The Annals of Applied Statistics
2021, Vol. 15, No. 4, 1742–1766
https://doi.org/10.1214/21-AOAS1441

A BAYESIAN NONPARAMETRIC APPROACH TO SUPER-RESOLUTION SINGLE-MOLECULE LOCALIZATION

BY MARIANO I. GABITTO[1,*], HERVE MARIE-NELLY[2,‡], ARI PAKMAN[3], ANDRAS PATAKI[1,†], XAVIER DARZACQ[2,§] AND MICHAEL I. JORDAN[4]

[1]*Center for Computational Biology, Flatiron Institute, Simons Foundation,* [*]*mgabitto@simonsfoundation.org;* [†]*apataki@flatironinstitute.org*

[2]*Li Ka Shing Center for Biomedical and Health Sciences, University of California, Berkeley,* [‡]*hervemn@berkeley.edu;* [§]*darzacq@berkeley.edu*

[3]*Department of Statistics and Center for Theretical Neuroscience, Columbia University, aripakman@gmail.com*

[4]*Department of Statistics, University of California, Berkeley, jordan@stat.berkeley.edu*

We consider the problem of single-molecule identification in super-resolution microscopy. Super-resolution microscopy overcomes the diffraction limit by localizing individual fluorescing molecules in a field of view. This is particularly difficult since each individual molecule appears and disappears randomly across time and because the total number of molecules in the field of view is unknown. Additionally, data sets acquired with super-resolution microscopes can contain a large number of spurious fluorescent fluctuations caused by background noise.

To address these problems, we present a Bayesian nonparametric framework capable of identifying individual emitting molecules in super-resolved time series. We tackle the localization problem in the case in which each individual molecule is already localized in space. First, we collapse observations in time and develop a fast algorithm that builds upon the Dirichlet process. Next, we augment the model to account for the temporal aspect of fluorophore photophysics. Finally, we assess the performance of our methods with ground-truth data sets having known biological structure.

1. Introduction. Light microscopes are the workhorse of cellular biology, enabling the study of molecular processes within the cell. The resolution of light microscopes is limited by the interaction of light with the microscope's optical system, due to the physics of diffraction. Diffraction causes a blur on each light point source (lps). The response of the imaging system to a lps was first described by Airy (Airy (1835)) and is represented by the point spread function (psf) of the microscope. The image of an object under a microscope is the superposition of all the lps comprising the object convolved with the psf (Figure 1A). If two lps are close enough, the finite size of the psf prevents their separate recognition. Using this fact in 1873, Abbe (Abbe (1873)) formalized the definition of resolution as the smallest distance between two objects that prevent their individual identification. In particular, Abbe established that two light-emitting sources can be distinguished only if they are separated by a distance of at least $d = \frac{\lambda}{2NA}$, where λ is the wavelength of incoming light and NA is the numerical aperture of the microscope. The resolution of conventional light microscopy is typically limited to around 200 nm.

Super-resolution microscopy (SRM) is an imaging methodology that allows researchers to overcome the diffraction limit imposed by conventional light microscopy (Betzig et al. (2006), Rust, Bates and Zhuang (2006)). SRM resolves photoswitchable fluorophores in a field of view by sparsely and randomly activating individual light emitters and then localizing them with subdiffraction precision (Figure 1B). This technique has revolutionized the

Received August 2020; revised January 2021.
Key words and phrases. Super-resolution microscopy, Bayesian nonparametrics, variational inference.

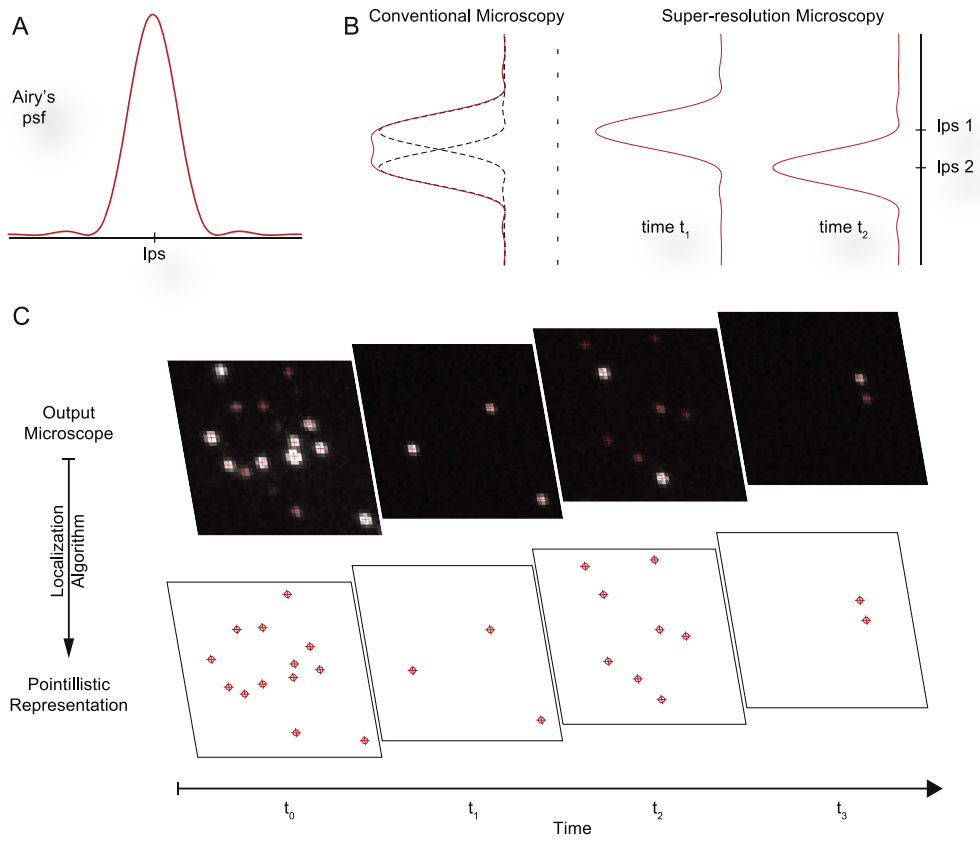

FIG. 1. ***Principles of conventional (CM) and super-resolution (SRM) light microscopy:*** *A) Light emitted by a light point source (lps) gets blurred after traversing the optics of microscopes due to the physics of diffraction. The point spread function (psf) characterizes this spatial light pattern. B) Comparison between CM and SRM in the case of a sample composed of two lps. Left, CM activates both lps preventing their individual identification. Right, in SRM, each lps is activated at different times enabling the localization of each lps circumventing the diffraction limit resolution. C) The result of an SRM experiment is a series of images at different times. A different set of fluorophores is activated at each time point. Localization algorithms transform images into a list of candidate fluorophore locations in space and time.*

field of cellular microscopy by enabling the study of intracellular proteins within cellular compartments at nanometer resolution (Hansen et al. (2018)), the organization of the actin cytoskeleton in neuronal axons and dendrites (Xu, Zhong and Zhuang (2013)), the structure of receptors and scaffolding proteins at synapses (Specht et al. (2013)) and the study of the protein complexes forming the nuclear pore (Szymborska et al. (2013)), to cite a few examples. The typical resolution of diffraction-unlimited microscopy is 10 nm.

Of the many techniques that have been developed to achieve super-resolution imaging, we focus on a class of methods that are generally referred to as *single-molecule-localization microscopy* (SMLM). These techniques include stochastic optical reconstruction microscopy (STORM) (Rust, Bates and Zhuang (2006)), photoactivation localization microscopy (PALM) (Betzig et al. (2006)) and other variants. SMLM techniques rely on the ability of photoactivatable fluorescent proteins (Lippincott-Schwartz and Patterson (2009)) or photoswitchable fluorophores (Dempsey et al. (2011), Heilemann et al. (2005), van de Linde, Heilemann and Sauer (2012)) to alternate between a fluorescence-emitting state and a dark state. By only activating a subset of the total number of fluorophores within the field of view,

the emitting molecules can be individually localized, thereby circumventing the diffraction limit of light.

SMLM experiments are performed routinely and experimental protocols have been honed. The latter includes sample preparation (Olivier et al. (2011)) (Kaplan and Ewers (2015)), fluorophore selection (Dempsey et al. (2011)) (Shcherbakova et al. (2014)) and data acquisition (Pertsinidis, Zhang and Chu (2010)) (Huang (2011)). An SMLM experiment produces data in the form of a time series of images, each of them representing a random and unknown number of fluorophores activated at each time (Figure 1C). Next, the first data-analysis challenge centers on the identification and localization of fluorophores in each image (Deschout (2014), Small and Stahlheber (2014)). Subsequently, a pointillistic representation of the data is built in which each point represents the location of a candidate fluorophore. At this stage, two challenges are intertwined. On the one hand, observations arising from the same fluorophore at different time points should be accounted for (Annibale et al. (2011a), Annibale et al. (2011), Veatch et al. (2012)). On the other hand, once fluorophores are identified, researchers need to extract biological insights from SMLM data by identifying spatial properties of the molecules under study (Kapoor-Kaushik et al. (2016), Owen et al. (2010), Rossy et al. (2014), Sengupta (2011)).

In principle, once observations arising from each fluorophore across time are correctly identified, SMLM is poised to enable accurate counting of single molecules. This process enables the study of protein stoichiometry and dynamics under physiological conditions (Košuta et al. (2020), Lee et al. (2012), Nicovich, Owen and Gaus (2017), Puchner et al. (2013)). Counting is, however, highly dependent on the characteristics of the molecule used as a fluorescent tag, and, in order to obtain an accurate count, several obstacles must be overcome. For example, fluorophore photophysics result in molecule overcounting when fluorescent molecules "blink" by transiently alternating between nonemitting dark and emitting light states (Heilemann et al. (2009), Roy et al. (2011)). Even more troublesome is the possibility that photoconvertible probes can be reactivated after a lengthy stay in the dark state (Annibale et al. (2010), Annibale et al. (2011)). Many algorithms have been devised to compensate for blinking, ranging from semiempirical approaches (Annibale et al. (2011)) to more robust procedures that account for a single dark state (Lee et al. (2012), Rollins et al. (2015)), many dark states (Hummer, Fricke and Heilemann (2016)) or the presence of many fluorophores and binding sites within a diffraction limited spot (Nino et al. (2017)). An ideal workflow, designed to analyze SRM data sets, would be capable of accounting for the different blinking properties of fluorophores while separating each fluorophore in space.

In this article we focus on two of the key statistical challenges in super-resolution data analysis: the localization of individual point-source fluorophores and the estimation of the identity of the observation arising from each fluorophore. These tasks must be carried out with only rough prior knowledge of the number of activated fluorophores and the dynamics of their activation. We devise a statistical approach to analyze STORM or PALM localizations computed by any conventional detection software (Holden, Uphoff and Kapanidis (2011), Ovesnỳ et al. (2014), Sergé et al. (2008)) and to infer the most likely fluorophore arrangement in the field of view. We cast this task as a doubly stochastic marked Poisson process. We use this framework to transform the problem into a density estimation model (Kottas and Sansó (2007), Taddy and Kottas (2012)) and use Bayesian nonparametric (BNP) priors to model the number of fluorophores present in the sample (Blei and Jordan (2006), Broderick, Jordan and Pitman (2013), Huggins and Wood (2014)).

To create a scalable approach able to analyze a high number of points, we avail ourselves of four strategies. First, we create an algorithm capable of creating a high-quality initial condition for our inference algorithm. To do so, we collapse observations across time and use a Dirichlet process (DP) mixture model to infer fluorophore events in the field of view. The DP

is a BNP prior with the flexibility of accounting for an unbounded number of fluorophores. Second, we perform approximate Bayesian posterior inference through the use of variational mean field methods. Variational inference uses numerical optimization to approximate the posterior distribution (Blei, Kucukelbir and McAuliffe (2017), Jordan et al. (1999)), resulting in a procedure which is often orders of magnitude faster than traditional Markov chain Monte Carlo methods. Third, to build a high-performance implementation capable of analyzing millions of points, we accelerate our posterior inference algorithm by exploiting a quadtree data structure (Finkel and Bentley (1974)). Fourth, we explore different fluorophore configurations through the use of fast state-space search techniques. The combination of these strategies creates a scalable inferential workflow able to process large data sets. Finally, to model time dependencies between fluorophore observations, we incorporate the photophysical properties of the fluorophores. We develop a model based on the Markovian Indian Buffet process (Gael, Teh and Ghahramani (2009)). This model is able to discover fluorophore positions, assign observations to each active fluorophore and estimate fluorophore dynamics jointly.

