to solve our problem.

To develop a method satisfying all criteria listed above, we propose to use tools within Bayesian nonparametrics, in particular the Beta-Bernoulli Process (BBP) [39, 40] which has never previously been used in PBSA. Nonparametrics are required because the number of fluorophores is unknown and the BBP allows the number of parameters, and thus the dimensionality of the model, to vary in a way that is warranted by the data [41, 42]. On account of these mathematical innovations, we will demonstrate that our method can accurately and reliably infer the number of fluorophores within an ROI while satisfying all four criteria and, by doing so, we can robustly count as many as 5 times more fluorophores than the state of the art.

1 Methods

Here we both set up the forward model, i.e., the model that describes the stochastic photobleaching process and other contributions to the data, and discuss the inference strategy required to learn the number of fluorophores from the data. In particular, we will learn the state of each fluorophore at each time, the transition probabilities between fluorophore states, the probability that a fluorophore starts bright, the fluorophore mean brightness, and the background mean brightness. As we seen in fig 3, learning these parameters is required to accurately infer the number of fluorophores in each ROI. As with all methods within the Bayesian paradigm, whether parametric or nonparametric, we provide not only a point estimate for the maximum *a posteriori* (MAP) estimate for each parameter, but also achieve full posterior inference with credible intervals.

1.1 Forward model

The forward model describes how the data are generated. We start with R diffraction limited ROIs, indexed by r for $r = 1, \ldots, R$. Each ROI has M_r fluorophores indexed by $m = 1, \ldots, M_r$. We record the brightness (measured in ADUs) of each ROI for N successive time levels, indexed by $n = 1, \ldots, N$. The brightness of the ROI r, at time level n is denoted w_n^r and is conditioned on the states of the fluorophores within the ROI at that time. The state of mth fluorophore in ROI r at time level n is labeled $s_n^{m,r}$. For simplicity, let us say that at each time level $s_n^{m,r}$ can be one of three states: dark, σ_D, active, σ_A, and photobleached, σ_B. We tackle the obstacle of multiple bright states in the SI.

At the first time level, each fluorophore in each ROI, will start either active or dark with probability mass given by $\boldsymbol{\pi}_0$ ($\boldsymbol{\pi}_0$ is an array with two elements: the probability of a fluorophore starting bright, π_{0A} and the probability of a fluorophore starting dark, π_{0D}). For each following time level, n, the state of each fluorophore is conditioned on the previous state of the fluorophore according to $\boldsymbol{\pi}$, the transition probability matrix. Each element, π_{ij}, of the matrix represents the probability that a fluorophore will be in state σ_j given that it was previously in state σ_i (see SI). These transitions can go like "dark to dark", "dark to bright", "bright to bright", "bright to dark", "bright to bleached", or "bleached to bleached" (See figure 2 and SI). In the presence of more than one bright state, we would have additional transitions such as "first bright to second bright" though, for simplicity alone we will limit our discussion below to one bright state.

This scheme is mapped here,

$$s_1^{m,r} \sim \textbf{Categorical}\left(\boldsymbol{\pi}_0\right) \tag{1}$$

$$s_n^{m,r} | s_{n-1}^{m,r} \sim \textbf{Categorical}\left(\boldsymbol{\pi}_{s_{n-1}^{m,r}}\right) \tag{2}$$

where $s_n^{m,r}$ is the state of fluorophore m in ROI r at time level n, $\sim$ means "is sampled from", $|$ means "given" or "conditioned on", $\textbf{Categorical}\left(x\right)$ means "the categorical distribution with probability mass x", and $\boldsymbol{\pi}_{s_{n-1}^{m,r}}$ means "the row of $\boldsymbol{\pi}$ corresponding the the state of $s_{n-1}^{m,r}$. The support for these categorical distributions is understood to be the set of possible states of the fluorophores, $\{\sigma_D, \sigma_A, \sigma_B\}$ meaning that for all n, m, and r, $s_n^{m,r} = \sigma_D$, $s_n^{m,r} = \sigma_A$, or $s_n^{m,r} = \sigma_B$.

At each time level, the fluorophores in each ROI give rise to the mean brightness of the ROI at the time level, μ_n^r. The mean brightness is the time step times by mean photons emitted per unit time for the time level, meaning it has units of photons per frame. We can decompose it into the sum of the mean background brightness of the ROI, μ_B^r, and the mean fluorophore brightness, μ_A, times the number of fluorophores in the ROI that are active,

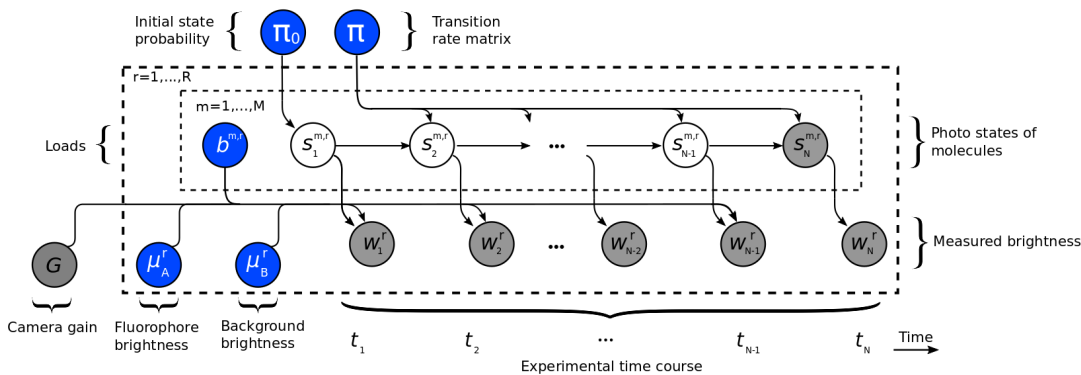$$\mu_n^r = \mu_B^r + \sum_{m=1}^{M_r} \mu_{s_n^{m,r}} \tag{3}$$

Figure 4: **The graphical model of the inverse model.** Nodes (circles) in the graphical model represent random variables. Blue nodes are the random variables we infer. White nodes are auxiliary random variables that "come along for the ride". Grey nodes are observations (data). The arrow between nodes indicates conditional dependence, meaning that if x is conditioned on y then we would draw an arrow going from node y to node x. The plates (dashed boxes) indicate that the random variables inside the plate repeat over the index in the top left of the plate. For example, the μ_B^r node is within the outer plate with index r meaning that be have one μ_B^r for each r.

where $\mu_{s_n^{k,r}}$ means "the brightness of the state corresponding to $s_n^{k,r}$", as in, if $s_n^{k,r} = \sigma_A$ then $\mu_{s_n^{k,r}} = \mu_A$ or if $s_n^{k,r} = \sigma_B$ then $\mu_{s_n^{k,r}} = 0$. We note that $\sum_{m=1}^{M_r} \mu_{s_n^{m,r}}$ simply counts how many fluorophores are in the active state in the ROI at the time level of interest. Here for simplicity we assume that background brightness μ_B^r remains constant in time. The time variant μ_B^r is treated in the SI.

For data obtained with an EMCCD camera the brightness that we measure, w_n^r, is conditioned on the mean brightness and the gain, G, through a gamma distribution [43]

$$w_n^r | \mu_n^r, G \sim \mathbf{Gamma}\left(\mu_n^r, G\right). \tag{4}$$

We notice that this conditional takes into account both the shot noise and the readout noise [44]. Substituting Eq. (3) into Eq. (4) we find

$$w_n^r | s_n^{1:M_r, r}, \mu_A, \mu_B^r, G \sim \mathbf{Gamma}\left(\mu_B^r + \sum_{k=1}^{K} \mu_{s_n^{k,r}}, G\right). \tag{5}$$

With this model, the mean expected readout is $\mu_n^r G$ ADU with a standard deviation of $\sqrt{\mu_n^r G^2}$ ADU. Thus the noise scales with the brightness.

This scheme, where many fluorophores give rise to a single measurement (the brightness at a time level), takes the form of a factorial hidden Markov model [45, 38].

1.2 Inverse model

The inverse model is the model we use to learn parameters from the data. We know the number of ROI, R, as well as the length of the time traces, N. Following the Bayesian paradigm, we place prior distributions on all

parameters whose posterior distribution we wish to determine from the data. The priors on the initial state probabilities and the priors on the rows of the transition probability matrix are Dirichlet distributions [38, 46, 47, 48]

$$\boldsymbol{\pi}_0 \sim \textbf{Dirichlet}\left(\alpha_0\right) \tag{6}$$

$$\boldsymbol{\pi}_{\sigma_i} \sim \textbf{Dirichlet}\left(\alpha_{\sigma_i}\right) \qquad\qquad (i = D, A, B) \tag{7}$$

where α_0 and α_{σ_i} are hyperparameters chosen to be as naive as possible (See SI). The Dirichlet distributions here are chosen for computational convenience alone as these are conjugate to the categorical distribution (Eq. (2)) meaning that they are deliberately chosen in such a way that the product of likelihood and prior lead to a distribution from which we can directly sample new values (See SI). The choice of prior becomes immaterial in the limit of a large amount of data.

The priors on the mean background brightness and mean fluorophore brightness are gamma distributions

$$\mu_A \sim \textbf{Gamma}\left(\phi_{\mu_A}, \psi_{\mu_A}\right) \tag{8}$$

$$\mu_B^r \sim \textbf{Gamma}\left(\phi_{\mu_B}, \psi_{\mu_B}\right) \tag{9}$$

The gamma distribution choice reflects our knowledge that these variables are positive, and that we have little knowledge of the range of values they take. By choosing small shape parameters ($\phi_{\mu_A}, \phi_{\mu_B} = 2$) the prior gives non-negligible probability to a wide of values range for μ_A and μ_B^r (choice of scale parameters $\psi_{\mu_A}, \psi_{\mu_B}$ is discussed in SI). Note that the parameters within the priors (ϕ_{μ_A}, α_0, etc.) are called hyperparameters and become decreasingly influential to the posterior as more data are collected [41, 49].

Because we cannot set a prior on the number of fluorophores in each ROI, M_r, we must use Bayesian nonparametrics in our analysis. Briefly, we will implement this using a nonparametric limit [50, 51]. That is, we will assume an exceedingly large number of model fluorophores in the ROI, $K \gg M_r$, indexing each fluorophore with $k = 1, \ldots, K$. We will then assign each model fluorophore a load $b^{k,r}$. If the load is on, $b^{k,r} = 1$, then we say that the fluorophore exists and contributes to the brightness of the ROI. If the load is off, $b^{k,r} = 0$, then the fluorophore is a virtual fluorophore which does not contribute to the brightness. Thus by adding up the loads of the model fluorophores we get the number of fluorophores that are located within the ROI. A load, $b^{k,r}$, is a random variable sampled from the Bernoulli distribution with hyperparameter γ

$$b^{m,r} \sim \textbf{Bernoulli}\left(\frac{\gamma}{K + \gamma - 1}\right). \tag{10}$$

This seemingly odd choice of probability mass is motivated by the Beta-Bernoulli Process [39, 40] (See SI). The result is that as K becomes large, formally at the limit $K \to \infty$, the probability distribution converges to a distribution in which an infinite number of fluorophores are considered [50, 51]. Furthermore, this choice of prior allows inference that is independent on our choice of loads. That is, we can choose K to be anything (so long as K is greater than the number of fluorophores in the ROI, see section 2.2) and it will not interfere with our counting.

