Statistical Science
2020, Vol. 35, No. 1, 92–111
https://doi.org/10.1214/19-STS753

Statistical Molecule Counting in Super-Resolution Fluorescence Microscopy: Towards Quantitative Nanoscopy

Thomas Staudt, Timo Aspelmeier, Oskar Laitenberger, Claudia Geisler, Alexander Egner and Axel Munk

Abstract. Super-resolution microscopy is rapidly gaining importance as an analytical tool in the life sciences. A compelling feature is the ability to label biological units of interest with fluorescent markers in (living) cells and to observe them with considerably higher resolution than conventional microscopy permits. The images obtained this way, however, lack an absolute intensity scale in terms of numbers of fluorophores observed. In this article, we discuss state of the art methods to count such fluorophores and statistical challenges that come along with it. In particular, we suggest a modeling scheme for time series generated by single-marker-switching (SMS) microscopy that makes it possible to quantify the number of markers in a statistically meaningful manner from the raw data. To this end, we model the entire process of photon generation in the fluorophore, their passage through the microscope, detection and photoelectron amplification in the camera, and extraction of time series from the microscopic images. At the heart of these modeling steps is a careful description of the fluorophore dynamics by a novel hidden Markov model that operates on two timescales (HTMM). Besides the fluorophore number, information about the kinetic transition rates of the fluorophore's internal states is also inferred during estimation. We comment on computational issues that arise when applying our model to simulated or measured fluorescence traces and illustrate our methodology on simulated data.

Key words and phrases: Molecule counting, super-resolution microscopy, quantitative nanoscopy, biophysics and computational biology, inhomogeneous hidden Markov models, statistical thinning.

1. INTRODUCTION

During the past decades cell biology has undergone a profound transition, shifting its character from qualitative work about basic cell activity to increasingly quantitative methods to study fine details like the role of individual proteins for signaling and transport. This trend was crucially supported by the advancement of super-resolution microscopy (nanoscopy) techniques, highlighted by the 2014 Nobel prize in chemistry, which have since become an indispensable tool for modern biomedical research [6, 21, 23, 49]. While previous imaging methods for cellular structures were either limited due to a lack of resolution

Thomas Staudt is a PhD student, Institute for Mathematical Stochastics, Georg-August University of Göttingen, Goldschmidtstraße 7, 37077 Göttingen, Germany (e-mail: thomas.staudt@uni-goettingen.de). Timo Aspelmeier is Assistant Professor, Institute for Mathematical Stochastics, Georg-August University of Göttingen, Goldschmidtstraße 7, 37077 Göttingen, Germany (e-mail: timo.aspelmeier@mathematik.uni-goettingen.de). Oskar Laitenberger is a researcher, Laser-Laboratorium Göttingen e.V., Hans-Adolf-Krebs-Weg 1, 37077 Göttingen, Germany (e-mail: oskar.laitenberger@llg-ev.de). Claudia Geisler is a researcher, Laser-Laboratorium Göttingen e.V., Hans-Adolf-Krebs-Weg 1, 37077 Göttingen, Germany (e-mail: claudia.geisler@llg-ev.de). Alexander Egner is Director, Laser-Laboratorium Göttingen e.V., Hans-Adolf-Krebs-Weg 1, 37077 Göttingen, Germany (e-mail: alexander.egner@llg-ev.de). Axel Munk is Felix-Bernstein Professor of Statistics, Institute for Mathematical Stochastics, Georg-August University of Göttingen, Goldschmidtstraße 7, 37077 Göttingen, Germany (e-mail: munk@math.uni-goettingen.de).

(like conventional light microscopy) or due to their invasiveness (like X-ray or electron microscopy), fluorescence nanoscopy enables high-resolution imaging of living cells to the nanometer scale without the necessity to prepare samples in ways that prohibit natural biochemical activity. The limits of super-resolution microscopy, both in principle and application, are still being explored as progress unfolds at a remarkable pace [3, 22].