The paper is organized as follows. In Section 2 we begin by presenting the SMLM data sets analyzed in this work. In Section 3 we review the Dirichlet process and discuss a time-independent model for pointillistic data. Then, in Section 4 we extend the model to include fluorophore photophysics and create a time-dependent formulation based on the Markovian Indian Buffet process. In Section 5 we explore the performance of our algorithms on biological data sets and assess the accuracy of our method. Next, in Section 6 we analyze a real biological data set of known geometrical structure. In Section 7 we describe related work. We conclude in Section 8.

A list of notational conventions and further details on our algorithms and experiments can be found in the Supplementary Material (Gabitto et al. (2021)).

2. Experimental data sets. In this work we consider real biological data sets obtained using STORM imaging. The first data set consists of a 3D DNA origami scaffold equipped with multiple handles for the attachment of different molecules (Zanacchi et al. (2017)). The scaffold is 225 nm long and consists of a 12-helix bundle with six inner and six outer helices. It contains 15 attachment points, separated by a distance of 14 nm, that project outward and provide site- and sequence-specific positions to which fluorophores or proteins of interest can be functionalized (Figure 2A). First, at handle position 14, TAMRA fluorophores are attached to enable identification of the DNA scaffold under wide field imaging. Next, complementary handle sequences labeled with Alexa Fluor 647 were attached to handles 1, 7 and 13 of helix 0 to permit identification of single fluorophores (Figure 2B–D). STORM data sets generated using the scaffold were kindly provided by the authors of the original work (Zanacchi et al. (2017)) for the demonstration of our method. We localized individual molecules present in single frames and used these positions as the input to our analysis. A second data set consists of super-resolution STORM imaging of the nuclear pore complex (NPC) in nuclear envelopes. NPCs provide access to the cell nucleus, thereby permitting the transport of proteins and RNA through the nuclear envelope. The function of the NPC is not limited to molecular trafficking; NPCs are also involved in diverse cellular processes (D'Angelo and Hetzer (2008)). The nuclear pore possesses a highly stereotyped configuration: proteins within NPCs are arranged in an eight-fold symmetric, cylindrical assembly consisting of approximately 30 different proteins of the nucleoporin (Nup) family (Kim et al. (2018)). Here, we analyzed a nuclear pore complex data set in which the Nup-107 protein is tagged with Alexa Fluor 647 in the nuclear pore membrane of U-2 OS cells and imaged using dSTORM on a commercial Leica SR GSD 3D microscope (kindly provided by the Reiss Lab, data from Li et al. (2018)) (Figure 3A). Nup-107 proteins belong to the best-studied module within the NPC, the Y-complex (its name describes the shape in which proteins in the module assemble). Nup-107

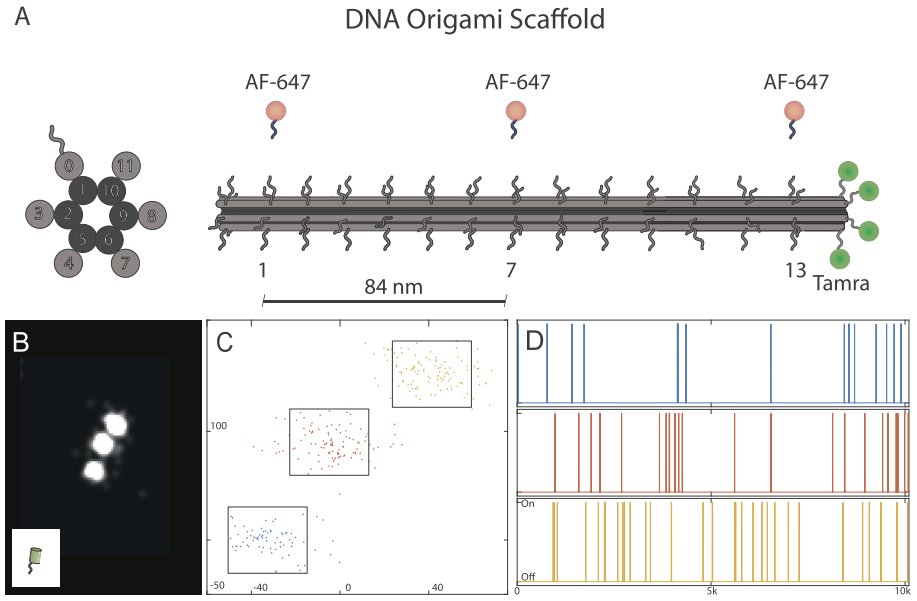

FIG. 2. ***Super-resolution imaging of a DNA origami scaffold*:** *a) DNA origami scaffold representation. Left, cross-section depicting the double helix barrels, sequence specific handles protrude outward from outside helices. Right, the scaffold is* 225 *nm long and possess* 15 *handles. Fluorophores with handle complementary sequence are attached to handles* 1, 7 *and* 13. *At handle position* 14, *TAMRA fluorophores are joined for visualization under wide field fluorescent microscopes. b-c-d) Example of super-resolution imaging of DNA origami scaffold in which sequence specific Alex Fluor* 647 *fluorophores are attached to handles* 1,7 *and* 13. *b) Fluorophore localizations smoothed with a Gaussian kernel* (*width* = 10 *nm*) *and aggregated across the time series. c) x-y positions of fluorophore observations. d) Time traces depicting the frames in which each of the three fluorophores is observed.*

surrounds the NPC, localizing to the nuclear rim and forming an eight-fold symmetrical ring (Beck and Hurt (2017)) (Figure 3B). This highly reproducible symmetrical ring provides a ground-truth structure to which we can compare in order to evaluate our algorithms.

3. A time-independent model of fluorophore locations. A variety of existing techniques produce pointillist representations of fluorophore centers (Holden, Uphoff and Kapanidis (2011), Ovesnỳ et al. (2014), Sergé et al. (2008)). These locations are prone to error due to their lack of correction for temporal effects (i.e., blinking) and their failure to integrate measurement uncertainty into the analysis. Moreover, spurious fluorophore locations can be created as the result of inaccurate modeling of the microscopic point spread function and the nuances of optimization algorithms. Any algorithm attempting to correct these errors faces a tremendous computational complexity, given that a typical SRM dataset contains $\sim$100k observations interspersed among $\sim$10k time points.

Our approach addresses these limitations within a Bayesian framework by precomputing an initial fluorophore configuration with a fast algorithm. This algorithm collapses time information and performs inference in a timeless data set, reducing the computational complexity of the problem. Moreover, to account for uncertainty in the number of fluorescent proteins in the sample, our modeling formulation is based on a Bayesian nonparametric approach. To perform posterior parameter inference at the scale of our application, we avail ourselves of variational inference methods. Once a robust initial fluorophore configuration is calculated, we use it as a seed to a more complex time-dependent model.

We begin this section by briefly reviewing the BNP prior that we used to model fluorophore centers follow by the description of our time-independent model and posterior parameter inference.

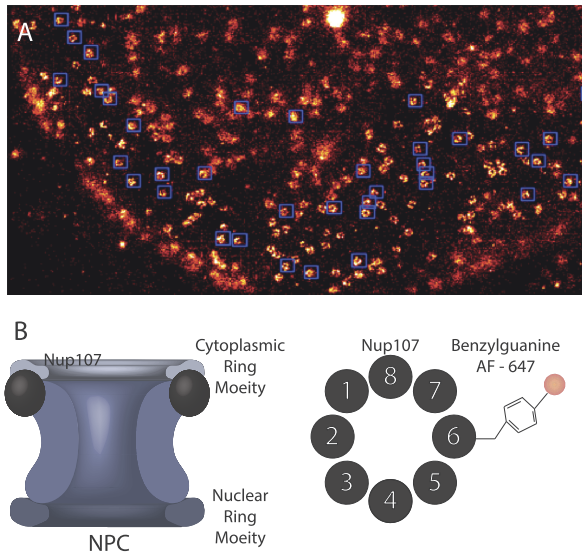

FIG. 3. ***Super-resolution imaging of the nuclear pore complex***: *a) Example field of view in which nuclear pore complexes are imaged with super-resolution microscopy. Nuclear pore complexes are identified by template cross-correlation and enclosed by blue rectangles. b) Nuclear pore complex schematic. Left, cross-section of the nuclear pore complex on the nuclear membrane indicating in black, Nup-*107*. Right, top view of the nuclear pore complex highlighting Nup-*107 *octagonal symmetry labeled with Alexa Fluor* 647.

3.1. *The Dirichlet process.* The Dirichlet process (DP) (Ferguson (1973)) is a distribution over probability measures. The DP is characterized by a scaling parameter $\alpha_0 \in \mathbb{R}^+$ and a base probability measure G_0, defined on the set Θ. A draw from a DP is an atomic probability measure. Sethuraman (1994) provided an explicit representation of a draw from a DP, via a *stick-breaking construction*,

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k}, \tag{3.1}$$

where δ_{θ_k} is an atom located at $\theta_k \in \Theta$, $\theta_k \sim G_0$ and the random weights $\bar{\pi} = [\pi_1, \pi_2, \dots]$ depend on α_0 as follows:

$$\begin{aligned} \nu_k &\sim \text{Beta}(1, \alpha_0), \\ \pi_k &= \nu_k \prod_{j<k} (1 - \nu_j), \quad k = 1, 2, \dots. \end{aligned} \tag{3.2}$$

The random sequence $\bar{\pi}$ sampled according to equation (3.2) is said to follow a GEM distribution (Ewens (1990)).

The atomic nature of the DP has been exploited by many authors by using a DP as a prior for Bayesian mixture models (Escobar and West (1995), Neal (1992), Rasmussen (2000)). In this setting the random weights of the atomic measure correspond to mixing proportions, and the locations of the atoms represent the parameters of the mixture components. Given observations (x_n) with parameters (θ_n), each θ_i is sampled from G, while each x_i has distribution $P(\theta_i)$ parametrized by θ_i. The following generative process describes the overall nonparametric mixture model:

$$\begin{aligned} G|\alpha_0, G_0 &\sim \text{DP}(\alpha_0, G_0), \\ \theta_i|G &\sim G, \\ x_i|\theta_i &\sim P(\theta_i). \end{aligned} \tag{3.3}$$

Clustering arises under this model, due to the fact that G is discrete, such that many θ_i can take the same value. An alternative way to capture this aspect of the model is by using a stick-breaking construction, including cluster assignment variables z_i sampled from $\bar{\pi}$. This respresentation is described as follows:

$$\begin{aligned} \bar{\pi}|\alpha_0 &\sim \mathrm{GEM}(\alpha_0), \\ z_i|\bar{\pi} &\sim \mathrm{Categorical}(\bar{\pi}), \\ \theta_k|G_0 &\sim G_0, \\ x_i|z_i, (\theta_k) &\sim P(\theta_{z_i}). \end{aligned} \tag{3.4}$$

3.2. *Model overview.* Our observed data is a set of two- or three-dimensional fluorophore locations x_n (i.e., putative fluorophores at different time frames). Each observation x_n has an associated localization accuracy σ_n^2 that informs confidence in the estimated location (σ_n^2 is obtained by typical preprocessing steps). We begin our analysis by collapsing observations across time and inferring fluorophore centers in this timeless collapsed data set. These inferences can be performed quickly, and they provide a seed for our more complex time-dependent model, described below in Section 4.

We model observations as arising from a nonhomogeneous marked spatial Poisson process (NHMSPP), defined on a space $R \times \mathbb{R}^+$, where R denotes the observation box $R = [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$. This NHMSPP represents a marked process in which each observation $x \in R$ contains a mark $\sigma^2 \in \mathbb{R}^+$. We define the conditional intensity function $\lambda(x|\sigma^2) : R \times \mathbb{R}^+ \longrightarrow \mathbb{R}^+$. For such random processes, the following hold true:

(i) For any bounded set $S \in R$, the number of points in S is Poisson distributed, $N(S) \sim \mathrm{Po}(\Lambda(S)) = \frac{\Lambda(S)^{N(S)} e^{-\Lambda(S)}}{N(S)!}$, where $\Lambda(S) = \int_S \lambda(x|\sigma^2)\, dx$.