In this scheme, the states are sampled the same way as in the forward model, but the brightness is now conditioned on the loads

$$s_1^{k,r} | \boldsymbol{\pi}_0 \sim \textbf{Categorical}\left(\boldsymbol{\pi}_0\right) \tag{11}$$

$$s_n^{k,r} | s_{n-1}^{k,r}, \boldsymbol{\pi} \sim \textbf{Categorical}\left(\boldsymbol{\pi}_{s_{n-1}^{k,r}}\right) \tag{12}$$

$$w_n^r | s_n^{1:K,r}, b^{1:K,r}, \mu_A, \mu_B^r, G \sim \textbf{Gamma}\left(\mu_B^r + \sum_{k=1}^{K} b^{k,r} \mu_{s_n^{k,r}}, G\right). \tag{13}$$

We note that $\sum_{k=1}^{K} b^{k,r} \mu_{s_n^{k,r}}$ counts how many fluorophores simultaneously have "on" loads and are in the active state in the ROI at the time level of interest. Furthermore, while this inverse model looks different from the forward model (because of the loads), the models are actually the same if we declare that $b^{m,r} = 1$ for all fluorophores in the forward model. However, for notational simplicity, we did not include loads in the forward model.

Lastly, in the PBSA experiments we consider, data is collected until all fluorophores are photobleached. As such, we have knowledge of the final states of the fluorophores. Put differently, the fluorophore states at the last time level are fixed at

$$s_N^{k,r} = \sigma_B. \tag{14}$$

1.3   Model summary

The model is summarized as follows.

$$\mu_A \sim \mathbf{Gamma}\left(\phi_{\mu_A}, \psi_{\mu_A}\right) \tag{15}$$

$$\mu_B^r \sim \mathbf{Gamma}\left(\phi_{\mu_B}, \psi_{\mu_B}\right) \tag{16}$$

$$\boldsymbol{\pi}_0 \sim \mathbf{Dirichlet}\left(\alpha_0\right) \tag{17}$$

$$\boldsymbol{\pi}_{\sigma_i} \sim \mathbf{Dirichlet}\left(\alpha_{\sigma_i}\right) \tag{18}$$

$$b^{m,r} \sim \mathbf{Bernoulli}\left(\frac{\gamma}{\gamma + K - 1}\right) \tag{19}$$

$$s_1^{k,r} \sim \mathbf{Categorical}\left(\boldsymbol{\pi}_0\right) \tag{20}$$

$$s_n^{k,r}|s_{n-1}^{k,r} \sim \mathbf{Categorical}\left(\boldsymbol{\pi}_{s_{n-1}^{k,r}}\right) \tag{21}$$

$$s_N^{k,r} = \sigma_B \tag{22}$$

$$w_n^r|s_n^{1:M_r,r}, \mu_A, \mu_B^r, G \sim \mathbf{Gamma}\left(\mu_B^r + \sum_{k=1}^{K} b^{k,r} \mu_{s_n^{k,r}}, G\right) \tag{23}$$

Note that in the following sections we include multiple bright states and exponentially decaying background in that analysis. This is treated in the SI, but for simplicity in notation is not shown here.

1.4   Sampling

The inverse model scheme (Eqs. (15)-(23)) allow us to develop the joint posterior of our model

$$\mathcal{P}\left(s_{1:N}^{1:K,1:R}, b^{1:K,1:R}, \mu_A, \mu_B^{1:R}, \boldsymbol{\pi}, \boldsymbol{\pi}_0|w_{1:N}^{1:R}\right) \tag{24}$$

$$\propto \mathcal{P}\left(w_{1:N}^{1:R}|s_{1:N}^{1:K,1:R}, b^{1:K,1:R}, \mu_A, \mu_B^{1:R}, \boldsymbol{\pi}, \boldsymbol{\pi}_0\right)$$

$$\times \mathcal{P}\left(s_{1:N}^{1:K,1:R}, b^{1:K,1:R}, \mu_A, \mu_B^{1:R}, \boldsymbol{\pi}, \boldsymbol{\pi}_0\right)$$

where the likelihood (middle term of Eq. (24)) is the product of Eq. (25) for all time levels and ROIs, which looks like

$$\mathcal{P}\left(w_{1:N}^{1:R}|s_{1:N}^{1:K,1:R}, b^{1:K,1:R}, \mu_A, \mu_B^{1:R}, \boldsymbol{\pi}, \boldsymbol{\pi}_0\right) \tag{25}$$

$$= \prod_{r=1}^{R}\prod_{n=1}^{N} \mathbf{Gamma}\left(w_n^r|\mu_B^{r,n} + \sum_{k=1}^{K} b^{k,r} \mu_{s_n^{k,r}}\right),$$

and the prior (final term) of can be further decomposed into

$$
\mathcal{P}\left(s_{1:N}^{1:K,1:R}, b^{1:K,1:R}, \mu_A, \mu_B^{1:R}, \boldsymbol{\pi}, \boldsymbol{\pi}_0\right) \tag{26}
$$
$$
\propto \mathcal{P}\left(b^{1:K,1:R}\right)\mathcal{P}\left(\mu_A\right)\mathcal{P}\left(\mu_B^{1:R}\right)\mathcal{P}\left(\boldsymbol{\pi}\right)\mathcal{P}\left(\boldsymbol{\pi}_0\right)
$$
$$
\times \mathcal{P}\left(s_{1:N-1}^{1:K,1:R}|\boldsymbol{\pi}, \boldsymbol{\pi}_0\right)
$$

which has two main parts, 1) the middle terms of Eq. (26) are the product of Eqs. (6)-(10) over all indexes 2) the last term of Eq. (26) is the product of $\pi_{0,D}$ to the power of the number of fluorophores that start dark (C_{0D} is the number of fluorophores that start dark) times $\pi_{0,A}$ to the power of the number of fluorophores that start bright (C_{0A} is the number of fluorophores that start bright) times π_{ij} to the power of the number of transitions that occurred from σ_i to σ_j (C_{ij} is the number of transitions from σ_i to σ_j)

$$
\mathcal{P}\left(s_{1:N-1}^{1:K,1:R}|\boldsymbol{\pi}, \boldsymbol{\pi}_0\right) = \pi_{O,D}^{C_{0D}}\pi_{0,A}^{C_{0A}}\prod_{i,j}\pi_{ij}^{C_{ij}} \tag{27}
$$

$$
C_{0D} = \sum_{r=1}^{R}\sum_{k=1}^{K}I[s_1^{k,r} = \sigma_D] \tag{28}
$$

$$
C_{0A} = \sum_{r=1}^{R}\sum_{k=1}^{K}I[s_1^{k,r} = \sigma_A] \tag{29}
$$

$$
C_{ij} = \sum_{r=1}^{R}\sum_{k=1}^{K}\sum_{n=2}^{N}I[s_n^{k,r} = \sigma_j]I[s_{n-1}^{k,r} = \sigma_i] \tag{30}
$$

where $I[x = y]$ is the indicator function that is equal to 1 if $x = y$ and 0 else. We note that in this scheme (Eqs. (24)-(27)), we are learning the number of fluorophores in many ROI simultaneously. This is important because by analyzing many ROI together, we effectively have more information to train on and therefore obtain more accurate results.

The posterior has no analytical solutions for the number of fluorophores in the ROI. Therefore we use Gibbs sampling to learn these the parameters of the model [52, 41, 53], $s_{1:N}^{1:K,1:R}, b^{1:K,1:R}, \mu_A, \mu_B^{1:R}, \boldsymbol{\pi}$, and $\boldsymbol{\pi}_0$. Briefly, Gibbs sampling starts with an initial set of values for the parameters that we wish to learn and then iteratively, one by one, samples new values for each parameter while holding the others constant. This is outlined below where (i) indexes the iteration in the Gibbs sampler.

- Step 1: Choose initial $s_{1:N}^{1:K,1:R,(0)}$, $b^{1:K,1:R,(0)}$, $\mu_A^{(0)}$, $\mu_B^{1:R,(0)}$, $\boldsymbol{\pi}^{(0)}$, and $\boldsymbol{\pi}_0^{(0)}$.

- Step 2: For many iterations, i:

  - A) Sample new states and loads from

    $$
    \mathcal{P}\left(s_{1:N}^{1:K,1:R,(i)}, b^{1:K,1:R,(i)}|\mu_A^{(i-1)}, \mu_B^{1:R,(i-1)}, \boldsymbol{\pi}^{(i-1)}, \boldsymbol{\pi}_0^{(i-1)}, w_{1:N}^{1:R}\right)
    $$

    which is the product of Eq. (25) and Eq. (27).

  - B) Sample new transition probabilities and initial state probabilities from

    $$
    \mathcal{P}\left(\boldsymbol{\pi}^{(i)}, \boldsymbol{\pi}_0^{(i)}|s_{1:N}^{1:K,1:R,(i)}, b^{1:K,1:R,(i)}, \mu_A^{(i-1)}, \mu_B^{1:R,(i-1)}, w_{1:N}^{1:R}\right)
    $$

    which is the product of Eqs. (6), (7), and (27). Here the choice of conjugate priors allow us to sample $\boldsymbol{\pi}_0$ and $\boldsymbol{\pi}$ directly (See SI).

   – C) Sample new camera parameters from

$$\mathcal{P}\left(\mu_A^{(i)}, \mu_B^{1:R,(i)} | s_{1:N}^{1:K,1:R,(i)}, b^{1:K,1:R,(i)}, \boldsymbol{\pi}^{(i)}, \boldsymbol{\pi}_0^{(i)}, w_{1:N}^{1:R}\right)$$

which is the product of Eq. (25) and the priors (Eqs. (8)-(9)).

For Step 1, the initial values are chosen to be the means of the prior. These means can in principle be anything, although, better guesses lead to faster convergence. We did notice the sampler is sensitive to the initial value for the fluorophore brightness. In other words we could sample, for example, half as many fluorophores with twice the brightness or three times the number of fluorophore with a third the brightness. To avoid this problem we set a sharp prior for the value of μ_A on the brightness of a single fluorophore which is clear at the end of the trace. This makes sure that our model is posterior is sharply peaked at reasonable numbers of fluorophores. We calibrate gain, G, using methods described in Hirsch et al [43]. See SI for how we choose the other parameters, $\mu_B^{1:R,(0)}$, $\boldsymbol{\pi}^{(0)}$, and $\boldsymbol{\pi}_0^{(0)}$.

In Step 2 A, for each ROI, we sample loads and states together by using the forward filter backwards sample algorithm [38]. To speed up computation we do this using a collapsed state space where we treat photobleached fluorophores and fluorophores with load off to be the same (See SI). For Step 2 C, we sample μ_A and $\mu_B^{1:R}$ each separately using a Metropolis-within-Gibbs step [54, 41, 53].

As the sampler runs, higher probability regions are sampled more often and lower probability regions are sampled less often [52, 41, 53]. After many iterations we can histogram the results to approximate the shape of our posterior. By looking at the histogram and not just the MAP or mean, we are able to get credible intervals for our estimates.

We note that the R ROIs in the base set are analyzed simultaneously. To be clear, we mean that each ROI has its own fluorophore states, loads, and mean background brightness, but the transition probabilities, initial state probability, and mean fluorophore brightness are global. This is advantageous because it allows the inference of each ROI to benefit from the others by providing more data on which to train the transition probabilities.

2   Results

In this section we evaluate the robustness of our method. We start by demonstrating the method on a simulated dataset generated with a base set of parameters. For the base set we have 50 ROIs containing on average 14 fluorophores. The traces are 1000 seconds long with brightness w_n^r collected every 50 ms, so 20000 total frames. The exact number of fluorophores in each ROI is sampled from a binomial distribution to mimic 20 binding sites with 70% labeling efficiency. The emGain used for the simulation was 50 [55]. The dimensionless background brightness parameters are $\mu_B^r = \mu_B^r{}' = 500$, $\tau = 2000$ (See SI for exponentially decaying background parameters). The fluorophores were simulated with two bright states with brightness given by $\mu_{A1} = 450$ and $\mu_{A2} = 350$ (plus one dark state and a photobleached state with brightness given by $\mu_D = \mu_B = 0$). The initial probability of starting bright was chosen to be .97 to reflect that most fluorophores are initially excited. For the number of loads, M, we choose 99. These parameters were chosen to mimic the parameters learned from real data which will be discussed later. We will refer to these parameters as the "base case." After the initial demonstration on the base case parameters, we change the parameters one by one to show how the results change. Then we will show our analysis on real data.