By now, many initial hurdles for the usage of fluorescence nanoscopy in various disciplines, like physiology, biology, and medicine, have been overcome: structures within living prokaryotic and eukaryotic cells are probed on unprecedented spatial scales in experiments [2, 31], and popular model organisms like fruit flies and mice are studied in vivo [5, 46]. It is hard to overstate the practical implications of bringing improved imaging resolution to these fields. From virology [10, 36, 40], immunology [38, 54] and neurology [11, 34] to cancer [9, 47] and plant biology [29], new ground in fundamental research is increasingly broken by means of nanoscopy. We exemplarily refer to [44] for an in-depth review about the unfolding role of super-resolution microscopy in cell biology.

The advancement of nanoscopy does not only raise new opportunities for experimentalists and lab scientists but also for statisticians. They are called to address a series of challenges that are highly relevant for exploiting the full potential of state-of-the-art fluorescence microscopy schemes (see [1], where further background on the underlying optics and the physical modeling of nanoscopy is given from a statistical perspective). Indeed, all current implementations of super-resolution microscopy are affected by the inherently stochastic behavior of fluorescent molecules, or fluorophores, which (randomly) emit photons if struck by incident light. In modes of nanoscopy that operate in a coordinate-targeted way (scanning observation points), like STED [21, 23] and RESOLFT [8, 18, 26], this stochasticity often plays a secondary role. Still, for quantitative analysis of the images (i.e., counting the actual number of fluorophores), the photon emission statistics turns out to be central. For example, individual fluorophores can be identified in STED nanoscopy by measuring the simultaneous arrival of emitted photons [28, 50]. In case of RESOLFT, an on-off Markov model for fluorophores has recently been demonstrated to be capable of extracting the contribution of single fluorophores in the total signal [16]. Other methodologies, like MINFLUX [3, 13], rely on a statistical treatment by design. In MINFLUX—fluorescence nanoscopy via minimal photon fluxes—a doughnut shaped laser intensity profile targeted to different spots on a biological sample is used to excite a fluorophore with unknown position. Based on the (approximately Poisson distributed) number of photons measured as response for each position of the excitation spot, the location of the fluorophore is inferred statistically, for example, via maximum likelihood estimation. Questions regarding the optimal measurement design—where to place the spots and which laser profile to use—naturally fit a Bayesian perspective and are still open for investigation.

The major focus of this article, however, will be another family of nanoscopy schemes, which exploit the fact that fluorophores have a tendency to blink over time, meaning that they (randomly) switch between active and inactive states. Under suitable conditions, fluorophores can thus be observed and localized individually even when clustered together. Methods that make use of this switching property are collected under the umbrella term single-marker-switching (SMS) nanoscopy, and include PALM [6], STORM [43], GSDIM [15], or variations thereof [12, 24, 52]. SMS nanoscopy works by recording a series of diffraction limited fluorescence images (or frames) in which only a small number of fluorophores is active and emits photons during the respective exposure. Spatially close molecules are therefore likely to be separated in time. As illustrated in Figure 1, the detected fluorophore positions from all frames can be used to create a pointillistic image with superior resolution. For a video that compares conventional fluorescence microscopy and SMS nanoscoy of a Rhodamine labeled microtubular network on the basis of 30,000 frames of experimental data, see stochastik.math.uni-goettingen.de/SMSData.

The stochastic nature of the frames recorded during SMS microscopy opens up a rich and fruitful field for statistical investigation. Indeed, major issues like the correction of spatial drifts in the image sequence, which originally required experimental intervention via so-called fiducial markers, have recently been tackled by fully statistical means—see [20] and the references therein. Another emergent topic is the temporal statistical modeling of the fluorophore dynamics, which also plays a crucial role for the present article. The most prominent approaches in this context are (hidden) Markov models, see [27, 35, 42, 51]. Recently, in [39], the photo switching behavior of fluorophores was characterized by a specifically tailored time-homogeneous hidden Markov model that reliably improves the estimation of kinetic transition rates from SMS data when compared to more basic methods, like exponential fitting of the dwell-times of the fluorophore in active/inactive states [33].