(ii) Given $N(S)$, the point locations within S are i.i.d. with density $\frac{\lambda(x|\sigma^2)}{\Lambda(S)}$.

By specifying the mean measure in terms of two independent factors—a total scalar intensity λ_0 and a spatial density $f(x|\sigma^2)$, $\lambda(x|\sigma^2) = \lambda_0 f(x|\sigma^2)$—we can write the Poisson process likelihood in separable form, thereby transforming the problem into one of density estimation,

$$\begin{aligned} p\big(x|\sigma^2, \lambda\big) &= p\big(x|\sigma^2, N\big) p(N) = e^{-\int_R \lambda_0 f(x|\sigma^2)} \prod_n \lambda_0 f\big(x|\sigma^2\big) \\ &= e^{-\lambda_0} \lambda_0^N \prod_n f\big(x|\sigma^2\big). \end{aligned} \tag{3.5}$$

Placing a gamma prior on λ_0 permits Bayesian inference in a manner that is independent of the spatial density.

Next, we transform the process into a doubly stochastic Poisson process (Cox process) by specifying a random process density $f(x|\sigma^2)$, making use of the Dirichlet process (Ickstadt and Wolpert (1999), Kottas and Sansó (2007)). Building on Rubin-Delanchy et al. (2015), we propose a mixture of two components: spatially random background noise with constant density $\frac{1}{\mathrm{Vol}(R)}$ (where $\mathrm{Vol}(R)$ represents the volume of the box R) and a DP containing a random unbounded number of fluorophores. The DP prior is consistent with the idea that not every fluorophore is observed during the imaging experiment and that the number of fluorophores would increase if the experiment would have continued. We have

$$\begin{aligned} f\big(x_i|\sigma_i^2, G, \pi_0\big) &= \pi_0 \frac{1}{\mathrm{Vol(R)}} + (1 - \pi_0) \int \mathcal{T}\, \mathrm{Normal}\big(x_i|\mu, \sigma_i^2\big)\, dG(\mu), \\ G|\alpha_0, G_0 &\sim \mathrm{DP}(\alpha_0, G_0), \end{aligned} \tag{3.6}$$

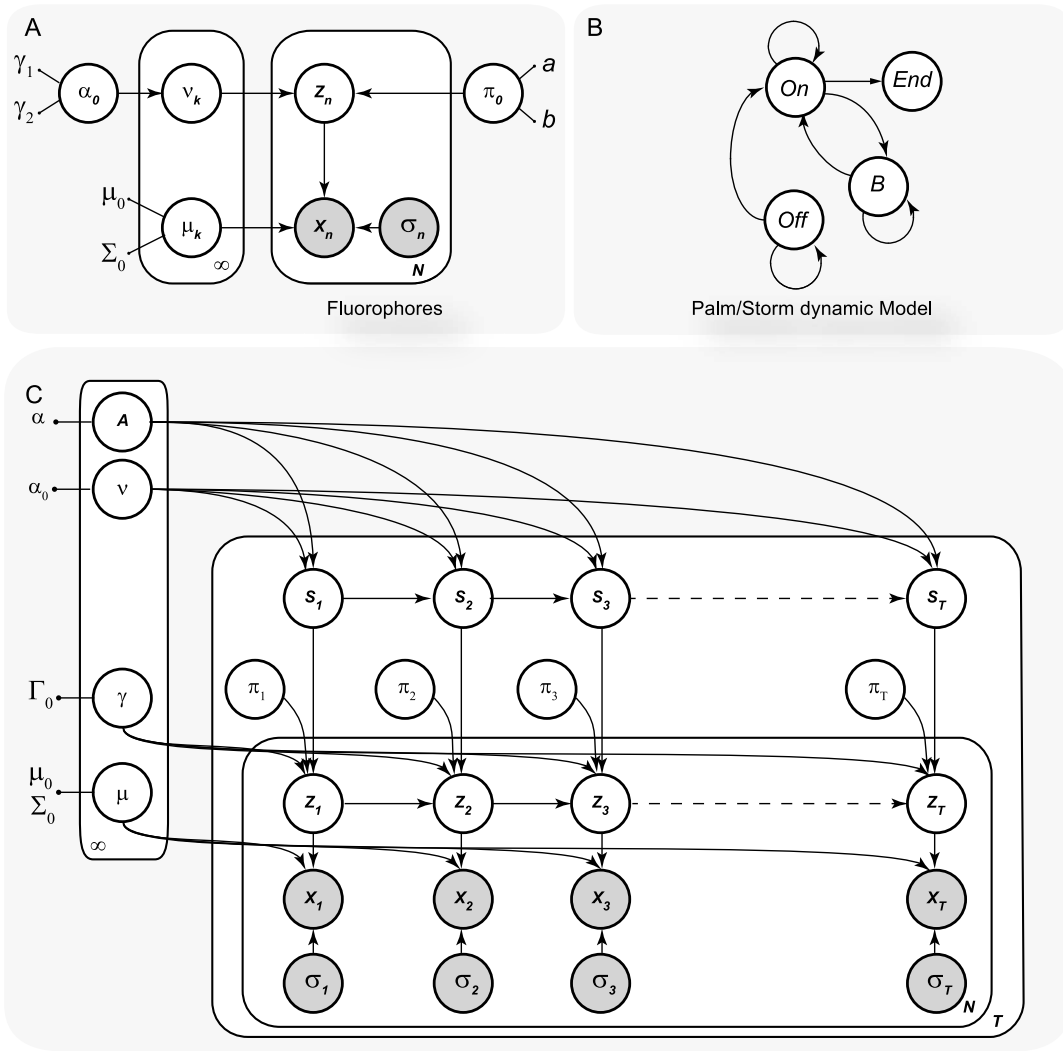

FIG. 4. ***Time-independent and time-dependent Graphical Models for super-resolution localization*****:** *a) Time-independent graphical model. b) Different states for the hidden Markov model used to model storm or palm data sets. c) Time-dependent graphical model.*

where π_0 is a mixing proportion guaranteeing integration over the volume to 1. In this construction we assume that every observation arises either from a constant background noise distribution or is a fluorophore which is distributed according to a truncated Gaussian distribution defined in R with mean μ, denoting the fluorophore position.

We define priors for the parameters of the model as follows. To infer a random number of fluorophores and random fluorophore location, we parametrize our DP prior using the stick-breaking (SB) construction. The graphical model representation is summarized in Figure 4, and the specification is as follows:

$$
\begin{aligned}
\bar{\pi}\,|\alpha_0 &\sim \mathrm{GEM}(\alpha_0),\\
\mu_k|\mu_0, \sigma_0^2 &\sim \mathrm{Normal}\big(\mu_0, \sigma_0^2\big), \quad k = 1, 2, \ldots,\\
\pi_0|a_0, b_0 &\sim \mathrm{Beta}(a_0, b_0),\\
z_n|\pi_0, \bar{\pi} &\sim \mathrm{Categorical}\big(\pi_0, (1-\pi_0)\bar{\pi}\big),
\end{aligned}
\tag{3.7}
$$

$$x_n|z_n = 0 \sim \frac{1}{\mathrm{Vol}(R)},$$

$$x_n|z_n = k, (\mu_k) \sim \frac{\mathrm{Normal}(\mu_k, \sigma_n^2)\mathbb{I}(x_n \in R)}{\prod_{i=x,y,z}[\Phi((b_i - \mu_k^i)/\sigma_n^i) - \Phi((a_i - \mu_k^i)/\sigma_n^i)]}, \tag{3.8}$$

where $n \in \{1, \ldots, N\}$ in the last three equations, $\mathbb{I}$ is an indicator function, Φ is the standard Gaussian cumulative function needed to normalize the truncated Gaussian and σ_n^2 denotes a diagonal covariance matrix which ensures the factorization of the normalization constant.

These modeling assumptions define a joint distribution, $p(x, z, \pi_0, \mu, \nu|\sigma)$, for the data and the latent structure of the model, where we use ν for the stick-breaking representation (3.2). After observing data (x, σ^2), our inferential goal is to obtain the posterior distribution of the fluorophore locations μ and assignments z. We refer to μ and the parameters of the stick-breaking prior as global parameters because they generalize to new observations, in contrast to the states z_n which are local to a specific observation.

3.3. *Mean field variational inference.* Exact posterior inference is infeasible in our model. The two most common approaches to perform approximate posterior inference are Markov chain Monte Carlo (MCMC) (Gelfand and Smith (1990)) and variational inference (VI) (Wainwright and Jordan (2008)). MCMC proceeds by sampling an ergodic Markov chain constructed such that its stationary distribution is the posterior; the posterior is then approximated by an empirical estimate. VI approximates the posterior distribution by projecting onto a family of candidate distributions, turning an integration problem into one of optimization.

While MCMC procedures have asymptotic convergence guarantees, VI can underestimate the variance of the posterior density (Turner and Sahani (2011)). However, in many large-scale problems empirical research has shown that VI does not produce inaccurate posterior predictive densities, and exhibits computational speedups that can be orders of magnitude faster than MCMC (Blei and Jordan (2006), Blei, Kucukelbir and McAuliffe (2017), Braun and McAuliffe (2010), Regier et al. (2019)). In our case the computational issues are critical—the inference algorithm must scale to millions of observations, and it should be able to quickly explore different fluorophore configurations and infer their fluorescent state (see next section).

We, therefore, resort to computing an approximation to the posterior distribution using VI. Let Q be a family of distributions on $\Theta = (\pi_0, \nu, z, \mu)$, the space of latent variables. For $q \in Q$, we search for a family of distributions that maximizes the evidence lower bound ("ELBO" $= \mathcal{L}(q)$),

$$\begin{aligned}\log p(x) &= \log \int p(x, \Theta)\, d\Theta = \log \int p(x, \Theta) \frac{q(\Theta)}{q(\Theta)}\, d\Theta = \log \mathbb{E}_q\left[\frac{p(x, \Theta)}{q(\Theta)}\right] \\ &\geq \mathbb{E}_q\big[\log p(x|\Theta)\big] - D_{\mathrm{KL}}\big(q(\Theta), p(\Theta)\big) \\ &=: \mathcal{L}(q).\end{aligned} \tag{3.9}$$

We seek a distribution q over the latent variables that is close to the true posterior and also lies within a factorized family, $q(\Theta) = q(\pi_0)q(\nu)q(\mu)q(z)$. Each of the factors belongs to a particular member of the exponential family, except for the truncated normal distribution that characterizes the fluorophore centers. For the latter we make use of the fact that our problem contains strong spatial information. In particular, we possess a priori information regarding the scale of the variance of each fluorophore center, given by the average uncertainty of the observations. By augmenting the size of the bounding box, we can assume that $\Phi(\frac{a-\mu}{\sigma}) \sim 0$ and $\Phi(\frac{b-\mu}{\sigma}) \sim 1$, removing the need to explicitly truncate. Numerical exploration of the

validity of this approximation is performed in the Supplementary Material (Gabitto et al. (2021)).

By approximating the generative model in an unconstrained space, we restrict our subsequent analysis to a variational distribution with the following factorized density:

$$q(\Theta) = q(\pi_0) \prod_{n=1}^{N} q(z_n) \prod_{k=1}^{K} q(v_k) q(\mu_k), \tag{3.10}$$

where

$$\begin{aligned} q(z_n) &= \text{Categorical}(r_{nk}), \\ q(v_k) &= \text{Beta}(\alpha_{k1}, \alpha_{k2}), \\ q(\mu_k) &= \text{Normal}(\tilde{\mu_k}, \tilde{\tilde{\sigma}_k}), \\ q(\pi_0) &= \text{Beta}(\tilde{a}, \tilde{b}). \end{aligned} \tag{3.11}$$

3.4. *Inference*. The computational task for our posterior inference algorithm is to find a set of parameters that maximizes the ELBO. Our algorithm updates free parameters in the variational distribution via coordinate ascent variational inference (CAVI). We present updates for global and local factors that converge to a local maximum. To simplify calculations, the ELBO is arranged into three terms that account for data generation, global variables characterizing our stick-breaking construction and an entropic term,

$$\begin{aligned} \mathcal{L}(q) &= \mathbb{E}_q\big[\log p(x|\Theta) + \log p(\Theta) - \log q(\Theta)\big] \\ &= \mathcal{L}(q)_{\text{Data}} + \mathcal{L}(q)_{\text{Global}} + \mathcal{H}, \end{aligned} \tag{3.12}$$

where

$$\begin{aligned} \mathcal{L}(x, r, \tilde{\mu})_{\text{Data}} &= \mathbb{E}_q\left[\log p(x|z, \mu) + \log \frac{p(\mu)}{q(\mu)}\right], \\ \mathcal{H}(z) &= -\mathbb{E}_q\big[\log q(z)\big], \\ \mathcal{L}(r, \tilde{\nu})_{\text{Global}} &= \mathbb{E}_q\left[\log p(z|\pi) + \log \frac{p(\nu)}{q(\nu)} + \log \frac{p(\pi_0}{q(\pi_0)}\right], \end{aligned}$$

where all expectations are taken with respect to the variational distribution. Due to the conjugate exponential family terms in the ELBO, the CAVI updates are easy to compute; see the Supplementary Material for the details (Gabitto et al. (2021)).