As a metric for analysis, we present our results in two ways: 1) we show the learned probability of the number of fluorophores per ROI. That is, we find $B_r = \sum_{k=1}^{K} b_k^r$ for each ROI and each sample in our Gibbs sampler, then we histogram the B_r's for $r = 1, .., R$. In the limit that the number of ROIs is large, this should converge to the ground truth distribution of fluorophore numbers convoluted with the uncertainty associated

with the number of fluorophores in a single ROI; 2) We find the error distribution of our estimate which is calculated by histogramming the difference between B_r and the ground truth number of fluorophores in ROI r.

2.1 Demonstration on base parameters

We evaluated our method with data simulated using the forward model (Eqs. (1)-(5)) using the base set of parameters. Figure 5 shows the results for our analysis. The first row shows the brightness over time for 5 ROIs, $w_{1:N}^{1:5}$. The bottom two rows show the inferred probability of the number of fluorophores in an ROI and the error distribution as explained above. For the distribution of the number of fluorophores in an ROI (middle row) we compare to a histogram of the ground truth number of fluorophores.

Overall distribution of the number of inferred fluorophores qualitatively matches with the ground truth. Error analysis shows that roughly half of the samples were equal to the ground truth. No samples were more than 2 fluorophores off. The variance on the number of fluorophores in an ROI was small ($\langle (B_r - \langle B_r \rangle)^2 \rangle =$ .35) meaning that once the sampler found a good value for B_r it did not drift far away from it.

2.2 Varying the number of loads

Before we move on, we must make sure that our choice of nonparametric limit in our beta-Bernoulli prior, Eq. (10), does not heavily the number of loads we choose. That is, we want to make sure that if 14 fluorophores are present in an ROI, then our choice of $K = 30$ or $K = 300$ from Eq. (10) is inconsequential. To address this we ran inference on the same simulated data, but with a different number of loads. The results are shown in figure 6.

The top row of figure 6 shows that if we do not provide a sufficient number of loads, then we cannot count the correct number of fluorophores. This is obvious because if the number of fluorophores is determined by how many loads are "on", then when there are not enough loads, we cannot turn on enough loads to account for all the fluorophores. For example, if we provide 9 loads, but there are 14 fluorophores, then even if all 9 loads are "on" we underestimate by 5 fluorophores. Aside from this constraint, we found that once a sufficient number of loads are chosen then the results are not dependent on the number of loads. As seen in the bottom three rows of figure 6, increasing the number of loads has negligible impact on the performance of the sampler.

Therefore, for the rest of this paper we will take for granted that we are free to set the number of loads as we please provided the value for K is large enough. A note however is in order. If we choose too many loads, then the computation becomes expensive and inefficient. As a nice balance between sufficiency and efficiency, we will choose to make the number of loads equal to two times the expected number of fluorophores. In the case that the ground truth number of fluorophores is unknown (like for real data) the best way forward would be to guess a number of loads and if it maxes out (i.e., the number of sampled fluorophores in an ROI is equal to the number of loads) then to double the number of loads and try again.

2.3 Varying the number of fluorophores

Now that we have shown that we can accurately count the number of fluorophores in an ROI, the next clear step forward is to see how high we can count. For this, we simulated data using the same base case parameters, but changed the way in which we sample the ground truth number of fluorophores. Here we simulate ten different experiments with 20, 40, ..., 200 fluorophore binding sites and 50% labeling efficiency so that the expected number of fluorophores for each ROI is 10, 20, ..., 100 respectively. In order to account for the additional fluorophores we also increase the number of loads to 198.

As seen in figure 7, the sampler does best at learning the number of fluorophores when there are few fluorophores. As the number of fluorophores increases, we start to experience a systematic underestimation
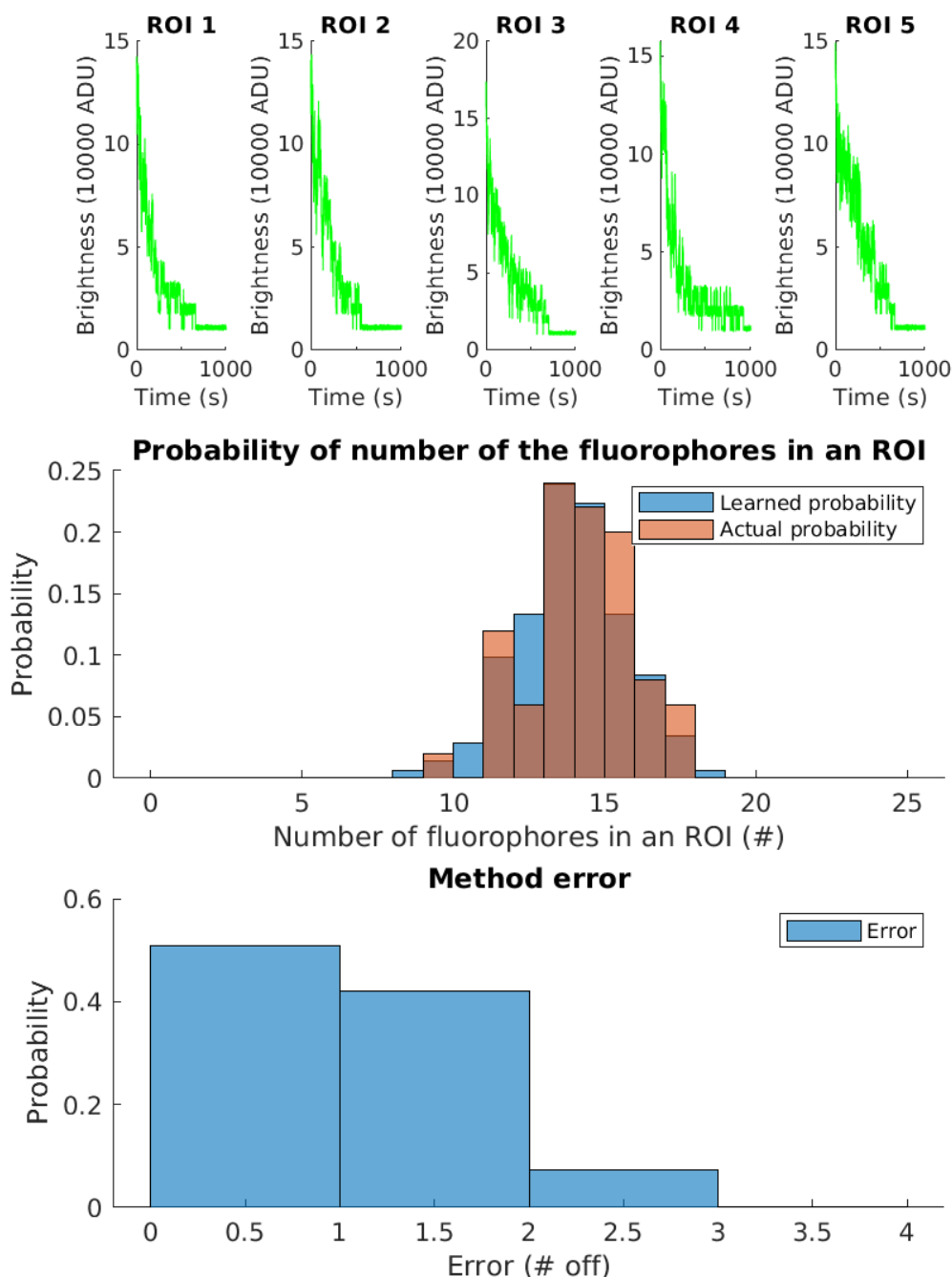
Figure 5: **Inference on the base set of parameters.** We infer the parameters for the 50 ROI in the base set simultaneously; we learn the number of fluorophores in each ROI as well as the global transition probability, gain, etc. The top row shows the input brightness vs time data for the first 5 ROIs. The second row shows the inferred probability distribution for the number of fluorophores in an ROI compared to the ground truth distribution. The third row shows the probability of the error from each sample, that is the absolute value of the ground truth number of fluorophores minus the sampled number of fluorophores for each sample and ROI.
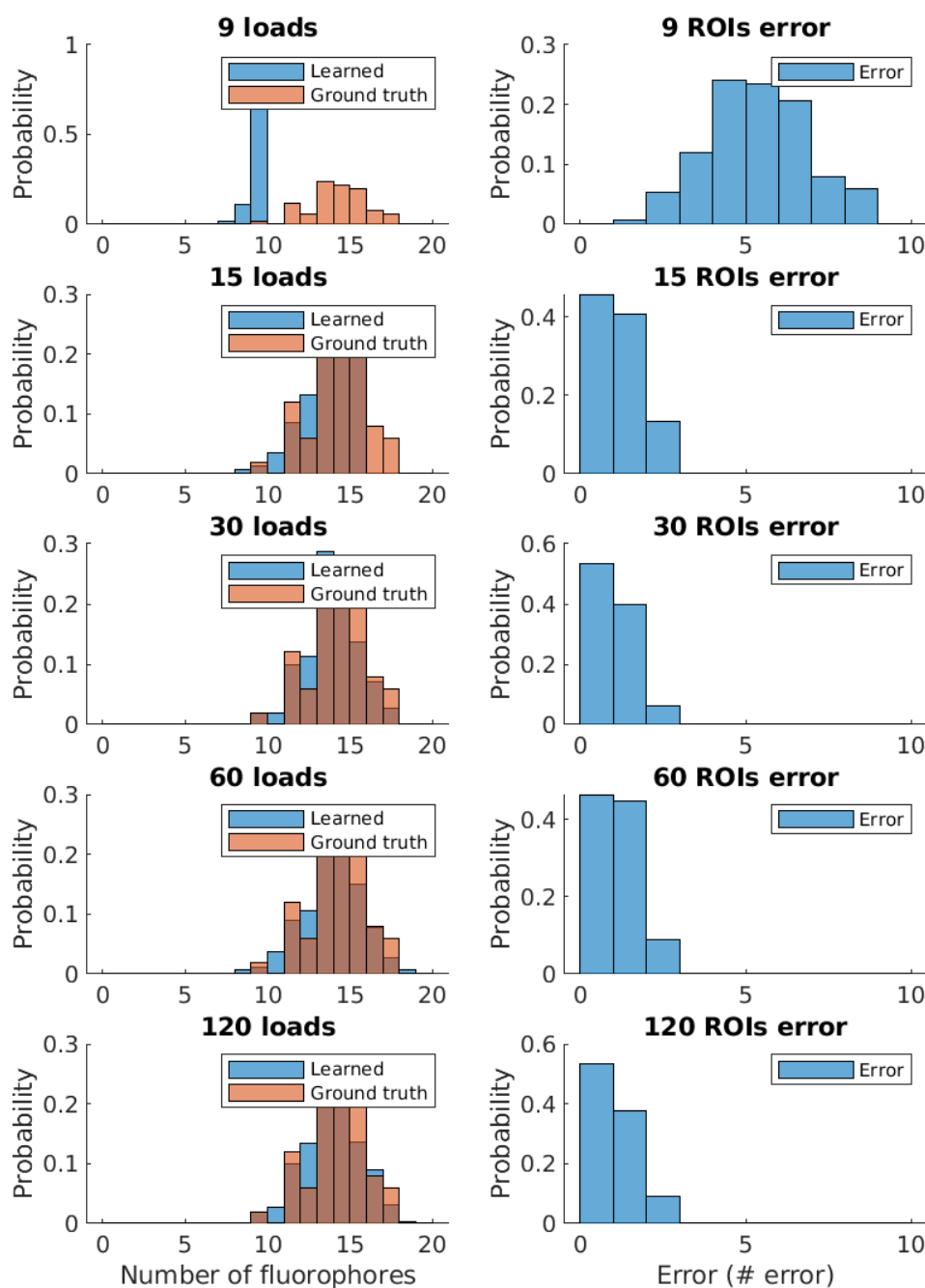
Figure 6: **Inference on fluorophore number from data simulated with varying number of loads.** Each row shows inference on data simulated with a different number of loads. The left column shows the learned number of fluorophores per ROI against the ground truth. The right column shows the probability of error for each sample.
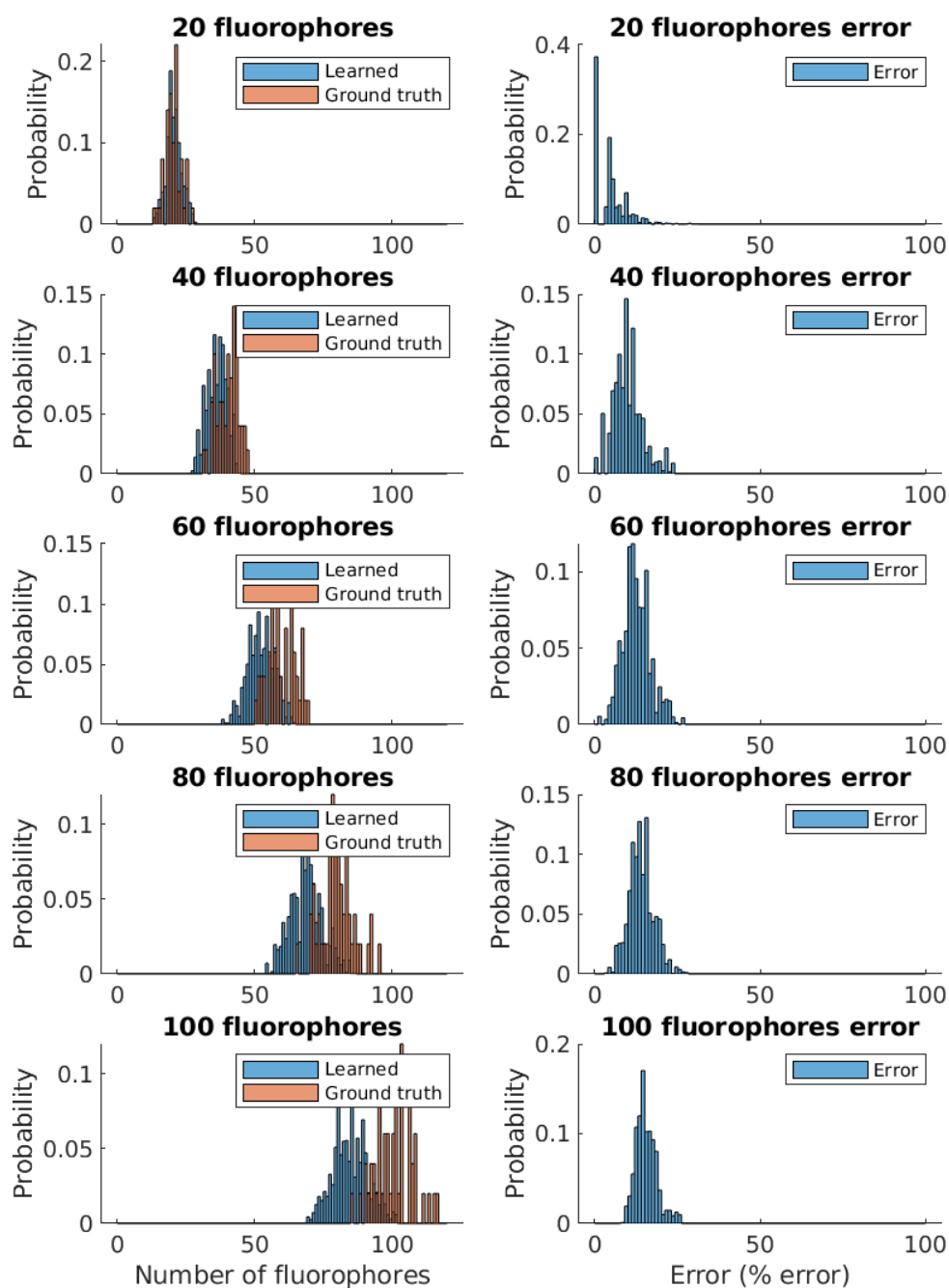
Figure 7: **Inference on fluorophore number from data simulated with varying fluorophore numbers.** Each row shows inference on data simulated with a different number of fluorophores per ROI. The left column shows the learned number of fluorophores per ROI against the ground truth. The right column shows the probability of error for each sample.