Moreover, Markov models allow for the refined extraction of quantitative information from SMS images, like counting the number of individual fluorophores in given image regions. This task of "quantitative nanoscopy" turns out to be much more involved than it appears at a first glance. The difficulty is that each fluorophore leaves a sophisticated intensity trace on the recorded image series, as it only causes a visible spot during frames in which the fluorophore is active—else it is invisible. Consequently, when a spot on the microscopic frames is lit up several

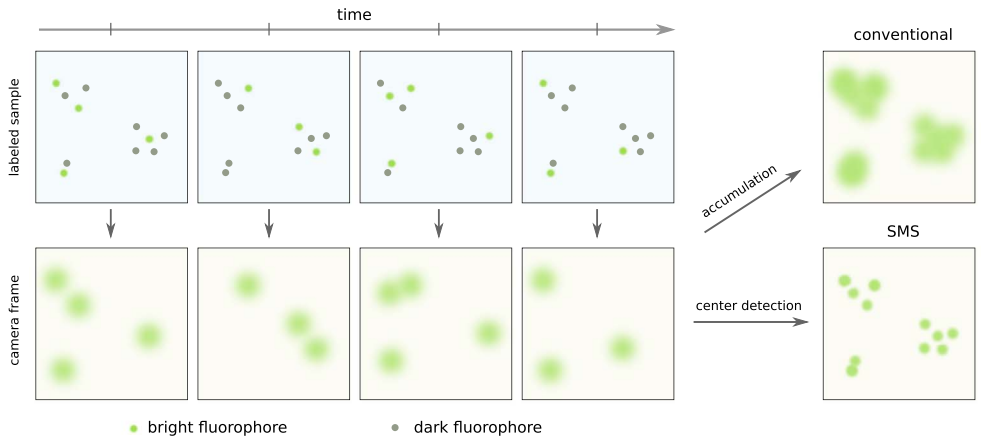

FIG. 1. *Principle of single-marker-switching microscopy. By exciting a biological sample that is labeled with fluorophores (top row) via a suitable laser, a temporal series of frames capturing fluorescent activity is recorded (bottom row). In each frame, only a sparse selection of fluorophores emits photons (green circles). The recorded images are blurry because of inevitable diffraction effects. Still, the center positions of the individual diffraction limited spots can be determined with higher precision due to spatial sparsity. This can be used to create a pointillistic nanoscopy image with superior resolution as compared to conventional fluorescence microscopy, where the photons emitted by all fluorophores would be recorded at the same time.*

times consecutively, it is not evident how many (close-by) fluorophores are responsible for the observed intensity pattern. This problem of mapping fluorescence intensity traces to the number of contributing molecules is of high practical relevance and poses a fundamental challenge for the application of SMS microscopy in quantitative biology.

In recent years, several methods to obtain such fluorophore numbers from fluorescence images have been proposed [27, 32, 42, 51]. They usually rely on the detection of switching events or on counting the number of steps during photobleaching (i.e., a fluorophore becoming irreversibly inactive). While these methods have been successfully applied to count 50 fluorophores and more in specific circumstances [32, 51], they can be prone to errors when misidentifying switching events or bleaching steps. These issues are particularly detrimental in the presence of many fluorophores within a diffraction limited region or when the fluorophore kinetics of bleaching and switching are fast in comparison to the image acquisition rate.

In this article, we lay the statistical foundations for a new method to estimate the number of fluorophores on SMS nanoscopy images introduced in [30]. Contrary to established methods, no step identification—which usually involves the choice of fluorescence levels or rate thresholds and depends on bleaching or switching—is necessary. This becomes possible by the careful statistical modeling and analysis of the whole imaging process: from photon generation in the fluorophore to signal amplification in the CCD camera. Our approach makes use of the full history of the recorded intensity information and exploits temporal correlations in the signal. The core component of the model is an accurate description of the fluorophore behavior in terms of a novel hidden Markov model that operates on two distinct timescales. It separates the fast dynamics that govern the emission of single photons during the exposure from the slow dynamics that describe fluorophore kinetics for states with dwell times longer than the exposure for a single frame. Although our Markov model is time-inhomogeneous, estimation of the fluorophore number and other kinetic parameters can be performed by applying the maximum likelihood principle to a simplified expression of the model's total likelihood. This simplification is based on a second-order approximation to the true likelihood and is derived by exploiting spectral properties of the model. Intriguingly, the inference takes place in an unusual setting: the quantity we want to estimate—the fluorophore number—is a feature of the initial state of the model and is lost in the asymptotic behavior for long times due to bleaching. In [30], the method has been experimentally verified on super-resolution images of DNA origami structures. This will be complemented by simulation results in the present work.