Finally, we treat the hyperparameter α_0 as random in both the generative model and the variational distribution (Blei and Jordan (2006)),

$$\alpha_0|\gamma_1, \gamma_2 \sim \text{Gamma}(\gamma_1, \gamma_2). \tag{3.13}$$

We explore the accuracy of our variational approximation in the Supplementary Material (Gabitto et al. (2021)).

3.5. *Scalable inference by exploiting spatial constraints*. Standard variational inference assigns to each data point a positive posterior probability r_{nk}, effectively assuming that each observation can arise from any fluorophore. This instantiates matrices that demand dense memory storage and computation that scales with the total number of clusters. Our problem presents certain advantages due to the locality of its fluorophore assignments. This advantage translates into near certainty that only a few clusters have meaningful posterior mass for any given observation. Unlike approaches that instantiate a fixed number of clusters (Rubin-Delanchy et al. (2015)), our approach assigns nonzero mass to clusters residing within a

certain spatial distance of each observation; this distance acts as a threshold and is the only tunable parameter.

We harness the spatial locality to speed up computations by making use of a *quadtree*, a tree-like data structure that recursively subdivides two-dimensional space into four quadrants (Finkel and Bentley (1974)). We exploit quadtree decomposition during the computation of local assignments. This step computes the posterior probability of assigning an observation n to fluorophore k for each observation r_{nk}. Our CAVI algorithm optimizes $\bar{r}_n$ by fixing global fluorophore parameters according to the following objective:

$$\mathcal{L}_n(r_n) = \mathbb{E}_q\big[\log p(x|\mu)\big] + \mathbb{E}_q[\log \pi] - r_n \log(r_n), \tag{3.14}$$

subject to the constraint that $\sum_k r_{n,k} = 1$ and $r_{n,k} > 0$. The first two terms represent the log posterior assignment of observation n to cluster k. Following Rubin-Delanchy et al. (2015), we replace the variational objective with a new objective that dynamically limits the number of clusters to which an observation can be assigned. Instead of instantiating r_n as a dense vector, we compute only the nonzero mass entries for each observation as determined by the clusters found within a certain distance of the observation. The new optimization problem can be written as

$$\begin{aligned} r_n^* &= \arg\max_{r_n} \mathcal{L}_n(r_n) \\ \text{s.t.} \quad & \sum_k r_{n,k} = 1, \quad r_{n,k} \geq 0, \forall k \in S_k, \end{aligned} \tag{3.15}$$

where the set S_k is computed dynamically at each iteration by querying a quadtree structure built on the locations of the K fluorophores. Observations that return no fluorophore centers are automatically designated as noise. Selecting the threshold distance results in a tradeoff between execution speed and inferential accuracy. In our case the tradeoff is a favorable one, due to the strong locality of the observations arising from each cluster. This new constrained optimization problem can be solved by exponentiation of the log posterior assignment and normalization of the subset of active fluorophores; see Figure 5 for a numerical example of quadtree scaling.

3.6. *Reliable Bayesian inference via state space adaptation.* Variational inference algorithms converge to local optima. We thus implement a multiple trial procedure that yields an algorithm that has the flexibility to improve initial assignment estimates. Specifically, we develop a series of fluorophore proposals aimed to obtain improvements in the ELBO. We interleave proposals that randomly split or merge observations assigned to fluorophores. We also create proposals that create or delete fluorophores by removing or assigning points to noise. To evaluate proposals rapidly, we simplify ELBO calculations for the gap between old and new fluorophores' configurations (see Figure 6). These calculations are reproduced in the Supplementary Material (Gabitto et al. (2021)).

4. A time-dependent model of fluorophore locations. We turn to the second phase of our single-molecule localization procedure. Here, we consider each individual observation in space and time, taking into account the photophysical properties of the emitting molecules. In this formulation a large number of fluorescent molecules are present in the sample, but only a fraction of them are visible at each time point. We design a statistical model capable of analyzing spatiotemporal localization by relating observations at each time point to a collection of K fluorescent time series, where K is unknown and subject to posterior inference. Major challenges include the need to distinguish observations from background noise and the possibility of assigning more than one observation to a given time series at each time

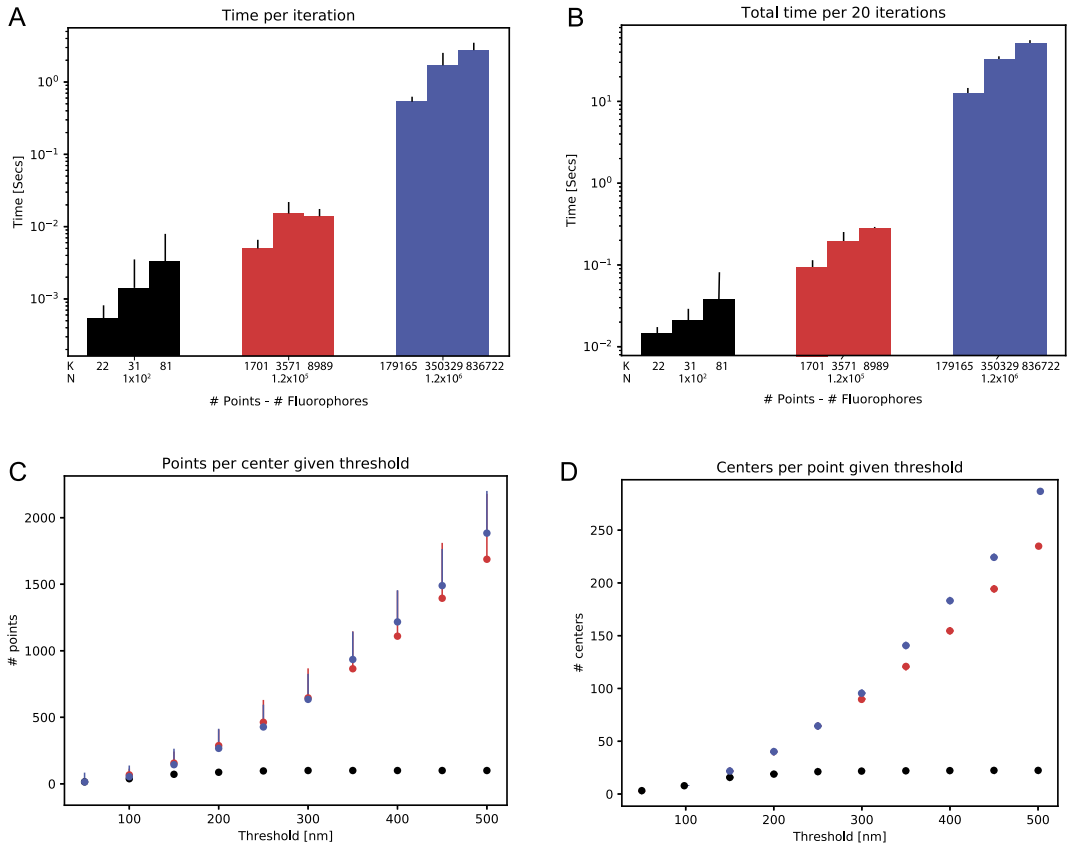

FIG. 5. ***Performance of quadtree acceleration.*** *For three different data sets, containing* 100, 12×10^3 *and* 1.2×10^6 *points, respectively, the algorithm was initialized with three initial conditions that gave rise to different numbers of initial clusters: a) Computational time for one iteration for each of the three data sets. b) Total time to compute* 20 *complete iterations, including local and global steps. By varying the threshold distance, every quadtree query returns a different number of clusters* $k' < K$ *for each point. c) Average number of points per center and d) average number of centers per point. Thresholds lower than* 150 *nm return an equal number of centers independently of the size of the data set.*

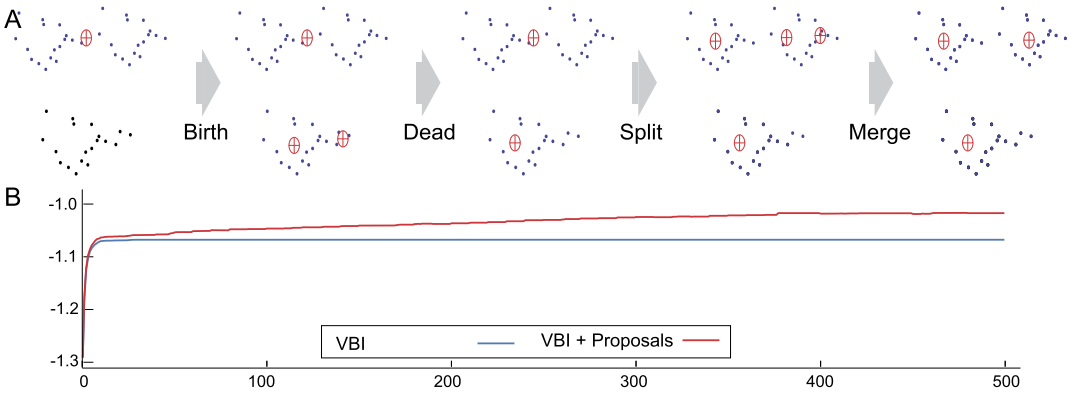

FIG. 6. ***State space exploration through different proposal configurations:*** *a) Cartoon representation showing how different configurations can evolve under birth, death, split and merge proposals. b). ELBO evolution with and without state space exploration on a data set of* 12,000 *points.*

point. To address these challenges, we based our model on a Bayesian nonparametric prior representing an unbounded number of components, each consisting of a Markov chain. At each time point, observations can be assigned to only one of the active chains. Our model is based on the Markovian Indian Buffet process (M-IBP) (Gael, Teh and Ghahramani (2009), Valera, Ruiz and Perez-Cruz (2015)) together with a time-dependent Dirichlet distribution.