of the number of fluorophores in an ROI. In general, it seems that in the high fluorophore limit, the error for an ROI is usually between 10 and 20 percent with an average error of about 16%. The source of the underestimation is that the width of the brightness levels is much wider than the separation between brightness levels (see figure 1).

2.4 Varying the number of states

We note that while our model is nonparametric in the number of fluorophores in an ROI (i.e., that the number of fluorophores in an ROI is not a fixed quantity but inferred), our model is parametric in the number of fluorophore states. That is, we must pre-specify the number of fluorophore states. It is difficult to be both nonparametric in the number of fluorophores and the number of states (as then, trivially, each transition could be considered to visit a new state of the fluorophore). In the previous sections, we have taken for granted that we knew the correct number of fluorophore states. We modeled that the fluorophores have two separate bright states, a dark state, and a photobleached state. Here we explore the photophysical state space further using different numbers of bright states. For conciseness, we will refer to these models using a naming convention that starts with 'D' followed by the number of dark states and then 'B' followed by the number of bright states: 1) D0B1, a model with no dark state and one bright state (i.e., the fluorophore is either bright or photobleached); 2) D1B1, a model with one dark state one bright state; 3) D1B2, a model with one dark state and two bright states; 4) D1B3, a model with one dark state and three bright states. In principle we could also explore using even more bright states or even multiple dark states, but the combinatorics on the possible arrangements on fluorophore states is too large and tangential to this current discussion. We note that the base case set of parameters uses a D1B2 model. Figure 8 shows inference for data simulated using the base set of parameters, D1B2, but inferred using other models. Figure 9 shows inference for data simulated using the D0B1 model, but inferred using other models.

We notice that the model with only two states, D0B1, is the same models as the one used in Garry et al. [22]. This model does well at counting fluorophores from data generated using the D0B1 model (figure 9 second row). On the other hand, the inference severely underestimates the number of fluorophores for the data generated using the D1B1 model (figure 8 second row). This is because, lacking a dark state, no possible fluorophore state trajectory under the D0B1 model can account for blinks where the brightness of the ROI goes down and then comes back up. Thus the only way for the sampler to account for blinks is to increase the variance (by increasing the background brightness, μ_B) to a level where a sudden drop in brightness is explained purely by noise. This in turn makes photobleaching steps also the same size as noise fluctuations (because the drop in brightness from a blink is the same as a drop in brightness from a photobleaching event), thus the sampler will "miss" the photobleaching steps and most brightness drops will be explained purely by the exponentially decaying background.

The model with only one bright state and one dark state, D1B1, can still accurately count the fluorophores from data generated with the D0B1 model (figure 9 third row). In this case the dark state is never visited despite being available. On the other hand, the D1B1 model over counts the number of fluorophores because it can only use bright-dark transitions to account for changes in brightness that are really due to a fluorophore switching between bright states (figure 8 row 3). For example, if a fluorophore starts in the brighter bright state, transitions to the dimmer bright state, then transitions back to the brighter bright stat, the model, only having one bright state, will explain the dip in brightness by starting two fluorophores in the bright state and having one fluorophore stay bright while the other goes into the dark state for a short while before it returns to the bright state. Thus the D1B1 model will often use 2 fluorophores to explain a trace made by one fluorophore (model also will infer the brightness to be lower than the ground truth).

The D1B2 and D1B3 models slightly overestimate the number of fluorophores in the data generated using the D0B1 model (figure 9). This is probably due to over fitting. Meanwhile, the D1B2 model accurately learns the number of fluorophores from the data generated with the D1B2 model because this model matches
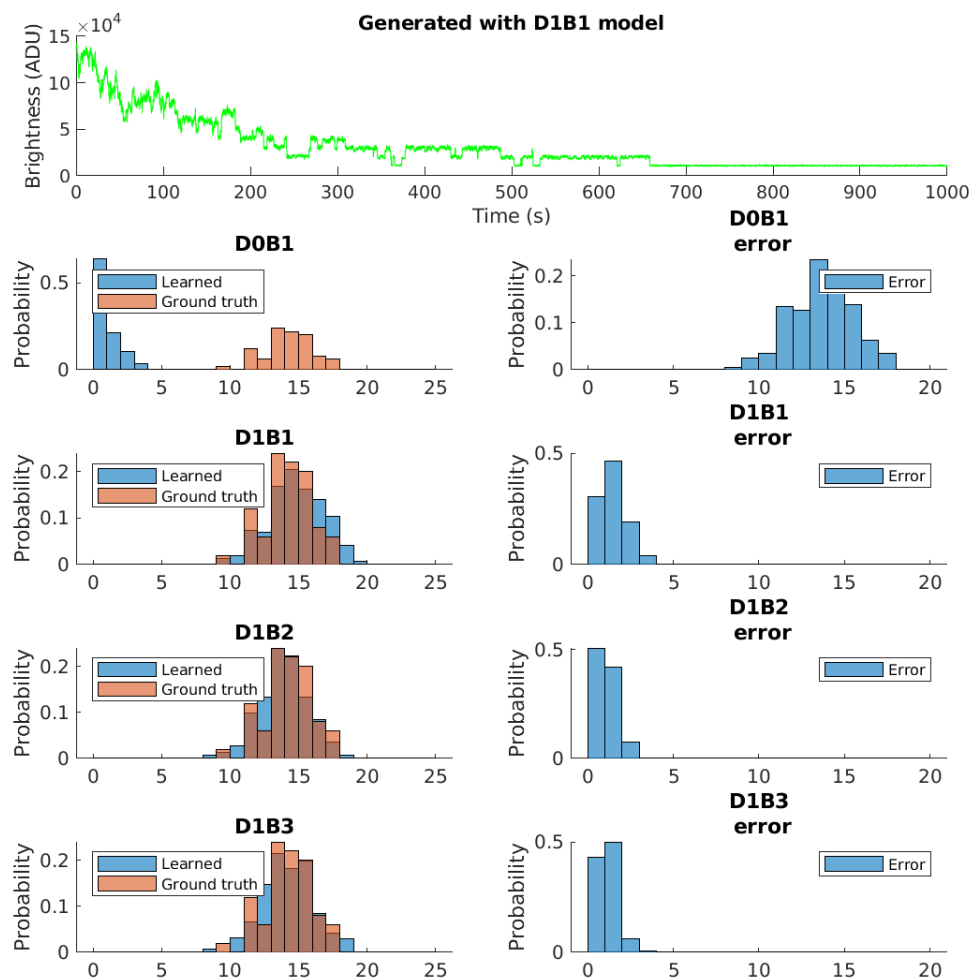
Figure 8: **Inference on fluorophore number from data simulated with varying number of states.** The top row shows an example trace from an ROI. Each following row shows inference using a specific model. The left column shows the learned number of fluorophores per ROI against the ground truth. The right column shows the probability of error for each sample.
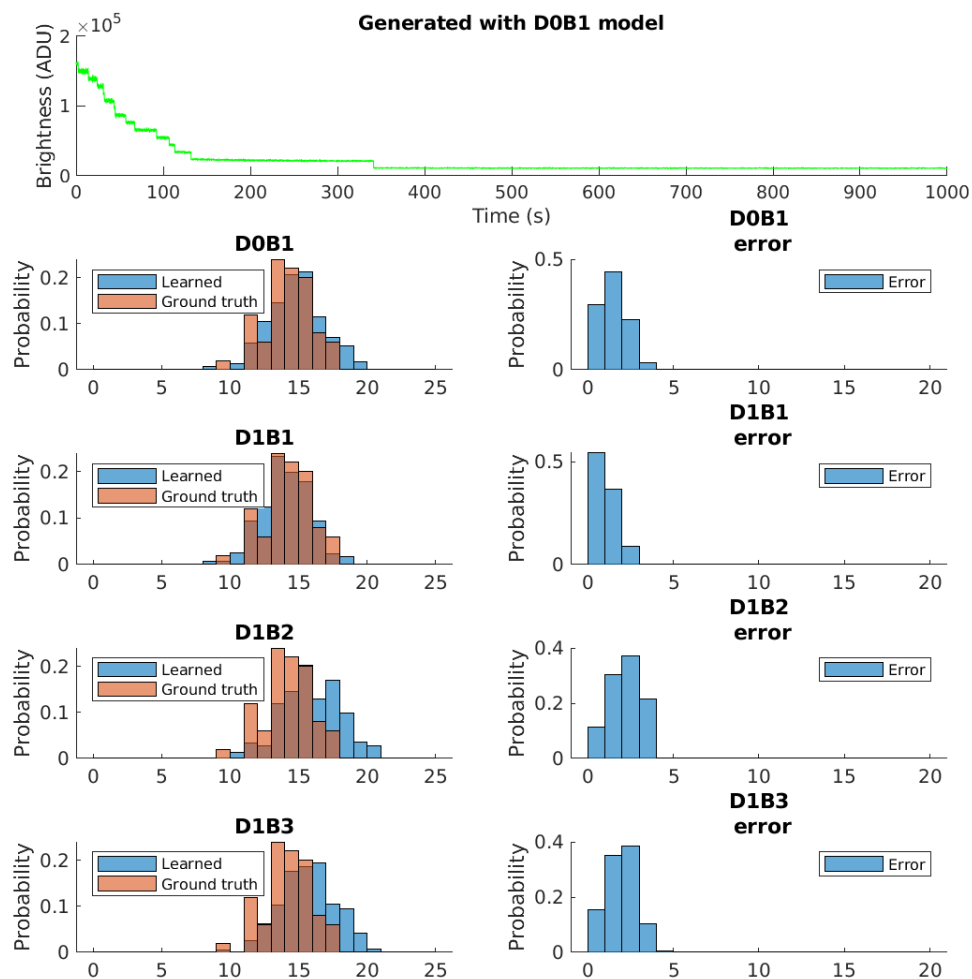
Figure 9: **Inference on fluorophore number from data simulated with varying number of states.** The top row shows an example trace from an ROI. Each following row shows inference using a specific model. The left column shows the learned number of fluorophores per ROI against the ground truth. The right column shows the probability of error for each sample.