The article is organized as follows. In Section 2, we provide an overview of the single modeling steps that contribute to our total model for the fluorescent time traces, and we briefly describe how we estimate the fluorophore number with it. Section 3 contains a detailed treatment of the fluorophore dynamics. We formulate the hidden two-timescale Markov model (abbreviated by HTMM) that is based on the description of fluorescent molecules as Markov chains, acting on different timescales with different transition rates. In particular, we derive expressions for the expectation and (co-)variance of the number of emitted photons in each frame, and provide results about spectral properties of the transition matrix which are useful for computational purposes (Appendix E and

F in the Supplementary Material [48]). In Section 4, we investigate how the number of emitted photons is transformed through (i) statistical thinning in the microscope and (ii) processing and amplification in the detector. To assist readability, the central notation that is introduced in Sections 3 and 4 is surveyed in Table A.1 in Appendix A of the Supplementary Material [48]. In Section 5, we then introduce the simplified "pseudo log-likelihood", comment on numerical issues for maximum likelihood estimation based on it, and present estimation results for simulated fluorescence intensity traces. Finally, in Section 6, we specialize our general model to the commonly used fluorophore Alexa 647. Section 7 contains a brief outlook that emphasizes open questions related to our work.

2. MODELING AND ESTIMATION

Super-resolution microscopy with single marker switching (SMS) relies on a series of fluorescence microscopy images, or frames, with only a small fraction of active fluorophores per image. This way, spatially close fluorophores are separated in time since they are unlikely to emit photons simultaneously. The resulting frames are used to localize the marker molecules with a superior precision on the nanometer scale [49]. The imaging is affected by the quantum physical behavior of the fluorophore, which leads to *switching* and *bleaching*, and by a series of subsequent manipulations of the emitted photons until they are detected by the camera and transformed to digital values [1]. Each step in this chain (depicted in Figure 2) modifies the original signal—photons emitted by the fluorophore—in a characteristic way and has to be taken into account. In the following, we present an outline of our approach to estimate the fluorophore number based on time traces extracted from a series of T camera images. More detailed considerations follow in subsequent sections.

Single fluorophore

Fluorophore dynamics is successfully modeled by Markov chains [35, 39]. The states of these chains roughly correspond to quantum physical states of the molecule (see Remark 1 below), which can exhibit very diverse lifetimes. The respective transition rates are governed by quantum mechanical kinetics that sensitively depend on the biochemical properties of the fluorophore's neighbourhood in the sample. Two of the states have a distinguished role in our model: the *bright* state, in which absorption and emission of photons is possible, and the *bleached* state, in which dyes have irreversibly lost their fluorescence functionality. Additionally, a number of temporary *dark* states, which, for example, correspond to triplet or redox states of the fluorophore [53], are usually necessary for a faithful description.

REMARK 1. In Markov chains, states with the same transition rates can be combined into a single state without losing the Markov property (see Appendix D for details). A reasonable fluorophore model does therefore not have to include every possible quantum physical state explicitly (like fine-structured rotational and vibrational substates). Rather, it only has to capture classes of states with similar dwell times and transition behavior. The number of such classes can be estimated from the data.

The phenomenon of fluorophores jumping between the bright and temporary dark states is denoted as *blinking* or *switching*. In our generic model for fluorescence, we finely resolve the fast dynamics inherent to the bright state, like single photon emissions, and model it as a Markov chain in its own right. This gives rise to a description that operates on two different time scales: a fast *inner model* that runs during the exposure time, and a slow *outer model* that captures states that are expected to persist over several frames. Figure 3 depicts our choice of states for the fluorophore Alexa 647, which we investigate more detailed in Section 6.