4.1. *A dynamic prior that shares features across time points.* In this section we derive a prior distribution over binary matrices with a finite number of columns, each column representing a Markov chain model of a fluorophore. This formulation is closely related to the M-IBP, having a temporal dynamics that is specialized to the time evolution of fluorophores. We begin by introducing a latent state variable, s_t^k, representing the state of the fluorophore k at time point t. The dynamics of this variable is governed by a transition matrix A such that

$$s_t^k | A, s_{t-1}^k = a_{s_{t-1}^k, s_t^k}, \quad A = \begin{pmatrix} a_{00} & a_{01} & 0 & 0 \\ 0 & a_{11} & a_{1b} & a_{1d} \\ 0 & a_{b1} & a_{bb} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{4.1}$$

with prior distributions given by

$$a_{01}^k | \alpha_0 \sim \text{Beta}(\alpha_0/K, 1), \quad a_i | \alpha \sim \text{Dir}(\alpha), i = 1, b. \tag{4.2}$$

The transition matrix models an inactive state (0), a light-emitting state (1), a blinking state (b) and a dead state (d). Next, we introduce a feature vector f_k^t associated with the presence ($f_k^t = 1$ or $s_t^k = 1$, on) or absence ($f_k^t = 0$ or $s_t^k = 0$, 2, 3 inactive, blinking or dead) of fluorophore k at time point t. This feature vector represents a binary matrix having as many columns as there are fluorophores present in the sample. Each feature vector follows a time evolution generated through the latent states,

$$P(f_t^k) \sim \text{Bernoulli}\big(a_{01}^{[S_{t-1}^k=0]} a_{b1}^{[S_{t-1}^k=b]} a_{11}^{[S_{t-1}^k=1]}\big). \tag{4.3}$$

By introducing a set of count variables, c, to indicate the number of transitions between two states (e.g., c_{01} is the number of transitions from the state 0 to 1), we can write the probability of the entire binary matrix as follows:

$$P(F|A) = \prod_k a_{01}^{c_{01}^k} (1 - a_{01})^{c_{00}^k} a_{b1}^{c_{b1}^k} (1 - a_{b1})^{c_{bb}^k} a_{11}^{c_{11}^k} a_{1b}^{c_{1b}^k} (1 - a_{1b} - a_{11})^{c_{1d}^k}. \tag{4.4}$$

In a manner similar to the M-IBP, we calculate the marginal over the matrix F by integrating out transition probabilities,

$$\begin{aligned} P(F|\alpha_0, \alpha) = {} & \frac{\alpha_0^K}{K^K} \prod_k \frac{\Gamma(c_{00}^k + 1)\Gamma(c_{01}^k + \frac{\alpha_0}{K})}{\Gamma(c_{00}^k + 1 + c_{01}^k + \frac{\alpha_0}{K})} \frac{\Gamma(2\alpha)\Gamma(3\alpha)}{\Gamma(\alpha)^5} \\ & \cdot \frac{\Gamma(c_{bb}^k + \alpha)\Gamma(c_{b1}^k + \alpha)}{\Gamma(c_{bb}^k + c_{b1}^k + 2\alpha)} \frac{\Gamma(c_{11}^k + \alpha)\Gamma(c_{1b}^k + \alpha)\Gamma(c_{1d}^k + \alpha)}{\Gamma(c_{11}^k + c_{1b}^k + c_{1d}^k + 3\alpha)}. \end{aligned} \tag{4.5}$$

This represents the probability of each binary matrix under our time-dependent prior. Next, we use this matrix to assign observations to one of the active chains at each time point ($f_t^k = 1$). We do so by associating to each chain a gamma variable (γ^k). Then, at each time point we form a Dirichlet distribution by normalizing the gamma variables over the space of active chains,

$$\pi_t^k | \gamma^k, f_t^k = \begin{cases} \dfrac{\gamma^k}{\sum_l \gamma^l f_t^l}, & \text{if } f_t^k = 1; \\ 0, & \text{otherwise.} \end{cases} \tag{4.6}$$

Finally, observations at time t are assigned to active chains by using a categorical distribution, Categorical(π_t).

In the Supplementary Material we present the infinite, that is, nonparametric, limit of our model (Gabitto et al. (2021)). To obtain this model, we need to define two infinite mathematical constructs. In the first one we derive a BNP prior over binary matrices that follows the fluorophore dynamics. The prior is exchangeable in the columns, and it is also Markov exchangeable in the rows. In the second construct we use the language of completely random measures (CRM) to show that a thinned CRM, based on the Gamma process, represents our model in the infinite case (Foti et al. (2013)).

4.2. *Model overview.* Our model relies on the Markovian dynamics of the time-dependent binary matrix prior to generate a sparse set of active fluorophores at each time point from which observations can be drawn. We present the time-dependent fluorophore model as a graphical model in Figure 4, and we provide the full specification as follows:

(1) Draw parameters

$$\pi_0|a_0, b_0 \sim \text{Beta}(a_0, b_0), \quad a_i|\alpha \sim \text{Dir}(\alpha), i = 1, b.$$

(2) For $k = 1, \ldots, K$, draw chain parameters

$$\mu_k|\mu_0, \sigma_0^2 \sim \text{Normal}(\mu_0, \sigma_0^2), \qquad \gamma_k|\Gamma_0 \sim \text{Gamma}\left(\frac{\Gamma_0}{K}, 1\right),$$

$$a_{01}^k|\alpha_0 \sim \text{Beta}\left(\frac{\alpha_0}{K}, 1\right).$$

(3) For each time point $t = 1, \ldots, T$:
For $k = 1, \ldots, K$:

$$s_t^k|s_{t-1}^k, \quad A^k \sim a_{s_t, s_{t-1}}^k$$

$$\pi_t^k|\gamma^k, \quad s_t^k = \begin{cases} \dfrac{\gamma^k}{\sum_l \gamma^l [s_t^l = 1]}, & \text{if } s_t^k = 1; \\ 0, & \text{otherwise.} \end{cases}$$

For $n_t = 1, \ldots, N_t$:

$$z_{n_t}^t|\pi_0, \bar{\pi}^t \sim \text{Categorical}(\pi_0, (1 - \pi_0)\bar{\pi}^t)$$

$$x_{n_t}^t|z_{n_t}^t = 1 \sim \frac{1}{\text{Vol}(R)}$$

$$x_{n_t}^t|z_{n_t}^t = k, \quad (\mu_k) \sim \text{Normal}(\mu_k, \sigma_{n_t}^{t,2}).$$

4.3. *Mean field variational inference.* In this section we develop a variational inference algorithm to approximate the posterior distribution of the temporal model. We leverage the fact that a correspondence exists between the assignment of an observation at a particular time point to a fluorophore, only if the fluorophore is active. Our approach employs the time-independent model as a seed, or initial condition, and we refine this solution through incremental move proposals. In particular, we propose a distribution from a factorized family, $q(\Theta) = q(\pi_0)q(\mu)q(\gamma)q(z)q(S)q(a)$, restricting our analysis to distributions q with the following dependence structure:

$$q(\Theta) = q(\pi_0)q(a) \prod_{k=1}^{K} q(S^k)q(a_i^k)q(\gamma_k)q(\mu_k) \prod_{t=1}^{T} \prod_{n_t=1}^{N_t} q(z_n^t), \tag{4.7}$$

where

$$
\begin{aligned}
&q\big(a_{01}^k\big) = \mathrm{Beta}(\tilde{\nu}_{k1}, \tilde{\nu}_{k2}), \qquad q(a_i) = \mathrm{Dir}\big(\tilde{\alpha}^k\big),\\
&q\big(\gamma^k\big) = \mathrm{Gamma}(\tilde{\eta}_{k1}, \tilde{\eta}_{k2}), \qquad q(\mu_k) = \mathrm{Normal}(\tilde{\mu}_k, \tilde{\sigma}_k),\\
&q(\pi_0) = \mathrm{Beta}\big(\tilde{\iota}_1^t, \tilde{\iota}_2^t\big), \qquad q\big(z_n^t\big) = \mathrm{Categorical}\big(r_{nk}^t\big).
\end{aligned} \tag{4.8}
$$

Finally, we fit the state-variable dynamics via a structured variational proposal with Markovian structure (Hughes, Kim and Sudderth (2015)),

$$
\begin{aligned}
q(S_k) &\sim \Bigg[\prod_{j=1}^{m} (\psi_{k0i})^{\delta S_k^0 = i}\Bigg]\Bigg[\prod_{t=1}^{T}\prod_{i=1}^{m}\prod_{j=1}^{m}\bigg(\frac{\phi_{ktij}}{\psi_{kti}}\bigg)^{\delta S_k^t = i, \delta S_k^{t-1} = j}\Bigg],\\
\psi_{kti} &= \sum_{j=1}^{m} \phi_{ktij},
\end{aligned} \tag{4.9}
$$

where the variational parameter ϕ_{ktij} represents the joint probability $\phi_{ktij} = q(s_k^{t+1=j}, s_k^{t=i})$ and ψ_{kti} defines the marginal probability $\psi_{kti} = q(s_k^{t=i})$.

4.4. *Inference*. We turn again to ELBO computation for the spatiotemporal model. In particular, we optimize the parameters of the fully factorized variational proposal via coordinate ascent (CAVI). We arrange the ELBO into three terms accounting for data generation, entropy and the KL divergence between our global parameter prior distributions and the corresponding variational proposals,

$$
\mathcal{L}(q) = \mathbb{E}_q\big[\log p(x|\Theta) + \log p(\Theta) - \log q(\Theta)\big] = \mathcal{L}(q)_{\mathrm{Data}} + \mathcal{H} + \mathcal{L}(q)_{\mathrm{Global}}, \tag{4.10}
$$

where

$$
\begin{aligned}
\mathcal{L}(x, \sigma, r, \tilde{\mu}, \tilde{\sigma}, \tilde{\eta}, \tilde{\iota})_{\mathrm{Data}} &= \mathbb{E}_q\big[\log p(x, \sigma, z|\gamma, z, \mu, S, \pi_0)\big],\\
\mathcal{H}(r) &= -\mathbb{E}_q\big[\log q(z)\big],\\
\mathcal{L}(\tilde{\mu}, \tilde{\sigma}, \tilde{\nu}, \tilde{\alpha}, \tilde{\eta}, \tilde{\iota})_{\mathrm{Global}} &= \mathbb{E}_q\bigg[\log\frac{p(\mu)}{q(\mu)} + \log\frac{p(\gamma)}{q(\gamma)} + \log\frac{p(a_{01})}{q(a_{01})}\\
&\qquad + \log\frac{p(\pi_0)}{q(\pi_0)} + \log\frac{p(S)}{q(S)}\bigg].
\end{aligned} \tag{4.11}
$$

We compute the ELBO for this new model and take partial derivatives with respect to each variational parameter to derive the coordinate ascent updates. Most of the updates are straightforward to compute, due to the conjugate exponential family factors. To optimize the local variables involved in the likelihood equation, we relax our generative model by introducing a variable $\epsilon \ll 1$,

$$
\begin{aligned}
P\big(x_n^t, \sigma_n^t, z_n^t|\gamma, s^t, \mu\big) &= \prod_{k=0}^{K} P\big(x_n^t, \sigma_n^t|\mu_{z_n^t}\big)^{[z_n^t = k][s_k^t = 1]} P\big(z_n^t|\gamma, s^t\big)^{[z_n^t = k][s_k^t]}\\
&= \bigg(\frac{1}{R}\pi_0\bigg)^{[z_n^t = 0]} \prod_{k=1}^{K} P\big(x_n^t, \sigma_n^t|\mu_{z_n^t}\big)^{[z_n^t = k][s_k^t = 1]}\\
&\quad * \Bigg[(1 - \pi_0)\frac{(\gamma^k)^{[s_k^t = 1]}\epsilon^{[1 - s_k^t = 1]}}{\sum_l \gamma^l[s_l^t = 1] + \epsilon[s_l^t = 0]}\Bigg]^{[z_n^t = k]}.
\end{aligned} \tag{4.12}
$$

This relaxation smooths the ELBO and helps convergence. Finally, as noted by Sun, Paisley and Liu (2017) in a similar setting, the gamma normalization term presents computational

difficulties, and we follow those authors in introducing an auxiliary variable ξ_t to further lower bound this term (Sun, Paisley and Liu (2017)),

$$
\begin{aligned}
&-\log\Big(\sum_l \gamma_l[s_l^t = 1] + \epsilon[1 - [s_l^t = 1]]\Big) \\
&\quad \geq -\log \xi_t - \frac{\sum_l \gamma_l[s_l^t = 1] + \epsilon[1 - [s_l^t = 1]] - \xi_t}{\xi_t}.
\end{aligned} \tag{4.13}
$$

To obtain the value of ξ_t, we differentiate the lower bound and set to zero which yields the following update rule:

$$
\xi_t = \sum_l \mathbb{E}_q\big(\gamma_l[s_l^t = 1] + \epsilon[1 - [s_l^t = 1]]\big). \tag{4.14}
$$

We reproduce the entire list of CAVI updates in the Supplementary Material (Gabitto et al. (2021)).

4.5. *Algorithmic work flow.* To summarize, the analysis of a particular data set, based on a given set of fluorophore localizations in space and time, proceeds as follows. First, we collapse observations in time and select an initialization. We refine this initialization through our time-independent model. We interleave birth-death, split-merge proposals to explore different configurations of the state space and prune existing clusters. The final configuration of our time-independent algorithm seeds our time-dependent model. Finally, clusters are refined by split-merge moves that help to separate time-overlapping clusters. We approximate the entire time evolution of each fluorophore state when splitting or merging fluorophores in this model. To efficiently propose split moves, we calculate the posterior number of fluorophores inside clusters with an unusually high number of observations and evaluate split proposals. The calculation of the posterior number of fluorophores, given blinking statistics, is detailed in the Supplementary Material (Gabitto et al. (2021)). Finally, at the end, for each fluorophore we compute the time evolution of each fluorescent trace.

5. Realistic simulation studies. In this section we present studies assessing the performance of our algorithm on data sets derived from the DNA origami platform and the nuclear pore complex. Both data sets were presented in Section 2.