the model used to generate the data (figure 8 row 4). Results from both the D1B2 and D1B3 models are similar for the data generated using the D1B2 model (figure 8 bottom row) because the extra state in the D1B3 model is rarely sampled.

2.5   Real data

In order to test our method on real data, we collected brightness traces from fluorophores bound to DNA origami. Data acquisition, which is already explained in Herten et al. [35], is briefly summarized here. Firstly we filmed the DNA origami. We then used a log of Gaussian filter to select ROIs. For each ROI at each time level we summed the brightness of every pixel in the ROI to get the total brightness for the ROI at each time level (i.e. the brightness time trace). We then manually filtered each trace to eliminate traces from ROIs in which free floating fluorophores had diffused through. We took time traces using two different types of of DNA origami constructed to have 20 and 35 binding sites respectively. For the 20 binding site origami, movies were taken for 1000 seconds at 50ms camera exposure (20000 frames) with an emGain of 50. For the 35 binding site origami, movies were taken for 300 seconds at 200ms camera exposure with an emGain of 10 (15000 frames). The expected distribution of fluorophores in this set up is a binomial distribution, thus we can compare the results from our counting method to a binomial distribution as a reliable ground truth.

Results are shown in figure 10. For the 20 binding site data set, the D0B1 model underestimates the number of fluorophores, and the D1B1 model overestimates the number of fluorophores, just like they did for the simulated data sets. The D1B2 and D1B3 models give similar results, but both models slightly overestimate the number of fluorophores. The overestimation could be due to accidentally selecting ROIs with two origami inside, additional sources of noise not accounted for in our model, or from excess fluorophores entering the ROI, that we were unable to filter out.

For the 35 binding site data (figure 11 bottom row), the overestimation from the D1B1 model is not as pronounced. This is because, as the number of fluorophores increases, the difference in brightness from brightness states at the beginning of the trace begins to become much smaller than the variance on the brightness (variance= $\mu_m^r G^2$ where μ_m^r is proportional to the number of active fluorophores). A larger number of fluorophores and a larger background brightness allows the variance to be large enough to smooth out the photophysical artifacts, blinks and extra bright states, without changing the number of fluorophores needed to explain the trace.

While our results for the 20 and 35 binding site data are reasonable, we wanted to test how high we could count. In order to create data sets with ROIs containing additional binding sites without running additional experiments, we decided to combine the ROIs we have to make new, larger ROIs. That is, we can draw our ROI boundary around two or more ROIs so that we have more fluorophores inside, or equivalently, at each time level we can add the brightnesses of two or more ROIs in order to create a bigger ROI with more fluorophores inside. Thus by adding together two ROIs with 20 and 35 binding sites, we can count fluorophores in ROIs with 40 and 70 total binding sites (figure 11). By adding together four ROIs with 20 and 35 binding sites, we can count fluorophores in ROI with 80 and 140 total binding sites (figure 12). Results for both these combined cases match qualitatively with the expected distributions.