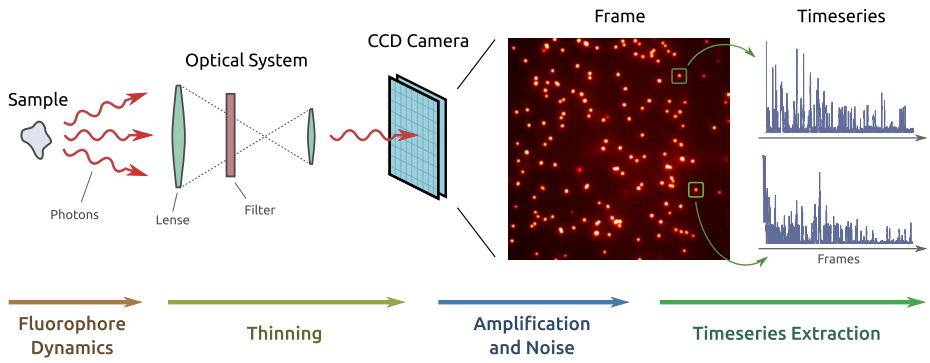

FIG. 2. *Overview of the modeling steps. During exposure, fluorophores within a labeled biological sample emit photons due to laser excitation. With a certain probability, these photons pass through the microscope (optical system) and are registered by a CCD camera. Over the course of the experiment, a series of camera frames is recorded. Summing up the intensities over a region of interest (ROI, green boxes) for each image yields a time series that captures the fluorescent activity in the respective ROI.*

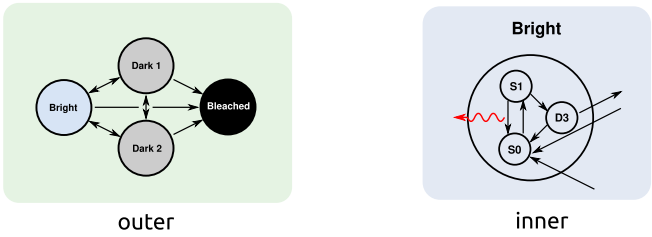

FIG. 3. *Exemplary inner and outer models that are used to describe the dynamics of the fluorophore Alexa* 647 (*see* [30]). *The wiggling red arrow indicates that transitions from the singlet state* S_1 *to the ground state* S_0 *cause the emission of a photon. The state* D_3 *is a short-lived dark state.*

Let X'_t denote the outer state of the fluorophore directly before the tth exposure, and let X_t denote its state after the exposure. The transition from X_{t-1} to X_t is depicted in Figure 4. It is modeled through applying one step of the outer dynamics on X_{t-1}, which yields X'_t, and then running the inner model, which changes X'_t to X_t, and which also yields a number Y_t of emitted photons during frame t. The distribution of Y_t depends on both X'_t and X_t. A complete description of our model for fluorophore dynamics is therefore given by a transition matrix M^ℓ for the long-time step, a matrix M^s for the short-time step, and the distribution $p_{xx'}$ of Y_t conditioned on $X_t = x$ and $X'_t = x'$, which we assume to be stationary. The combined chain

$$(X_0, X'_1, X_1, X'_2, \ldots)$$

of outer states is an inhomogeneous Markov chain with alternating transition matrices M^ℓ and M^s, while the individual chains $(X_t)_t$ and $(X'_t)_t$ are homogeneous with transition matrices $M^s M^\ell$ and $M^\ell M^s$, respectively.

In experiment, the states X_t and X'_t cannot be observed directly. We only obtain outputs of the measurement device (e.g., a CCD camera) generated through the Y_t emitted photons. This makes our ansatz a hidden Markov model. In Section 3, we derive the generating function of the process $Y = (Y_t)_{t=1}^T$ and obtain its expectation μ and the covariance Σ, which are eventually used to estimate the number of fluorophores in Section 5.

REMARK 2 (Notation). We refer to the model outlined above as hidden two-timescale Markov model, or HTMM. The observable part of this model, Y_t, denotes the number of photons that are emitted in the time interval between X'_t and X_t. One can therefore think of X'_t as the state X_{t^-} directly prior to X_t, and Y_t as an observation that accumulates from t^- to t.

Microscope and camera

Photons emitted by fluorescent dyes are directed randomly and may fail to enter the microscope, such that they are lost for the experiment. In addition, a photon may be absorbed by lenses, filters, or mirrors within the optical path. Consequently, each emitted photon has a probability $p_c < 1$ to reach the camera. When it reaches the camera, the position of the photon on the CCD sensor is randomized due to diffraction: light originating from a point source is spread to a blurred spot on the detector interface. From the viewpoint of classical physics, where light is modeled as a wave of electromagnetic radiation, this blurring is described by a convolution of the light intensity distribution with a nonnegative point spread function h (see [7, 17] for the underlying physics and [1] for a treatment in the context of statistics). In the quantum mechanical interpretation of light as photons, $h(z)$ denotes the probability that a photon emitted at the origin of the sample incides at pixel z on the detector, which leads to a multinomial distribution of incident photons to pixel locations. When the photon arrives at a pixel z, it is absorbed with a certain probability p_a and a so-called photo electron, that is, an electron ejected from the detector material due to energy transfer from the photon, is released. The total chance for a photon to reach the detector at any pixel and be absorbed is denoted by $p_d = p_c p_a$.