5.1. *Data preprocessing and data set construction.* To create realistic SMLM data sets, we make use of the DNA origami platform introduced in Section 2. Raw images of every data set were preprocessed with Thunderstorm (Ovesnỳ et al. (2014)). Briefly, raw images were imported into FIJI (Schindelin et al. (2012)), and the Thunderstorm plugin was run with camera parameters and default approximate and subpixel molecule localization parameters. Next, observations with an unusual variance, uncertainty or intensity value (five standard deviations above or below the mean) were filtered out. We use these raw localizations to generate realistic data sets.

We isolated single fluorophores present in the data set in which fluorophores are attached to handle complementary sequences (Figure 3A). We verified that observations are localized with the reference TAMRA signal, that three sets of cloud points co-localized and that cloud distances were close to 84 nm (distance between handles 1–7 and 7–13). When these conditions were met, we then isolated individual clouds of points and considered each of them as an isolated fluorophore. For each extracted fluorophore, we computed its posterior fluorophore location, according to our algorithm, for just one fluorophore. This is the ground-truth fluorophore location against which we test our algorithm.

5.2. *Model hyperparameters*. For all subsequent computational experiments, we used weakly informative hyperpriors. We placed a Gamma(1, 0.01) prior on the concentration parameters α_0. The parameters of the base measure were set from the data, with μ_0 was chosen to be the center of the field of view and σ_0 the maximum distance between observations. To set the prior on π_0, we reasoned that, in a real experiment, most of the points arise from real fluorophores. Accordingly, we chose $a_0 = 1$, $b_0 = 100$. We chose the augmentation factor of the bounding box to be 1.25 times the average standard deviation of the points. This value seemed to be robust across the different data sets that we analyzed. For the time-dependent model, Γ_0 was given a Gamma(100, 1) prior. The prior on the transition matrix was given via pseudo-counts $\alpha_{on,on/\text{blink/dead}} \in \{1, 5, 10\}$ and $\alpha_{\text{blink},on/\text{blink}} \in \{1, 10\}$.

5.3. *Computational experiments*. In this section we present several different experimental scenarios that we employed to test the limits of our algorithm. In particular, we aimed to assess: 1) when our algorithm fails to distinguish nearby fluorophores, 2) how well our move proposals explore the ELBO, finding global optima and 3) how well our procedures scales with an increasing number of fluorophores in the field of view.

To achieve this goal, we simulated fluorophore observations by randomly sampling a DNA origami fluorophore and placing its observations in a ground-truth position. To contaminate data sets with noise, we randomly selected observations from any fluorophore and randomly positioned them in the field of view. We quantified the noise level by measuring the ratio of noise points over points that belong to a ground-truth fluorophore. By these means we were able to construct realistic simulated data sets with realistic ground-truth observations. We judged performance with reference to a variety of metrics: the algorithm's ability to correctly segment the data according to the underlying fluorophore location, the robustness to the algorithm to different number of fluorophores and noise level as well as fluorophore detection based on fluorophore proximity.

5.4. *Identifying nearby fluorophores*. We simulated two fluorophores at different distances and different noise levels and compared our algorithms (both the time-dependent and time-independent algorithms) to DBScan (Ester et al. (1996)) in Figure 7a. We explored the ELBO through different move proposals and returned the best configuration seen. In all cases, automatic DBScan settings explored through optics (Schubert et al. (2017)) failed to identify the correct number of fluorophores. As seen in the figure, our algorithm performed significantly better than DBScan. Our algorithm performed poorly when observations are nearer than two times the standard deviation of the observations, although the dispersion of the localizations of each fluorophore also played a role in the performance. This limit is extended through the use of our time-dependent formulation. This limit is an intrinsic property of our model, as revealed by the average ELBO gap between the true fluorophore configuration and one in which fluorophores are merged (see Figure 7b). The gap trace switches sign at the limit, indicating a preferred incorrect fluorophore configuration. Our algorithm exhibits performance above this limit that is robust to noise. Below this limit, even if data-driven proposals can identify more than one fluorophore in the cloud of points, their location cannot be correctly determined.

We explored robustness to different noise scenarios, revealing that performance is maintained even when the noise level reaches a value of two. This noise corruption means that, locally, two out of three points belonged to noise. Experimentally, it is highly unlikely to encounter such scenarios, and we did not observe it in our nuclear pore experiments. Finally, our time-dependent algorithm seems to be more accurate than our time-independent model in correctly localizing fluorophores' true positions (Figure 7c).

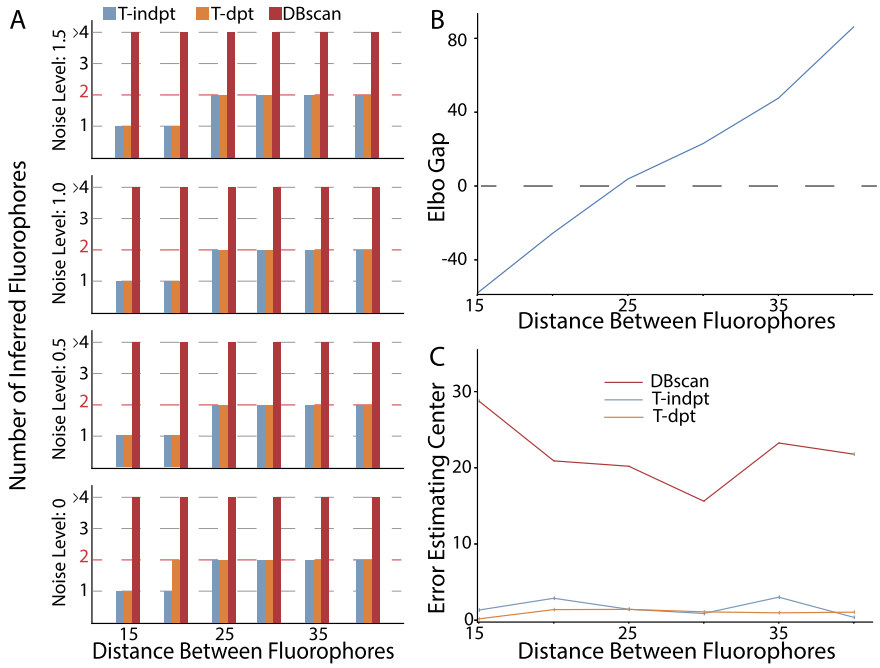

FIG. 7. ***Realistic simulated experiments, distinguishing two individual fluorophores:*** *a) Number of inferred fluorophores under different noise regimes, varying the distance between them. b) For a noise level of* 0.6, *ELBO gap computed as the difference between a model with two fluorophores and a model with one fluorophore. When the gap becomes negative, the model prefers the incorrect configuration of just one fluorophore. c) Error of inferred fluorophore positions.*

5.5. *Scaling fluorophore numbers.* Next, we simulated an increasing number of fluorophores at different distances, localizing them on a grid. We imposed a realistic level of noise of 0.5 (Figure 8a–b). We varied fluorophore distances while remaining above our identification limit. We then applied our algorithms to every condition, randomly simulating the conditions six times and computing averages of inferred fluorophore numbers. As seen in Figure 8c, our algorithm is robust to an increasing number of fluorophores as long as we increase the number of proposals explored. There was a mild decrease in fluorophore recovery when the distances approached the detection limit. As expected, true fluorophore localization degraded as fluorophores approached each other (see Figure 8d).

6. Application to nuclear pore complex data. Finally, we present results of applying our algorithms to the nuclear pore complex, a real biological data set of known structure. To focus on the NPC, a field of view needs to be preprocessed and NPCs isolated. To select several instances of the imaged NPCs, we proceeded by isolating a few candidates in the image and creating a searching template (Figure 9a). This template was cross-correlated against the entire image, and candidates' data sets were created from regions in which the correlation score exceeded a threshold (Figure 3a). These NPC datasets with localized emitters served as input to our analysis (examples of NPCs are reproduced in Figure 9b).

We applied our algorithms to 369 NPC datasets, running them for a fixed budget of iterations (although convergence was recognized well before exhausting the budget) and identifying underlying fluorophores (Figure 9c–d)). We were specifically interested in assessing the validity of our inferred fluorophores. Therefore, we made use of the aforementioned NPC symmetry and counted fluorophore distance to the closest neighbor. Given the NPC octagonal shape, we thereby obtained a minimal fluorophore distance of 41.5 nm, where the raw data minimal distance is close to three nm (as illustrated in Figure 9e–f)). Moreover, as seen

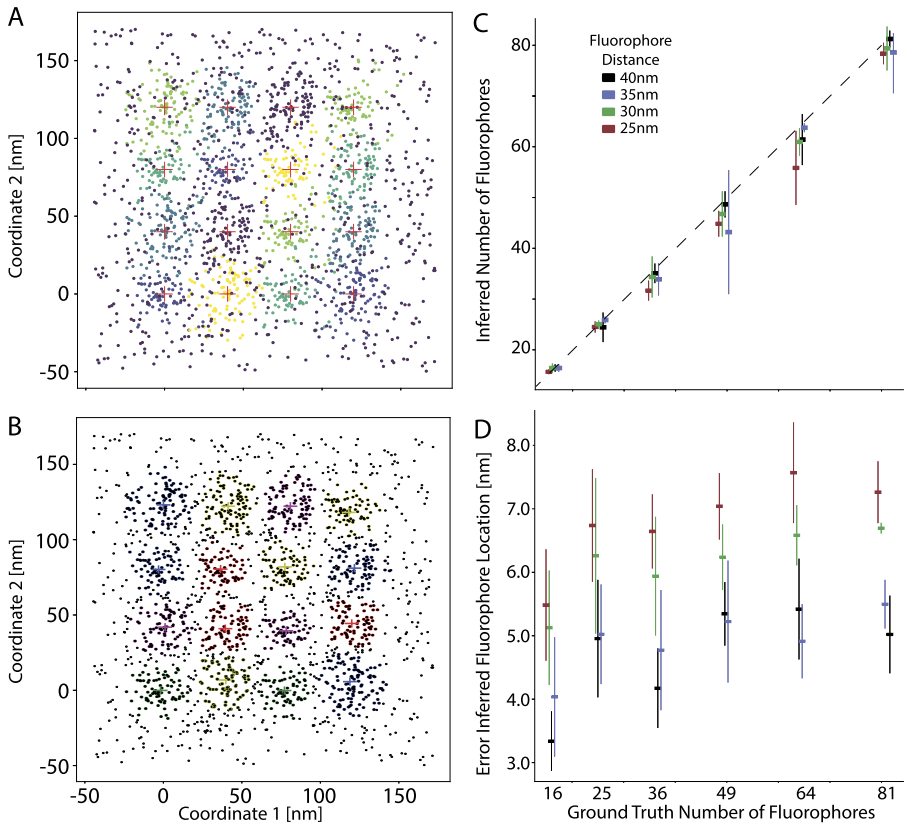

FIG. 8. ***Realistic simulated experiments, scaling the number of fluorophores:*** *a) True configuration for an example of* 16 *fluorophores in the field of view. b) For a configuration such as in a), inferred fluorophore positions. c) Number of inferred fluorophores when the ground truth number in the field of view is increased. d) Average error* $\pm$ *standard deviation for the inferred fluorophores' positions.*

in Figure 9g–h, the inferred time dependent parameters were in accordance with published results with the blinking probability exhibiting a clear exponential decay (Lee et al. (2012)).

7. Related work. There has been a great deal of previous work on the development to software to deconvolve super-resolution movies (Holden, Uphoff and Kapanidis (2011), Ovesnỳ et al. (2014), Sergé et al. (2008), Small and Stahlheber (2014)). Most of this work focuses on computationally efficient algorithms for detecting the Gaussian shape of the point spread functions without attempting to explicitly model latent temporal dynamics. Software that incorporates temporal information is extremely computationally costly and is prone to producing artifacts (Rosten, Jones and Cox (2013)). Recently, deep generative models have been used to identify fluorophores from SR images, taking into account different PSF shapes (Nehme et al. (2019), Speiser, Turaga and Macke (2015), Sun, Archer and Paninski (2017)). Some of these approaches are complementary to ours and of similar computational complexity; others are significantly more costly computationally. Our software builds upon fast single-frame deconvolution algorithms to incorporate temporal information into the localization analysis. Furthermore, we use spatially sensitive data structures to speed up calculations and facilitate scalability.