3   Discussion

Learning the number of fluorophores within an ROI is a key step toward unraveling processes below the diffraction limit [5, 12, 2, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 56]. In order to go beyond the state of the art and enumerate as many as five times more fluorophores, we must accurately account for all the following: 1) fluorophore photophysics; 2) models for shot and detector noise; 3) provide full credible intervals (error bars) about the estimates; and 4) exploit every data point available without pre- or post-processing. Put differently, we must jointly learn the photophysical transition
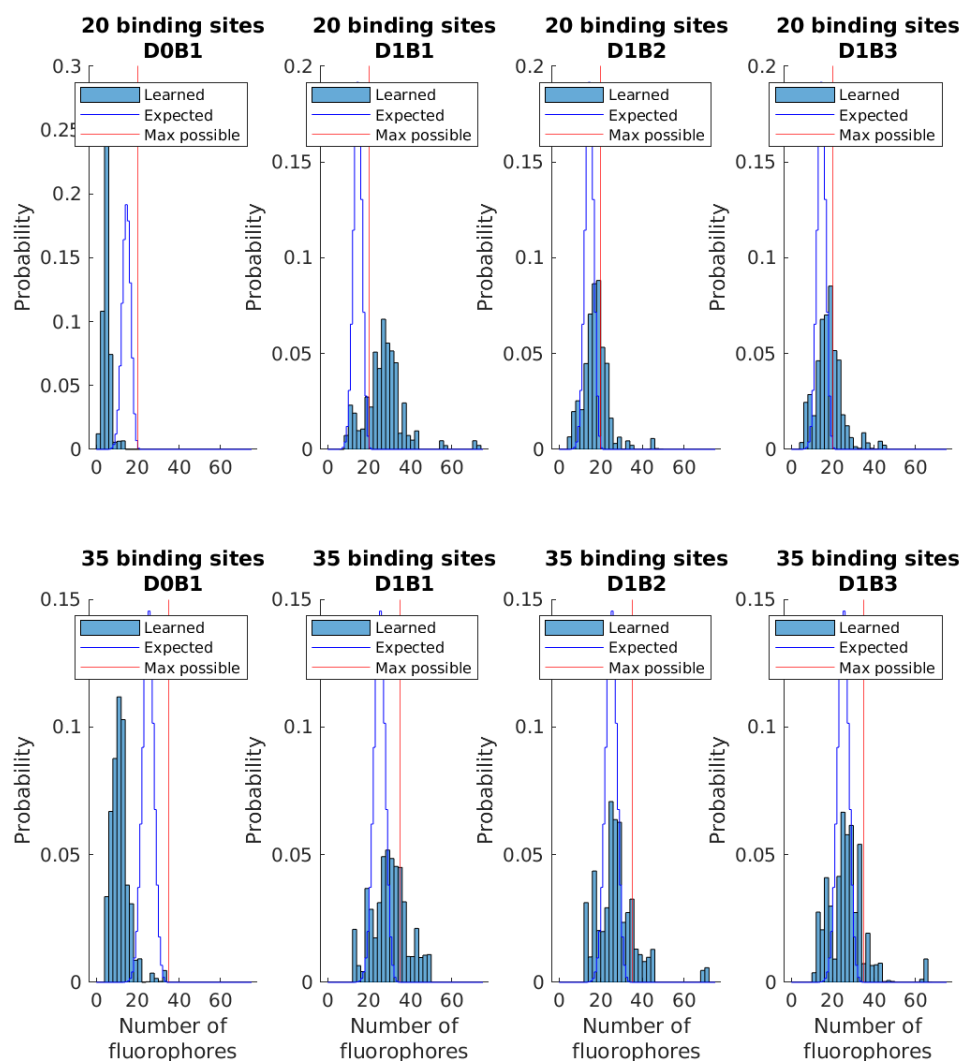
Figure 10: **Inference on real data.** Here we count the fluorophores from real data. Each column shows inference on the data using a different fluorophore model. The top row analyzes data from experiments using DNA origami with 20 binding sites. The bottom row analyzes data from experiments using DNA origami with 35 binding sites. Also plotted are the expected distribution of fluorophores (a binomial distribution), and a vertical line showing the maximum possible number of fluorophores.
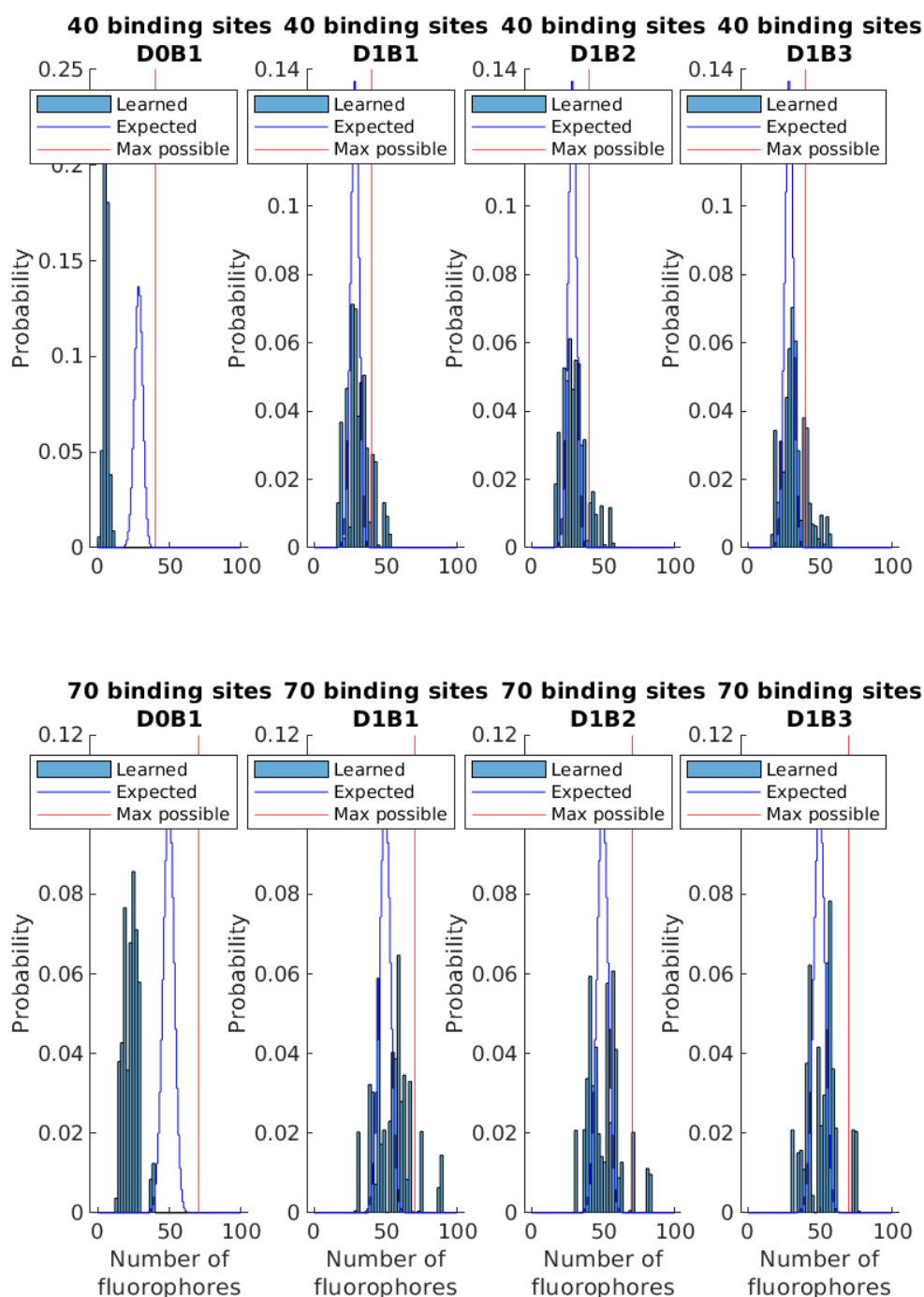
Figure 11: **Inference on semi-synthetic data.** Here we count the fluorophores from real data where we combine two ROIs into one. Each column shows inference on the data using a different fluorophore model. The top row analyzes data from experiments using DNA origami with 20 binding sites. The bottom row analyzes data from experiments using DNA origami with 35 binding sites. Also plotted are the expected distribution of fluorophores (a binomial distribution), and a vertical line showing the maximum possible number of fluorophores.
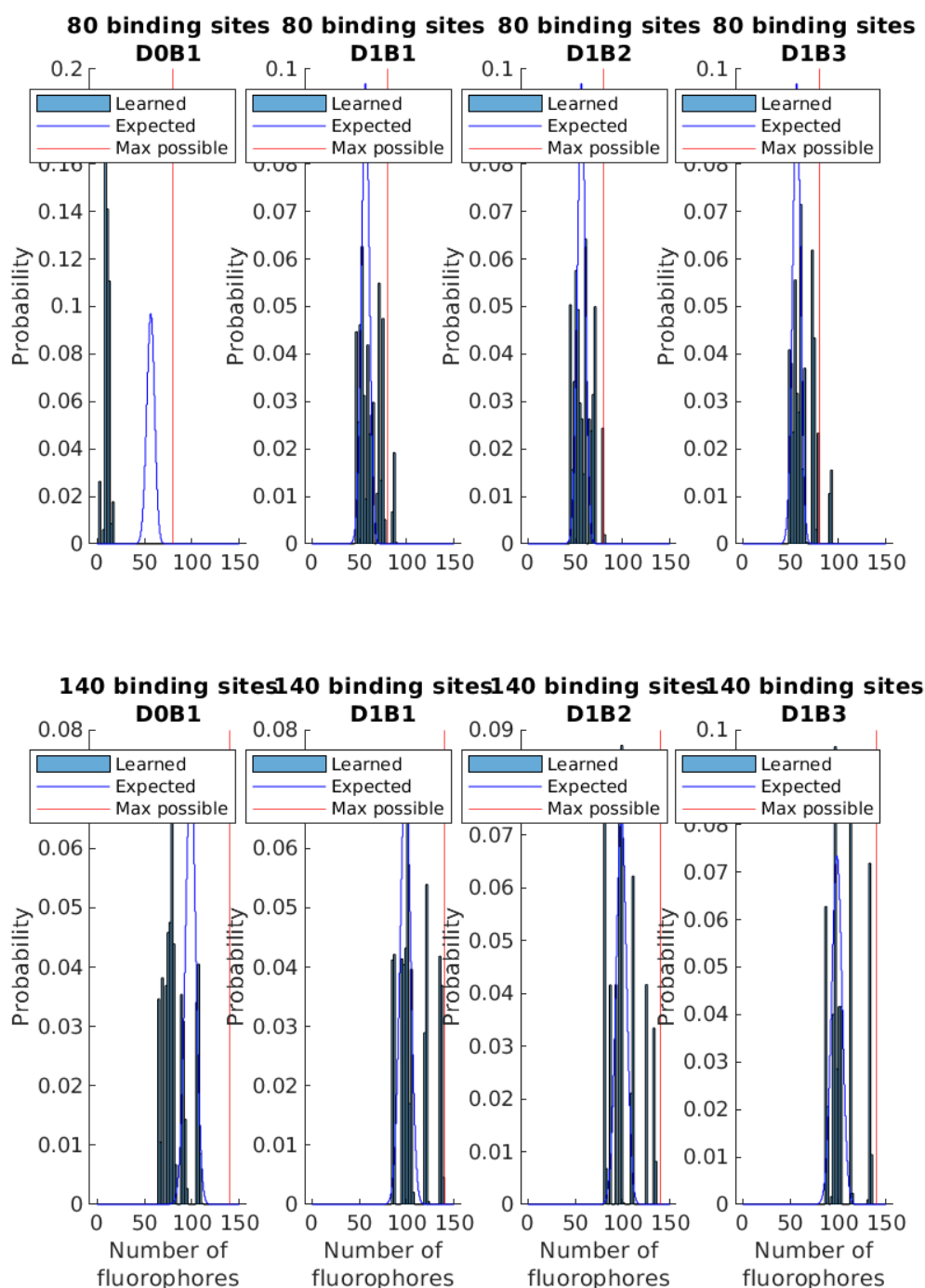
Figure 12: **Inference on more semi-synthetic data.** Here we count the fluorophores from real data where we combine four ROIs into one. Each column shows inference on the data using a different fluorophore model. The top row analyzes data from experiments using DNA origami with 20 binding sites. The bottom row analyzes data from experiments using DNA origami with 35 binding sites. Also plotted are the expected distribution of fluorophores (a binomial distribution), and a vertical line showing the maximum possible number of fluorophores.

rates and brightness parameters along with the number of fluorophores. In order to achieve this, we had to consider multiple ROIs simultaneously.

Despite the improvement afforded by Bayesian nonparametrics in treating fluorophore photophysics, there is an added computational cost. For example, counting as many as a hundred fluorophores can take as much as a week on a typical PC. This can time can be shortened if sampling states for each ROI is done in parallel. As such, the time to completion can be drastically reduced when running the same calculation a high performance computing cluster.

Moving forward, there are a number of ways in which our method could be improved. Firstly, beyond parallelizing the treatment of the code, we can reduce the computational burden by allowing K, the number of loads, to adapt during our sampler's burn in phase. This will optimize K on the fly. Furthermore, any method attempting to count as many fluorophores as we are, including ours, will be sensitive to the choice of photophysical model for the fluorophores (the number of bright and dark states) that needs to be appropriately calibrated. The calibration could be done by isolating single fluorophores and finding state levels using an infinite hidden Markov model [46].

More broadly, however, questions of fixed stoichiometry, as we have explored them herein, are only a small part of deducing life's events within small confined regions. Of greater interest, in the horizon, is the possibility of monitoring time-dependent events within diffraction-limit regions.

References

[1] Jung, S.-R., Fujimoto, B. S. & Chiu, D. T. Quantitative microscopy based on single-molecule fluorescence. *Current opinion in chemical biology* **39**, 64–73 (2017).

[2] Lee, A., Tsekouras, K., Calderon, C., Bustamante, C. & Pressé, S. Unraveling the thousand word picture: an introduction to super-resolution data analysis. *Chemical reviews* **117**, 7276–7330 (2017).

[3] Leung, B. O. & Chou, K. C. Review of super-resolution fluorescence microscopy for biology. *Applied spectroscopy* **65**, 967–980 (2011).

[4] Kolin, D. L. & Wiseman, P. W. Advances in image correlation spectroscopy: measuring number densities, aggregation states, and dynamics of fluorescently labeled macromolecules in cells. *Cell biochemistry and biophysics* **49**, 141–164 (2007).

[5] Coffman, V. C. & Wu, J.-Q. Counting protein molecules using quantitative fluorescence microscopy. *Trends in biochemical sciences* **37**, 499–506 (2012).

[6] Arant, R. J. & Ulbrich, M. H. Deciphering the subunit composition of multimeric proteins by counting photobleaching steps. *ChemPhysChem* **15**, 600–605 (2014).

[7] Michelini, F. *et al.* From "cellular" rna to "smart" rna: multiple roles of rna in genome stability and beyond. *Chemical reviews* **118**, 4365–4403 (2018).

[8] Grußmayer, K. S., Yserentant, K. & Herten, D.-P. Photons in-numbers out: Perspectives in quantitative fluorescence microscopy for in situ protein counting. *Methods and applications in fluorescence* (2018).

[9] Aspelmeier, T., Egner, A. & Munk, A. Modern statistical challenges in high-resolution fluorescence microscopy (2015).

[10] Verdaasdonk, J. S., Lawrimore, J. & Bloom, K. Determining absolute protein numbers by quantitative fluorescence microscopy. In *Methods in cell biology*, vol. 123, 347–365 (Elsevier, 2014).

[11] Ulbrich, M. H. & Isacoff, E. Y. Subunit counting in membrane-bound proteins. *Nature methods* **4**, 319 (2007).

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

[13] Hines, K. E. Inferring subunit stoichiometry from single molecule photobleaching. *The Journal of general physiology* **141**, 737–746 (2013).

[14] McGuire, H., Aurousseau, M. R., Bowie, D. & Blunck, R. Automating single subunit counting of membrane proteins in mammalian cells. *Journal of Biological Chemistry* **287**, 35912–35921 (2012).

[15] Zhang, H. & Guo, P. Single molecule photobleaching (smpb) technology for counting of rna, dna, protein and other molecules in nanoparticles and biological complexes by tirf instrumentation. *Methods* **67**, 169–176 (2014).

[16] Kalafut, B. & Visscher, K. An objective, model-independent method for detection of non-uniform steps in noisy signals. *Computer Physics Communications* **179**, 716–723 (2008).

[17] Chen, Y., Deffenbaugh, N. C., Anderson, C. T. & Hancock, W. O. Molecular counting by photobleaching in protein complexes with many subunits: best practices and application to the cellulose synthesis complex. *Molecular biology of the cell* **25**, 3630–3642 (2014).

[18] Nino, D., Djayakarsana, D. & Milstein, J. N. Nanoscopic stoichiometry and single-molecule counting. *Small Methods* 1900082 (2019).

[19] Xu, J. *et al.* Automated stoichiometry analysis of single-molecule fluorescence imaging traces via deep learning. *Journal of the American Chemical Society* **141**, 6976–6985 (2019).

[20] Staudt, T. *et al.* Statistical molecule counting in super-resolution fluorescence microscopy: Towards quantitative nanoscopy. *arXiv preprint arXiv:1903.11577* (2019).

[21] Liesche, C. *et al.* Automated analysis of single-molecule photobleaching data by statistical modeling of spot populations. *Biophysical journal* **109**, 2352–2362 (2015).

[22] Garry, J., Li, Y., Shew, B., Gradinaru, C. C. & Rutenberg, A. D. Bayesian counting of photobleaching steps with physical priors. *The Journal of Chemical Physics* **152**, 024110 (2020).