We call a region R on the image that captures the blurred spot created by one (or several close-by) fluorophores a region of interest (ROI). The total number Y'_t of detected photons in R is given by

$$Y'_t = \sum_{z \in R} Y'_{t,z},$$

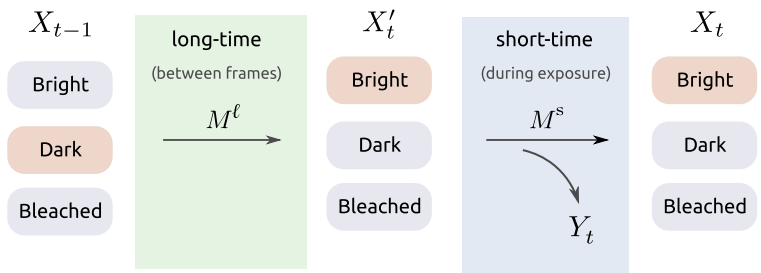

FIG. 4. *Single fluorophore model. At time* $t-1$, *the fluorophore can be in one of several distinct outer states* X_{t-1}. *When transitioning from* X_{t-1} *to* X_t, *we apply the long-time matrix* M^ℓ *to describe the dynamics that takes place between separate frames, and then apply the short-time matrix* M^s *for the effects of the fast dynamics on the outer state during exposure. In the latter step, the fluorophore emits a number* Y_t *of photons in frame* t.

where $Y'_{t,z}$ is the number of photons detected at pixel $z \in R$. Since we assume that the electrical circuits underlying individual pixels are identical in their properties, we can ignore the spatial distribution of photons within one ROI and work with $Y'_t \sim \mathrm{Bin}(Y_t, p_\mathrm{d})$ directly. This amounts to a binomial thinning of Y_t [19].

Since the electrical charge of a photoelectron is too small to be detected reliably, cameras employ an electron multiplying system that operates stochastically [25, 41]. Let $\mathcal{D}$ denote the distribution for the number of electrons after amplification of the incoming electron in the CCD. Then the final camera output value $\widetilde{Y}_t$, when summed over R, is given by

$$\widetilde{Y}_t = c \sum_{k=1}^{Y'_t} U_{t,k} + \epsilon_t + o, \tag{2.1}$$

with $U_{t,k} \sim \mathcal{D}$ i.i.d. for all t and k. The constant factor $c > 0$ results from the analog-to-digital conversion of the accumulated electron charge in the pixels, and the random variables ϵ_t collect different contributions of inevitable additional randomness—like background photons, thermal electrons in the electronics, or readout noise. Additionally, a constant positive offset o is added to the camera output to avoid noise induced fluctuations into the negative domain.

Multiple fluorophores

Each fluorophore in the bright state produces a diffraction-limited spot during exposure, as seen in the frame shown in Figure 2. The major difficulty for estimating the number of fluorophores reliably results from the fact that several fluorophores can contribute to the same spot if their mutual distance is small and if they are bright simultaneously. The core contribution of this article is to use the information from a temporal series of frames to estimate the total number m of fluorophores that are present in a given region of interest R.

REMARK 3. The unknown number m of fluorophores is the major quantity of interest to be estimated in a single ROI. By combining estimates for m from different ROIs, one can obtain quantitative information on the spatial fluorophore density in the entire image.