More broadly, identifying the number of different time series and assigning observations at each time point is a difficult task when the number of observations does not match the number of time series. This problem has been partially addressed by different authors. Several BNP approaches that capture time evolution—most based on the hierarchical Dirichlet

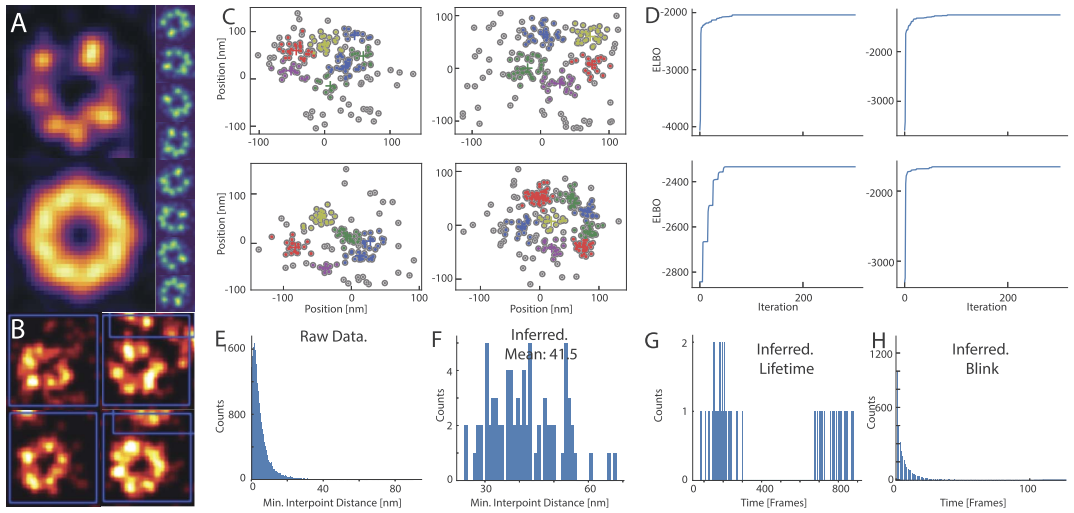

FIG. 9. ***Analysis of nuclear pore datasets*:** *a) Top left, the image of a nuclear pore convolved with a Gaussian filter. Bottom left, the nuclear pore complex is rotated eight times, and these rotated images are aggregated to generate a nuclear pore complex template. Right, an eight-fold rotation of the nuclear pore complex. b) Nuclear pore complexes isolated through template matching. c) Localizations extracted from nuclear pore complex data sets. Inferred fluorophore centers highlighted with a cross in different colors. d) ELBO evolution during our inference procedure for the four datasets shown in c). e) A histogram depicting the distance for each point to its closest neighbor. f) Same as in c) but this time for each inferred fluorophore location. g) Inferred fluorophore durations. h) Inferred time interval between subsequent observations of each fluorophore.*

process (HDP) (Teh et al. (2006))—have been developed in the setting of topic models. The HDP assumes that the probability of topics and the proportion of words explained within each document are coupled. This is undesirable, however, when there are rare topics explained by a large proportion of words in a small number of documents. A similar problem has been encountered in sparse topic modeling (Archambeau, Lakshminarayanan and Bouchard (2014), Faisal et al. (2012), Williamson et al. (2010)), where it is important to distinguish the probability that a topic belongs to a document from the probability of inclusion of the topic into the analysis. To address this issue, the authors proposed the use of a compound Indian buffet process-Dirichlet process. However, their work did not consider the dependency of the features across time and is limited by the sampling scheme developed.

The use of Bayesian nonparametric feature models for the modeling of time series was initiated by Fox et al. (2009). Their work focused on motion capture data, a domain without the complexities and the scale of single-molecule imaging domain that is our focus. There has been significant follow-up work in this vein, including the use of a time-dependent beta process as a Bayesian nonparametric prior for feature allocation models (Perrone et al. (2017)). Again, however, the focus has been on small-scale problems, and the methods are not directly applicable to the single-molecule imaging problem.

8. Discussion. We have presented a Bayesian nonparametric method for the identification of fluorescent molecules in super-resolution experiments. To obtain a procedure that is viable at the scale of realistic experiements, we developed a statistical methodology that proceeds in two phases. The first phase is based on a model that analyzes localization microscopy observations by collapsing temporal information. This model relies on the Dirichlet process as a prior on the underlying number of fluorophores present in the sample. To speed calculation, we used spatial data structures (quadtrees) to obtain individual fluorophore assignments. Inference in this model is performed using mean field variational inference, and the feature space is explored using state space adaptation techniques. Next, we developed a statistical

model that incorporates temporal information into the analysis and accounts for fluorophore photophysics. Taking the infinite limit of this model defines a nonparametric prior that is comprised of an infinite factorial hidden Markov model and a dependent Gamma process. This prior allows the assignment of different probabilities to the inclusion of a fluorophore at each time point and determines the probability of assigning an observation to each active fluorophore. To refine the inferred fluorophore numbers, we incorporate data-driven split-merge moves that split fluorophores based on fluorophore blinking statistics.

We demonstrated the utility of our model using realistic simulated data and a real data set in which the underlying biological structure is known. In this real data set we were able to correctly infer fluorophore localization consistent with the geometry of the sample, as demonstrated by the interfluorophore position distributions. By using realistic simulated data, we showed that our model is robust to noise conditions encountered in real experiments. We expect our method to perform poorly in cases where fluorescent molecules are out of focus which results in an inferred position that differs greatly from the true location.

A key feature of our algorithm is that it can be used as a postprocessor for any software pipeline that extracts raw fluorophore localizations from data. Our approach aims to integrate temporal information into the analysis and correct mistakes produced during single-molecule identification. We illustrated our method using the external software package Thunderstorm to process our images. In practice, there is evidence that such localization software produces mistakes when two nearby fluorescent molecules are active in the same frame due to a failure to correctly infer molecule locations. Future work could consider alternative ways of preprocessing imaging datasets. In this case, to further improve localization accuracy, it might be desirable to incorporate modeling of the point spread function into the model and to directly process raw images within a more sophisticated statistical framework.

Software. Our method is implemented in python with a C++ kernel, available on github at https://github.com/marianogabitto/Lightning and in the Supplementary Material (Gabitto et al. (2021)).

Acknowledgments. We thank Melike Lakadamyali, Francesca Cella, Jonas Reis and Yiming Li for providing published data sets to validate our methods.

Funding. The third author was supported by ONR N00014-17-1-2843, NSF NeuroNex Award DBI-1707398 and The Gatsby Charitable Foundation.

We also wish to acknowledge support from the Army Research Office under contract/grant number W911NF-16-1-0368.

SUPPLEMENTARY MATERIAL

Supplement to "A Bayesian nonparametric approach to super-resolution single molecule localization" (DOI: 10.1214/21-AOAS1441SUPPA; .pdf). We provide additional material to support the results in this paper. This Supplementary Material includes detailed derivations of parameter update rules, cluster refinement procedures and possible extensions of our model to different time distributions.

Source code to "A Bayesian nonparametric approach to super-resolution single molecule localization" (DOI: 10.1214/21-AOAS1441SUPPB; .zip). We provide python code and a C++ kernel for the time independent model.

REFERENCES

ABBE, E. (1873). Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung. *Arch. Mikrosk. Anat.* **9** 413–418.

AIRY, G. B. (1835). On the diffraction of an object-glass with circular aperture. *Trans. of the Cambridge Philosoph. Soc.* **5** 283–291.

ANNIBALE, P., SCARSELLI, M., KODIYAN, A. and RADENOVIC, A. (2010). Photoactivatable fluorescent protein mEos2 displays repeated photoactivation after a long-lived dark state in the red photoconverted form. *J. Phys. Chem. Lett.* **1** 1506–1510.

ANNIBALE, P., VANNI, S., SCARSELLI, M., ROTHLISBERGER, U. and RADENOVIC, A. (2011a). Identification of clustering artifacts in photoactivated localization microscopy. *Nat. Methods* **8** 527–528.

ANNIBALE, P., VANNI, S., SCARSELLI, M., ROTHLISBERGER, U. and RADENOVIC, A. (2011). Quantitative photo activated localization microscopy: Unraveling the effects of photoblinking. *PLoS ONE* **6** e22678. https://doi.org/10.1371/journal.pone.0022678

ARCHAMBEAU, C., LAKSHMINARAYANAN, B. and BOUCHARD, G. (2014). Latent IBP compound Dirichlet allocation. *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 321–333.

BECK, M. and HURT, E. (2017). The nuclear pore complex: Understanding its function through structural insight. *Nat. Rev. Mol. Cel. Biol.* **18**.

BETZIG, E., PATTERSON, G. H., SOUGRAT, R., LINDWASSER, O. W., OLENYCH, S., BONIFACINO, J. S., DAVIDSON, M. W., LIPPINCOTT-SCHWARTZ, J. and HESS, H. F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313** 1642–1645.

BLEI, D. M. and JORDAN, M. I. (2006). Variational inference for Dirichlet process mixtures. *Bayesian Anal.* **1** 121–143. MR2227367 https://doi.org/10.1214/06-BA104

BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *J. Amer. Statist. Assoc.* **112** 859–877. MR3671776 https://doi.org/10.1080/01621459.2017.1285773

BRAUN, M. and MCAULIFFE, J. (2010). Variational inference for large-scale models of discrete choice. *J. Amer. Statist. Assoc.* **105** 324–335. MR2757203 https://doi.org/10.1198/jasa.2009.tm08030

BRODERICK, T., JORDAN, M. I. and PITMAN, J. (2013). Cluster and feature modeling from combinatorial stochastic processes. *Statist. Sci.* **28** 289–312. MR3135534 https://doi.org/10.1214/13-STS434

D'ANGELO, M. A. and HETZER, M. W. (2008). Structure, dynamics and function of nuclear pore complexes. *Trends Cell Biol.* **18** 456–466.

DEMPSEY, G. T., VAUGHAN, J. C., CHEN, K. H., BATES, M. and ZHUANG, X. (2011). Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nat. Methods* **8** 1027.

DESCHOUT, H. . (2014). Precisely and accurately localizing single emitters in fluorescence microscopy. *Nat. Methods* **11** 253–266.

ESCOBAR, M. D. and WEST, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90** 577–588. MR1340510

ESTER, M., KRIEGEL, H.-P., SANDER, J. and XU, X. (1996). A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. KDD'*96 226–231. AAAI Press, Menlo Park.

EWENS, W. J. (1990). Population genetics theory—the past and the future. In *Mathematical and Statistical Developments of Evolutionary Theory* (*Montreal*, *PQ*, 1987). *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.* **299** 177–227. Kluwer Academic, Dordrecht. MR1108002

FAISAL, A., GILLBERG, J., PELTONEN, J., LEEN, G. and KASKI, S. (2012). Sparse nonparametric topic model for transfer learning. In *European Symposium on Artificial Neural Networks*.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230. MR0350949

FINKEL, R. A. and BENTLEY, J. L. (1974). Quad trees a data structure for retrieval on composite keys. *Acta Inform.* **4** 1–9.

FOTI, N., FUTOMA, J., ROCKMORE, D. and WILLIAMSON, S. (2013). A unifying representation for a class of dependent random measures. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research* **31**.

FOX, E., JORDAN, M. I., SUDDERTH, E. B. and WILLSKY, A. S. (2009). Sharing features among dynamical systems with beta processes. In *Advances in Neural Information Processing Systems* 549–557.

GABITTO, M. I., MARIE-NELLY, H., PAKMAN, A., PATAKI, A., DARZACQ, X. and JORDAN, M. I. (2021). Supplement to "A Bayesian nonparametric approach to super-resolution single-molecule localization." https://doi.org/10.1214/21-AOAS1441SUPPA, https://doi.org/10.1214/21-AOAS1441SUPPB

GAEL, J. V., TEH, Y. W. and GHAHRAMANI, Z. (2009). The infinite factorial hidden Markov model. In *Advances in Neural Information Processing Systems* 1697–1704.

GELFAND, A. E. and SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409. MR1141740

HANSEN, A. S., CATTOGLIO, C., DARZACQ, X. and TJIAN, R. (2018). Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* **9** 20–32. https://doi.org/10.1080/19491034.2017.1389365

HEILEMANN, M., MARGEAT, E., KASPER, R., SAUER, M. and TINNEFELD, P. (2005). Carbocyanine dyes as efficient reversible single-molecule optical switch. *J. Am. Chem. Soc.* **127** 3801–3806.