[23] Shi, X. *et al.* Counting quantum dot aggregates for the detection of biotinylated proteins. *Chemical Communications* **51**, 2353–2356 (2015).

[24] Delport, F. *et al.* Improved method for counting dna molecules on biofunctionalized nanoparticles. *Langmuir* **26**, 1594–1597 (2010).

[25] Herbert, K. M. *et al.* A heterotrimer model of the complete microprocessor complex revealed by single-molecule subunit counting. *Rna* **22**, 175–183 (2016).

[26] Simonson, P. D. *et al.* Counting bungarotoxin binding sites of nicotinic acetylcholine receptors in mammalian cells with high signal/noise ratios. *Biophysical journal* **99**, L81–L83 (2010).

[27] Durisic, N. *et al.* Stoichiometry of the human glycine receptor revealed by direct subunit counting. *Journal of Neuroscience* **32**, 12915–12920 (2012).

[28] Cawte, A. D., Unrau, P. J. & Rueda, D. S. Live cell imaging of single rna molecules with fluorogenic mango ii arrays. *bioRxiv* 840082 (2019).

[29] Agrawal, P. *et al.* Stoichiometric analyses of soluble cd4 to native-like hiv-1 envelope by single-molecule fluorescence spectroscopy. *Cell Reports* **29**, 176–186 (2019).

[30] Lee, A. A.-T. *Improving the resolution and accuracy of optical tweezers through algorithmic and instrumental advances.* Ph.D. thesis, UC Berkeley (2018).

[31] Li, Y. *Insane in the Membrane: The Functional Assembly of a G Protein Coupled Receptor at the Single-Molecule Level.* Ph.D. thesis, University of Toronto (2018).

[32] Prakash, V. *et al.* Quantitative mapping of endosomal dna processing by single molecule counting. *Angewandte Chemie International Edition* **58**, 3073–3076 (2019).

[33] Tutkus, M. *et al.* Fluorescence microscopy of single liposomes with incorporated pigment–proteins. *Langmuir* **34**, 14410–14418 (2018).

[34] Hariri, A. A. *et al.* Stoichiometry and dispersity of dna nanostructures using photobleaching pair-correlation analysis. *Bioconjugate chemistry* **28**, 2340–2349 (2017).

[35] Hummert, J., Yserentant, K., Fink, T., Euchner, J. & Herten, D.-P. Photobleaching step analysis for robust determination of protein complex stoichiometries. *bioRxiv* (2020).

[36] Cisse, I. I. *et al.* Real-time dynamics of rna polymerase ii clustering in live human cells. *Science* **341**, 664–667 (2013).

[37] Cho, W.-K. *et al.* Mediator and rna polymerase ii clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).

[38] Bishop, C. M. *Pattern recognition and machine learning* (springer, 2006).

[39] Zhou, M., Hannah, L., Dunson, D. & Carin, L. Beta-negative binomial process and poisson factor analysis. In *Artificial Intelligence and Statistics*, 1462–1471 (2012).

[40] Thibaux, R. & Jordan, M. I. Hierarchical beta processes and the indian buffet process. In *Artificial Intelligence and Statistics*, 564–571 (2007).

[41] Gelman, A. *et al. Bayesian data analysis* (Chapman and Hall/CRC, 2013).

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

[43] Hirsch, M., Wareham, R. J., Martin-Fernandez, M. L., Hobson, M. P. & Rolfe, D. J. A stochastic model for electron multiplication charge-coupled devices–from theory to practice. *PloS one* **8**, e53671 (2013).

[44] Huang, F. *et al.* Video-rate nanoscopy using scmos camera–specific single-molecule localization algorithms. *Nature methods* **10**, 653 (2013).

[45] Gael, J. V., Teh, Y. W. & Ghahramani, Z. The infinite factorial hidden markov model. In *Advances in Neural Information Processing Systems*, 1697–1704 (2009).

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

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

[48] Kilic, Z., Sgouralis, I. & Presse, S. Residence time analysis of rna polymerase transcription dynamics: A bayesian sticky hmm approach. *bioRxiv* (2020).

[49] Sivia, D. & Skilling, J. *Data analysis: a Bayesian tutorial* (OUP Oxford, 2006).

[50] Tavakoli, M. *et al.* Pitching single-focus confocal data analysis one photon at a time with bayesian nonparametrics. *Physical Review X* **10**, 011021 (2020).

[51] Jazani, S. *et al.* An alternative framework for fluorescence correlation spectroscopy. *Nature communications* **10**, 1–10 (2019).

[52] Gilks, W. R. & Wild, P. Adaptive rejection sampling for gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **41**, 337–348 (1992).

[53] Robert, C. & Casella, G. *Monte Carlo statistical methods* (Springer Science & Business Media, 2013).

[54] Gilks, W. R., Best, N. G. & Tan, K. Adaptive rejection metropolis sampling within gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **44**, 455–472 (1995).

[55] Lu, H., Zhang, H., Yang, S. & Zheng, Z. Camera parameters auto-adjusting technique for robust robot vision. In *2010 IEEE International Conference on Robotics and Automation*, 1518–1523 (IEEE, 2010).

[56] Casanova, D. *et al.* Counting the number of proteins coupled to single nanoparticles. *Journal of the American Chemical Society* **129**, 12592–12593 (2007).

[57] Glembockyte, V., Lin, J. & Cosa, G. Improving the photostability of red-and green-emissive single-molecule fluorophores via ni2+ mediated excited triplet-state quenching. *The Journal of Physical Chemistry B* **120**, 11923–11929 (2016).

4 Supplemental Information

4.1 Acknowledgements

4.2 Competing Interests

The authors declare that they have no competing financial interests.

4.3 Correspondence

Correspondence and requests for materials should be addressed to Steve Pressé. (email: spresse@asu.edu).

4.4 Remark on transition probabilities

As shown in figure 2 A, fluorophores can transition like "dark to bright", "bright to dark", and "bright to bleached" as well as self transition like "dark to dark", "bright to bright", and "bleached to bleached." This is captured by the layout of $\boldsymbol{\pi}$, where rows correspond to "old state" and columns correspond to "new state".

$$\boldsymbol{\pi} = \begin{bmatrix} \pi_{\sigma_D \to \sigma_D} & \pi_{\sigma_D \to \sigma_A} & 0 \\ \pi_{\sigma_A \to \sigma_D} & \pi_{\sigma_A \to \sigma_A} & \pi_{\sigma_A \to \sigma_B} \\ 0 & 0 & 1 \end{bmatrix} \tag{31}$$

Here, $\pi_{\sigma_B \to \sigma_B} = 1$ because there are no transitions out of the bleached state. Similarly, $\boldsymbol{\pi}_0$ looks like

$$\boldsymbol{\pi}_0 = \begin{bmatrix} \pi_{0 \to \sigma_D} & \pi_{0 \to \sigma_A} & 0 \end{bmatrix}. \tag{32}$$

4.5 Hyperparameter on loads

Each load is sampled from a Bernoulli distribution with a parameter $q^{m,r}$. This parameter $q^{m,r}$ is sampled from a beta distribution with hyperparameters γ and K, where γ is a hyperhyperparameter on all $q^{m,r}$'s and K is the total number of load on and load off fluorophores in ROI r.

$$b^{m,r} \sim \textbf{Bernoulli}\left(q^{m,r}\right) \tag{33}$$

$$q^{m,r} \sim \textbf{Beta}\left(\frac{\gamma}{K}, \frac{K-1}{K}\right) \tag{34}$$

This is the Beta-Bernoulli process (BBP) [39, 40]. The distribution on $q^{m,r}$ is constructed in such a way that as K goes to infinity, $q^{m,r}$ goes to zero. This is in line with our assumption that $K \gg M_r$. It turns out that we can marginalize over the $q^{m,r}$'s to get a Bernoulli distribution for $b^{m,r}$ in terms of γ and K.
Proof:
(for clarity, indices are dropped)

$$
\begin{aligned}
\mathcal{P}(b) &= \int_0^1 dq \, \mathcal{P}(b|q)\,\mathcal{P}(q) \\
&= \int_0^1 dq \, \textbf{Bernoulli}(b|q)\,\textbf{Beta}\left(q|\frac{\gamma}{K}, \frac{K-1}{K}\right) \\
&= \int_0^1 dq \left(q^b (1-q)^{1-b}\right)\left(\frac{\Gamma\left(\frac{\gamma}{K} + \frac{K-1}{K}\right)}{\Gamma\left(\frac{\gamma}{K}\right)\Gamma\left(\frac{K-1}{K}\right)} q^{\frac{\gamma}{K}} (1-q)^{\frac{K-1}{K}}\right) \\
&= \frac{\Gamma\left(\frac{\gamma}{K} + \frac{K-1}{K}\right)}{\Gamma\left(\frac{\gamma}{K}\right)\Gamma\left(\frac{K-1}{K}\right)} \int_0^1 dq \, q^{\frac{\gamma}{K}+b-1}(1-q)^{\frac{K-1}{K}-b+1-1} \\
&= \frac{\Gamma\left(\frac{\gamma}{K} + \frac{K-1}{K}\right)}{\Gamma\left(\frac{\gamma}{K}\right)\Gamma\left(\frac{K-1}{K}\right)} \frac{\Gamma\left(\frac{\gamma}{K}+b\right)\Gamma\left(\frac{K-1}{K}-b+1\right)}{\Gamma\left(\frac{\gamma}{K}+\frac{K-1}{K}+1\right)} \int_0^1 dq \, \textbf{Beta}\left(q|\frac{\gamma}{K}+b, \frac{K-1}{K}-b+1\right) \\
&= \frac{K}{\gamma+K-1} \frac{\Gamma\left(\frac{\gamma}{K}+b\right)}{\Gamma\left(\frac{\gamma}{K}\right)} \frac{\Gamma\left(\frac{K-1}{K}-b+1\right)}{\Gamma\left(\frac{K-1}{K}\right)} \\
&= \begin{cases} \frac{K-1}{\gamma+K-1} & b = 0 \\ \frac{\gamma}{\gamma+K-1} & b = 1 \end{cases} \\
&= \textbf{Bernoulli}\left(\frac{\gamma}{\gamma+K-1}\right)
\end{aligned}
$$

According to our assumption, $K \gg M_r$, $\gamma/(\gamma + K - 1)$ must be a small number (we use $\gamma/(\gamma + K - 1) = .1$ with $K = 100$), but other than that we can set γ by hand knowing that since γ is a hyperhyperparameter its influence on the distribution will matter less and less as data are collected. All together we get

$$b^{m,r} \sim \textbf{Bernoulli}\left(\frac{\gamma}{\gamma + K - 1}\right). \tag{35}$$

4.6 Multiple bright states

Fluorophores like ATTO647N may have multiple bright states [57]. Here instead of having one active state, σ_A, we will have L active states, σ_{A1}, $\ldots$, σ_{AL} each with brightness $\mu_{A1}, \ldots \mu_{AL}$. Note that L must be chosen before hand. The transition matrix is now $L + 2$ by $L + 2$

$$\boldsymbol{\pi} = \begin{bmatrix} \pi_{\sigma_D \to \sigma_D} & \pi_{\sigma_D \to \sigma_{A1}} & \pi_{\sigma_D \to \sigma_{A2}} & \cdots & 0 \\ \pi_{\sigma_{A1} \to \sigma_D} & \pi_{\sigma_{A1} \to \sigma_{A1}} & \pi_{\sigma_{A1} \to \sigma_{A2}} & \cdots & \pi_{\sigma_{A1} \to \sigma_B} \\ \pi_{\sigma_{A2} \to \sigma_D} & \pi_{\sigma_{A2} \to \sigma_{A1}} & \pi_{\sigma_{A2} \to \sigma_{A2}} & \cdots & \pi_{\sigma_{A2} \to \sigma_B} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{36}$$

We note that when we add these bright states we must add L different random variables $(\mu_{A1}, \ldots \mu_{AL})$. In the results section (Section 2.4), we evaluate the results of inference using different numbers for L.