A crucial assumption that we use to model multi-fluorophore systems is *statistical independence*, that is, that no (relevant) physical interactions between the single fluorophores take place. We also assume that all m fluorophores are *identical* in their physical behavior, meaning that they can be described by a common fluorophore model with a common set of parameters. Then, the total number of emitted photons is given by the sum of m independent copies $Y_t^1, \dots, Y_t^m$ of the process Y_t,

$$Y_t^{(m)} = \sum_{k=1}^{m} Y_t^k. \tag{2.2}$$

Similarly, the time series obtained by summing the CCD values over the region R is composed of m independent versions $\widetilde{Y}_t^1, \dots, \widetilde{Y}_t^m$ of $\widetilde{Y}_t$. Therefore, the total signal we observe is encoded in the process

$$\widetilde{Y}_t^{(m)} = \sum_{k=1}^{m} \widetilde{Y}_t^k.$$

REMARK 4. The assumptions of independence and identical distribution are approximations that are justified for many typical experimental situations. Still, they can be violated, for example, if the spatial distance of neighbouring fluorophores is very small (<10 nm). Then, interactions like FRET (Förster Resonance Energy Transmission) become likely. The experimental study [30] highlights that our model indeed produces inconsistent results in this case.

Estimation

Our objective is to estimate m from a realization y of the process $\widetilde{Y}^{(m)} = (\widetilde{Y}_t^{(m)})_t$. Besides m, there are several other parameters that may have to be estimated, like the transition rates in the HTMM, or the initial distribution of the outer state. These parameters depend sensitively on details of the experimental setting, like the fluorophore type, the biochemical conditions in the sample, or the applied laser wavelengths and intensities. Some of these properties may vary from ROI to ROI. Furthermore, different types of fluorophores may even require different inner or outer models. This poses an interesting problem in model selection—which we will, however, not address in this article.

Since the number of unknown parameters, which we call γ for the moment, is typically small (e.g., at most 16 for our Alexa 647 model with three dark states, see Section 6), it is near at hand to employ maximum likelihood estimation. However, the computation of the MLE requires that we can evaluate the log-likelihood $l_y(\gamma)$ of the full model, which is unfeasible for two reasons: first, the number of terms in $l_y(\gamma)$ turns out to be overwhelming even for a moderate number T of frames; and second, we lack information about how the signal is transformed in the camera, since manufacturers usually only provide information about the first two moments of the camera-statistics $\mathcal{D}$. We therefore choose an approach that is based on approximating $\widetilde{Y}^{(m)}$ by a Gaussian process with the same expectation $\mu = \mu(\gamma)$ and covariance $\Sigma = \Sigma(\gamma)$ as $\widetilde{Y}^{(m)}$. This leads to the *pseudo log-likelihood*

$$\tilde{l}_y(\gamma) = -\frac{1}{2}\big[(y-\mu)\Sigma^{-1}(y-\mu) + \log\det\Sigma\big],$$

and parameter estimation reduces to maximizing $\tilde{l}_y(\gamma)$, which is still challenging but becomes numerically feasible. In particular, we have fewer degrees of freedom:

for Alexa 647, only 11 (compared to 16) independent parameters are necessary to fully describe μ and Σ, since the first two moments do not rely on all transition probabilities of the HTMM individually. In Section 5, we address a number of subtleties that come along with this approach, like nonlinear constraints on the parameter space and practical complications with the numerical optimization. As a proof-of-concept of our approach, we also include exemplary estimation results on simulated data.

We stress that other methods of estimation are certainly of interest, too. In particular, a Bayesian approach becomes feasible when prior knowledge on parameters is available. We do, however, not pursue this issue further in this article and contain ourselves to (pseudo) maximum likelihood based statistical analysis.

3. FLUOROPHORE DYNAMICS

The dynamics of fluorophores is at the heart of our model for the imaging process in Figure 2. Due to the i.i.d. assumption when modeling multiple fluorophores (see Remark 4), this effectively amounts to modeling a single fluorophore. In this section, we will treat the short-time dynamics as a "black box" with as few assumptions as is necessary for our intentions. Only later, when we specialize the model to the fluorophore Alexa 647 in Section 6, we elaborate in detail on a concrete inner model. As a guidance for the derivations that follow, consulting Table A.1 in Appendix A might prove helpful, as it summarizes the relevant notation that is introduced in this and the next section.