HEILEMANN, M., DEDECKER, P., HOFKENS, J. and SAUER, M. (2009). Photoswitches: Key molecules for subdiffraction-resolution fluorescence imaging and molecular quantification. *Laser Photonics Rev.* **3** 180–202.

HOLDEN, S. J., UPHOFF, S. and KAPANIDIS, A. N. (2011). DAOSTORM: An algorithm for high-density super-resolution microscopy. *Nat. Methods* **8** 279–280. https://doi.org/10.1038/nmeth0411-279

HUANG, Z. L. . (2011). Localization-based super-resolution microscopy with an sCMOS camera. *Opt. Express* **19** 19156–19168.

HUGGINS, J. H. and WOOD, F. (2014). Infinite structured hidden semi-Markov models. Preprint. Available at arXiv:1407.0044.

HUGHES, M., KIM, D. I. and SUDDERTH, E. (2015). Reliable and scalable variational inference for the hierarchical Dirichlet process. In *Artificial Intelligence and Statistics* 370–378.

HUMMER, G., FRICKE, F. and HEILEMANN, M. (2016). Model-independent counting of molecules in single-molecule localization microscopy. *Mol. Biol. Cell* **27** 3637–3644. https://doi.org/10.1091/mbc.E16-07-0525

ICKSTADT, K. and WOLPERT, R. L. (1999). Spatial regression for marked point processes. In *Bayesian Statistics*, 6 (*Alcoceber*, 1998) 323–341. Oxford Univ. Press, New York. MR1723503

JORDAN, M. I., GHARAMANI, Z., JAAKKOLA, T. and SAUL, L. (1999). An introduction to variational methods for graphical models. *Mach. Learn.* **37** 183–233.

KAPLAN, C. and EWERS, H. (2015). Optimized sample preparation for single-molecule localization-based superresolution microscopy in yeast. *Nat. Protoc.* **10** 1007–1021.

KAPOOR-KAUSHIK, N. . (2016). Distinct mechanisms regulate Lck spatial organization in activated T cells. *Front. Immunol.* **7** 83.

KIM, S. J., FERNANDEZ-MARTINEZ, J., NUDELMAN, I., SHI, Y., ZHANG, W., RAVEH, B., HERRICKS, T., SLAUGHTER, B. D., HOGAN, J. A. et al. (2018). Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555** 475.

KOŠUTA, T., CULLELL-DALMAU, M., ZANACCHI, F. C. and MANZO, C. (2020). Bayesian analysis of data from segmented super-resolution images for quantifying protein clustering. *Phys. Chem. Chem. Phys.* **22** 1107–1114. https://doi.org/10.1039/c9cp05616e

KOTTAS, A. and SANSÓ, B. (2007). Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis. *J. Statist. Plann. Inference* **137** 3151–3163. MR2365118 https://doi.org/10.1016/j.jspi.2006.05.022

LEE, S.-H., SHIN, J. Y., LEE, A. and BUSTAMANTE, C. (2012). Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM). *Proc. Natl. Acad. Sci. USA* **109** 17436–17441.

LI, Y., MUND, M., HOESS, P., DESCHAMPS, J., MATTI, U., NIJMEIJER, B., SABININA, V. J., ELLENBERG, J., SCHOEN, I. et al. (2018). Real-time 3D single-molecule localization using experimental point spread functions. *Nat. Methods* **15** 367–369. https://doi.org/10.1038/nmeth.4661

LIPPINCOTT-SCHWARTZ, J. and PATTERSON, G. H. (2009). Photoactivatable fluorescent proteins for diffraction-limited and super-resolution imaging. *Trends Cell Biol.* **19** 555–565.

NEAL, R. M. (1992). Bayesian mixture modeling. In *Proceedings of the Workshop on Maximum Entropy and Bayesian Methods of Statistical Analysis* **11** 197–211.

NEHME, E., FREEDMAN, D., GORDON, R., FERDMAN, B., WEISS, L. E., ALALOUF, O., ORANGE, R. and MICHAELI, T. (2019). DeepSTORM3D: Dense three dimensional localization microscopy and point spread function design by deep learning. *Int. J. Image Video Process.*

NICOVICH, P. R., OWEN, D. M. and GAUS, K. (2017). Turning single-molecule localization microscopy into a quantitative bioanalytical tool. *Nat. Protoc.* **12** 453–460. https://doi.org/10.1038/nprot.2016.166

NINO, D., RAFIEI, N., WANG, Y., ZILMAN, A. and MILSTEIN, J. N. (2017). Molecular counting with localization microscopy: A Bayesian estimate based on fluorophore statistics. *Biophys. J.* **112** 1777–1785.

OLIVIER, N., KELLER, D., GÖNCZY, P. and MANLEY, S. (2011). Resolution doubling in 3D-STORM imaging through improved buffers. *PLoS ONE* **8** e69004.

OVESNÝ, M., KŘÍŽEK, P., BORKOVEC, J., ŠVINDRYCH, Z. and HAGEN, G. M. (2014). ThunderSTORM: A comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics* **30** 2389–2390.

OWEN, D. M. . (2010). PALM imaging and cluster analysis of protein heterogeneity at the cell surface. *J. Biophotonics* **3** 446–454.

PERRONE, V., JENKINS, P. A., SPANÒ, D. and TEH, Y. W. (2017). Poisson random fields for dynamic feature models. *J. Mach. Learn. Res.* **18** Paper No. 127, 45. MR3763761

PERTSINIDIS, A., ZHANG, Y. and CHU, S. (2010). Subnanometre single-molecule localization, registration and distance measurements. *Nature* **466** 647–651.

PUCHNER, E. M., WALTER, J. M., KASPER, R., HUANG, B. and LIM, W. A. (2013). Counting molecules in single organelles with superresolution microscopy allows tracking of the endosome maturation trajectory. *Proc. Natl. Acad. Sci. USA* **110** 16015–16020.

RASMUSSEN, C. E. (2000). The infinite Gaussian mixture model. *Adv. Neural Inf. Process. Syst.* **12**.

REGIER, J., MILLER, A. C., SCHLEGEL, D., ADAMS, R. P., MCAULIFFE, J. D. and PRABHAT (2019). Approximate inference for constructing astronomical catalogs from images. *Ann. Appl. Stat.* **13** 1884–1926. MR4019161 https://doi.org/10.1214/19-AOAS1258

ROLLINS, G. C., SHIN, J. Y., BUSTAMANTE, C. and PRESSÉ, S. (2015). Stochastic approach to the molecular counting problem in superresolution microscopy. *Proc. Natl. Acad. Sci. USA* **112** E110–E118.

ROSSY, J., COHEN, E., GAUS, K. and OWEN, D. M. (2014). Method for co-cluster analysis in multichannel single-molecule localisation data. *Histochem. Cell Biol.* **141** 605–612.

ROSTEN, E., JONES, G. E. and COX, S. (2013). ImageJ plug-in for Bayesian analysis of blinking and bleaching. *Nat. Methods* **10** 97.

ROY, A., FIELD, M. J., ADAM, V. and BOURGEOIS, D. (2011). The nature of transient dark states in a photoactivatable fluorescent protein. *J. Am. Chem. Soc.* **133** 18586–18589.

RUBIN-DELANCHY, P., BURN, G. L., GRIFFIÉ, J., WILLIAMSON, D. J., HEARD, N. A., COPE, A. P. and OWEN, D. M. (2015). Bayesian cluster identification in single-molecule localization microscopy data. *Nat. Methods* **12** 1072–1076. https://doi.org/10.1038/nmeth.3612

RUST, M. J., BATES, M. and ZHUANG, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3** 793–795. https://doi.org/10.1038/nmeth929

SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E., KAYNIG, V., LONGAIR, M., PIETZSCH, T., PREIBISCH, S., RUEDEN, C., SAALFELD, S. et al. (2012). Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9** 676.

SCHUBERT, E., SANDER, J., ESTER, M., KRIEGEL, H.-P. and XU, X. (2017). DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Trans. Database Syst.* **42** Art. 19, 21. MR3693646 https://doi.org/10.1145/3068335

SENGUPTA, P. . (2011). Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis. *Nat. Methods* **8** 969–975.

SERGÉ, A., BERTAUX, N., RIGNEAULT, H. and MARGUET, D. (2008). Multiple-target tracing (MTT) algorithm probes molecular dynamics at cell surface. Protocol Exchange, https://doi.org/10.1038/nprot.2008.128.

SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4** 639–650. MR1309433

SHCHERBAKOVA, D. M., SENGUPTA, P., LIPPINCOTT-SCHWARTZ, J. and VERKHUSHA, V. V. (2014). Photocontrollable fluorescent proteins for superresolution imaging. *Annu. Rev. Biophys.* **43** 303–329.

SMALL, A. and STAHLHEBER, S. (2014). Fluorophore localization algorithms for super-resolution microscopy. *Nat. Methods* **11** 267–279. https://doi.org/10.1038/nmeth.2844

SPECHT, C. G., IZEDDIN, I., RODRIGUEZ, P. C., EL BEHEIRY, M., ROSTAING, P., DARZACQ, X., DAHAN, M. and TRILLER, A. (2013). Quantitative nanoscopy of inhibitory synapses: Counting gephyrin molecules and receptor binding sites. *Neuron* **79** 308–321.

SPEISER, A., TURAGA, S. C. and MACKE, J. H. (2019). Teaching deep neural networks to localize sources in super-resolution microscopy by combining simulation-based learning and unsupervised learning. Available at arXiv:abs/1907.00770.

SUN, R., ARCHER, E. and PANINSKI, L. (2017). Scalable variational inference for super resolution microscopy. In *Proceedings of the* 20*th International Conference on Artificial Intelligence and Statistics* 1057–1065.

SUN, S., PAISLEY, J. and LIU, Q. (2017). Location dependent Dirichlet processes. In *International Conference on Intelligent Science and Big Data Engineering* 64–76. Springer, Berlin.

SZYMBORSKA, A., DE MARCO, A., DAIGLE, N., CORDES, V. C., BRIGGS, J. A. and ELLENBERG, J. (2013). Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. *Science* **341** 655–658.

TADDY, M. A. and KOTTAS, A. (2012). Mixture modeling for marked Poisson processes. *Bayesian Anal.* **7** 335–361. MR2934954 https://doi.org/10.1214/12-BA711

TEH, Y. W., JORDAN, M. I., BEAL, M. J. and BLEI, D. M. (2006). Hierarchical Dirichlet processes. *J. Amer. Statist. Assoc.* **101** 1566–1581. MR2279480 https://doi.org/10.1198/016214506000000302

TURNER, R. E. and SAHANI, M. (2011). Two problems with variational expectation maximisation for time series models. In *Bayesian Time Series Models* 104–124. Cambridge Univ. Press, Cambridge. MR2894235

VALERA, I., RUIZ, F. J. and PEREZ-CRUZ, F. (2015). Infinite factorial unbounded-state hidden Markov model. *IEEE Trans. Pattern Anal. Mach. Intell.* **38** 1816–1828.

VAN DE LINDE, S., HEILEMANN, M. and SAUER, M. (2012). Live-cell super-resolution imaging with synthetic fluorophores. *Annu. Rev. Phys. Chem.* **63** 519–540.

VEATCH, S. L. . (2012). Correlation functions quantify super-resolution images and estimate apparent clustering due to over-counting. *PLoS ONE* **7** e31457.

WAINWRIGHT, M. J. and JORDAN, M. I. (2008). Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.* **1** 1–305.

WILLIAMSON, S., WANG, C., HELLER, K. A. and BLEI, D. M. (2010). The IBP compound Dirichlet process and its application to focused topic modeling. In *Proceedings of the* 27*th International Conference on Machine Learning* (*ICML*-10) 1151–1158. Citeseer.

XU, K., ZHONG, G. and ZHUANG, X. (2013). Actin, spectrin, and associated proteins form a periodic cytoskeletal structure in axons. *Science* **339** 452–456.

ZANACCHI, F. C., MANZO, C., ALVAREZ, A. S., DERR, N. D., GARCIA-PARAJO, M. F. and LAKADAMYALI, M. (2017). A DNA origami platform for quantifying protein copy number in super-resolution. *Nat. Methods* **14** 789–792. https://doi.org/10.1038/nmeth.4342