4.7 Time dependent background brightness

Observations of the brightness over time of the ROIs reveal that there is a slight exponentially decaying background brightness [35]. This makes sense because light from fluorophores outside the ROI may bleed into the ROI. Because these fluorophores photobleach we expect that the brightness should go down with time. In the continuum limit, the background brightness from fluorophores outside of the ROI will decay like an exponential. Thus to model the background brightness at each time level, n, we can break the background contribution into two pieces

$$\mu_B^{r,n} = \mu_B^r + \mu_B^{r\prime} \exp(-n/\tau) \tag{37}$$

where $\mu_B^{r,n}$ is the background brightness at the time level, μ_B^r is the background brightness in the absence of external fluorophores (fluorophores outside the ROI), $\mu_B^{r\prime}$ is the contribution to background brightness due to external fluorophores, and τ is the decay time of the background fluorophores (in units of frames or time levels). We note that each ROI will have its own $\mu_B^{r\prime}$, but they share a common τ. Plugging Eq. (37) into Eq. (4) we have

$$w_n^r | s_n^{1:M_r,r}, \mu_A, \mu_B^r, G \sim \textbf{Gamma}\left(\mu_B^r + \mu_B^{r\prime} \exp(-n/\tau) + \sum_{k=1}^{K} b^{k,r} \mu_{s_n^{k,r}}, G\right). \tag{38}$$

To learn $\mu_B^{r\prime}$ and τ we place gamma distribution priors on both

$$\mu_B^{r\prime} \sim \textbf{Gamma}\left(\phi_{\mu_B'}, \psi_{\mu_B'}\right) \tag{39}$$

$$\tau \sim \textbf{Gamma}\left(\phi_\tau, \psi_\tau\right) \tag{40}$$

4.8   The collapsed state space

Under our model, each fluorophore is described by two different random variables, its load and its state. However, we instead describe it with one random variable by lumping together fluorophores with off loads and photobleached fluorophores with on loads. That is, we consider a new state space defined by

$$\sigma'_D = \{b^{m,r} = 1 \text{ and } s_n^{m,r} = \sigma_D\} \tag{41}$$

$$\sigma'_A = \{b^{m,r} = 1 \text{ and } s_n^{m,r} = \sigma_A\} \tag{42}$$

$$\sigma'_B = \{(b^{m,r} = 1 \text{ and } s_n^{m,r} = \sigma_B) \text{ or } (b^{m,r} = 0)\} \tag{43}$$

In this "collapsed state space", we ignore the loads entirely, but allow for fluorophores to start from the photobleached state with probability $(K-1)/(\gamma + K - 1)$ (the probability that a fluorophore has load off). This in turn causes $\boldsymbol{\pi}_0$ to scale down by $\gamma/(\gamma + K - 1)$ as well,

$$\boldsymbol{\pi}_0 = \left[\frac{\gamma}{\gamma + K - 1} \pi_{0D}, \frac{\gamma}{\gamma + K - 1} \pi_{0A}, \frac{K-1}{\gamma + K - 1} \pi_{0B} \right] \tag{44}$$

but notice that $\boldsymbol{\pi}$ is not affected (as off load fluorophores cannot become on and vice versa) nor are any of the other random variables. Thus we can simplify our calculations by simply allowing load off fluorophores to be considered photobleached so long as we keep track of $\boldsymbol{\pi}_0$.

4.9   Remark on sampling fluorophore states

When sampling the fluorophore states, we use the collapsed state space (explained above). Additionally, rather than sample each $s_n^{m,r}$ individually, we sample $s_n^{m,r}, s_n^{m+1,r}, \ldots s_n^{m+J,r}$ together (for $J = 4$ in our case). We additionally shuffle the fluorophore indices m so that at each sampler iteration, different fluorophores are sampled together. Sampling fluorophore states together in this way helps the sampler mixing.

The joint microstate, $S_n^{m,r}$, specifies the states of the fluorophores when we consider them together. For example, one realization of a joint microstate might look like

$$
\begin{aligned}
S_n^{m,r} &= \Sigma_{17} \\
S_n^{m,r} &= \{s_n^{m,r} = \sigma'_D, \\
&\quad\quad s_n^{m+1,r} = \sigma'_A, \\
&\quad\quad s_n^{m+2,r} = \sigma'_B, \\
&\quad\quad s_n^{m+3,r} = \sigma'_A, \\
&\quad\quad s_n^{m+4,r} = \sigma'_D, \\
&\quad\quad s_n^{m+5,r} = \sigma'_B\},
\end{aligned}
$$

or using a notation $\Sigma_1, \Sigma_2, \ldots$ to represent the possible joint microstates that $S_n^{m,r}$ can take,

$$
\begin{aligned}
S_n^{m,r} &= \Sigma_{17} \\
\Sigma_{17} &= \{\sigma'_D, \sigma'_A, \sigma'_B, \sigma'_A, \sigma'_D, \sigma'_B\},
\end{aligned}
$$

where the index, 17, on Σ_{17} is used simply to illustrate an example. The number of values $S_n^{m,r}$ can take $(\Sigma_1, \Sigma_2, \ldots)$ is the Kronecker product of the individual state spaces $(\{\sigma'_D, \sigma'_A, \sigma'_B\})$.

4.10 Conjugate prior on transition probabilities

The choice of Dirichlet prior on the rows of $\boldsymbol{\pi}$ combined with the likelihood give back a Dirichlet distribution [38].

Proof:

Using Bayes's theorem, and dropping all terms that do not depend on $\boldsymbol{\pi}_0$ and $\boldsymbol{\pi}$, the likelihood of our FHMM is

$$
\mathcal{P}\left(s_{1:N}^{1:K,1:R}, b^{1:K,1:R}, \mu_A^{1:R}, \mu_B^{1:R}, G, \boldsymbol{\pi}, \boldsymbol{\pi}_0 | w_{1:N}^{1:R}, \Omega\right)
$$

$$
\propto \mathcal{P}\left(w_{1:N}^{1:R} | s_{1:N}^{1:K,1:R}, b^{1:K,1:R}, \mu_A^{1:R}, \mu_B^{1:R}, G, \boldsymbol{\pi}, \boldsymbol{\pi}_0, \Omega\right) \times \mathcal{P}\left(\boldsymbol{\pi}, \boldsymbol{\pi}_0\right)
$$

$$
\propto \left(\prod_{r=1}^{R}\prod_{m=1}^{K} \mathcal{P}\left(s_1^{m,r} | \boldsymbol{\pi}_0\right) \prod_{n=2}^{N} \mathcal{P}\left(s_n^{m,r} | s_{n-1}^{m,r}, \boldsymbol{\pi}\right)\right) \left(\mathbf{Dirichlet}\left(\pi_0 | \alpha_{\pi_0}\right) \prod_i \mathbf{Dirichlet}\left(\pi_{\sigma_i} | \alpha_{\pi_{\sigma_i}}\right)\right)
$$

$$
\propto \left(\prod_{r=1}^{R}\prod_{m=1}^{K} \pi_{0,\sigma_D}^{[s_1^{m,r}=\sigma_D]} \pi_{0,\sigma_A}^{[s_1^{m,r}=\sigma_A]} \prod_{n=2}^{N}\prod_{i,j} \pi_{i,j}^{[s_{n-1}^{m,r}=\sigma_i][s_n^{m,r}=\sigma_j]}\right) \left(\pi_{0,\sigma_D}^{\alpha_{0,\sigma_D}-1} \pi_{0,\sigma_A}^{\alpha_{0,\sigma_A}-1} \prod_{i,j} \pi_{\sigma_i,\sigma_j}^{\alpha_{\sigma_i,\sigma_j}-1}\right)
$$

$$
\propto \left(\pi_{0,\sigma_D}^{C_{0,D}} \pi_{0,\sigma_A}^{C_{0,A}} \prod_{i,j} \pi_{i,j}^{C_{i,j}}\right) \left(\pi_{0,\sigma_D}^{\alpha_{0,\sigma_D}-1} \pi_{0,\sigma_A}^{\alpha_{0,\sigma_A}-1} \prod_{i,j} \pi_{\sigma_i,\sigma_j}^{\alpha_{\sigma_i,\sigma_j}-1}\right)
$$

$$
\propto \pi_{0,\sigma_D}^{C_{0,D}+\alpha_{0,\sigma_D}-1} \pi_{0,\sigma_A}^{C_{0,A}+\alpha_{0,\sigma_A}-1} \prod_{i,j} \pi_{\sigma_i,\sigma_j}^{C_{i,j}+\alpha_{\sigma_i,\sigma_j}-1}
$$

$$
\propto \mathbf{Dirichlet}\left(\pi_0 | C_0 + \alpha_{\pi_0}\right) \prod_i \mathbf{Dirichlet}\left(\pi_{\sigma_i} | C_i + \alpha_{\pi_{\sigma_i}}\right)
$$

where $C_{0,D}$ and $C_{0,A}$ are the number of fluorophores that started in states σ_D, σ_A, and $C_{i,j}$ is the number of fluorophores that transitioned from σ_i to σ_j, C_0 and C_i are the arrays containing the counts $C_{i,j}$, and the products indexed by i and j go over σ_D, σ_A, and σ_B. So the Dirichlet prior on the initial state probability and transition probabilities allow us to get back independent Dirichlet distributions in the posterior.

Notice that all elements of $\boldsymbol{\pi}_0$ and $\boldsymbol{\pi}$ that are equal to zero also have zero counts in the trace. Because $0^0 = 1$, these components do not affect the probability.

4.11 Notes on initial sampler guess and hyperparameters.

We estimate the background for each ROI by looking at the end of the brightness time trace, when all fluorophores are presumed to be photobleached, for each ROI. This can be estimated using the closed form parameter estimators for the gamma distribution,

$$
\mu_B^{(0),r} = \frac{(N-N')\sum_{n=N'}^{N} w_n^r}{(N-N')\sum_{n=N'}^{N} w_n^r \log(w_n^r) - \sum_{n=N'}^{N}\log(w_n^r)\sum_{n=N'}^{N} w_n^r} \tag{45}
$$

where N' is the time level which we guess all fluorophores are bleached by.

For $\mu_A^{(0)}$, we go through the data by eye and find the brightness drop from one well resolved photobleaching event. Additional bright states are initialized at 90% this value, 80% this value, etc.

For $\boldsymbol{\pi}_0^{(0)}$ we take the maximally naive approach and choose the starting probability for bright and dark to be 50/50. We can choose $\boldsymbol{\pi}^{(0)}$ using this approach, but we find that choosing self transitions to be larger

speeds up the convergence significantly (by about 100-500 iterations)

$$\boldsymbol{\pi}^{(0)} = \begin{bmatrix} .9 & .1 & 0 \\ .1 & .8 & .1 \\ 0 & 0 & 1 \end{bmatrix}$$

For $s_{1:N}^{(0),1:R,1:K}$ we guess that there are no fluorophores in any ROI, i.e that for all loads, $b^{m,r} = 0$.

In our analysis, we chose the hyperparameters to coincide with the initial guess. So $\alpha_0 = [.5, .5, 0]$, $\alpha_{\sigma_D} = [.9, .1, 0]$, $\alpha_{\sigma_A} = [.1, .8, .1]$, $\alpha_{\sigma_B} = [0, 0, 1]$. For the mean background brightness and mean fluorophore brightness, we chose the initial guess to be the mean of the prior by scaling the scale parameter by the shape parameter,

$$\psi_{\mu_A} = \mu_A^{(0)} / \phi_{\mu_A}$$
$$\psi_{\mu_B}^r = \mu_B^{r,(0)} / \phi_{\mu_B}$$