Fluorophore model

The outer state space of a fluorophore is described by one bright and $r \in \mathbb{N}$ dark states, including the bleached one. See Figure 3 for an example. We name the state space $\mathcal{S}$ and denote its elements by $x \in \mathcal{S} = \{0, \dots, r\}$, with $x = 0$ being the bright and $x = r$ being the bleached state. On $\mathcal{S}$, we consider two coupled time-discrete Markov chains $(X_t)_{t=0}^T$ and $(X'_t)_{t=1}^T$, where $T \in \mathbb{N}$ is the number of frames. The evolution of X_t and X'_t is given by the (stationary) transition matrices M^ℓ and M^{s}, where

$$M^\ell_{x'x} = \mathbb{P}(X'_t = x' \mid X_{t-1} = x) \quad \text{and} \quad M^{\mathrm{s}}_{xx'} = \mathbb{P}(X_t = x \mid X'_t = x') \tag{3.1a}$$

for $x, x' \in \mathcal{S}$. We interpret X'_t as state of the fluorophore directly before exposure in frame t, and X_t as state directly after exposure (see Remark 2 and Figure 4). The full transition matrix for X_t is given by

$$M = M^{\mathrm{s}} M^\ell. \tag{3.1b}$$

The transition of the fluorophore from X'_t to X_t during exposure is governed by the inner model, which also determines the number Y_t of photons that are emitted in the corresponding frame. We characterize the photon statistics of the inner model by the conditional distributions

$$p_{xx'}(y) = \mathbb{P}(Y_t = y \mid X_t = x, X'_t = x') \tag{3.2}$$

for $y \in \mathbb{N}_0$, which we assume to be time-stationary. The probabilities in (3.2) are collected in the matrix $P(y) = (p_{xx'}(y))_{x,x' \in \mathcal{S}}$. We furthermore use the symbol $\nu = (\nu_x)_{x \in \mathcal{S}}$ to denote the initial distribution, that is, the distribution of X_0. In total, specification of M^{s}, M^ℓ, P, and ν completely defines a single-fluorophore model.

DEFINITION 1. Any observable process $Y = (Y_t)_t$ as constructed above, with conditional distributions P in (3.2), transition matrices M^{s} and M^ℓ defined in (3.1), and initial (hidden) distribution ν, is denoted as hidden two-timescale Markov model, or HTMM.

LEMMA 1. *The likelihood of an HTMM under observation of a time series $y = (y_t)_{t=1}^T$ is*

$$l_y(M^{\mathrm{s}}, M^\ell, P, \nu) = \sum_{x \in \mathcal{S}^{T+1}} \left(\prod_{t=1}^T \sum_{x' \in \mathcal{S}} p_{x_t x'}(y_t) M^{\mathrm{s}}_{x_t x'} M^\ell_{x' x_{t-1}} \right) \nu_{x_0}, \tag{3.3}$$

where the outer sum covers all tuples $x = (x_0, x_1, \dots, x_T) \in \mathcal{S}^{T+1}$.

PROOF. A single transition step under observation of $Y_t = y$ photons is described by

$$\mathbb{P}(Y_t = y, X_t = x \mid X_{t-1} = z) = \sum_{x' \in \mathcal{S}} p_{xx'}(y) M^{\mathrm{s}}_{xx'} M^\ell_{x'z},$$

and the probability to observe the full time series $(y_t)_{t=1}^T$ can be written as

$$\mathbb{P}(Y_t = y_t \text{ for all } t) = \sum_{x \in \mathcal{S}^{T+1}} \left(\prod_{t=1}^T \mathbb{P}(Y_t = y_t, X_t = x_t \mid X_{t-1} = x_{t-1}) \right) \nu_{x_0}.$$

Combining these two equations yields the stated result. □

REMARK 5. The outer sum in equation (3.3) contains $(r+1)^{T+1}$ terms, and we are not aware of a way to significantly simplify the expression under general circumstances. Since there is at least one dark outer state (meaning $r \geq 1$) and the number T of frames for SMS microscopy is often well above 1000, it is impossible to directly evaluate the likelihood l_y numerically ($2^{1000} > 10^{300}$).

REMARK 6. We will eventually extend the HTMM in Definition 1 by (i) the generalization to multiple i.i.d. fluorophores, and (ii) the additional statistical modeling of the imaging process (see Section 4). These enhanced models will for convenience also be referred to as HTMMs, since the context usually clarifies which specific fluorophore model is meant.

Model restrictions

We introduce several restrictions on our model in order to reflect physical fluorophore properties and to make the analysis of (3.3) viable. The most evident constraint is that the bleached state $x = r$ acts as an absorbing state for both M^{s} and M^{ℓ}. We also assume that the fluorophore can leave the bright state $x = 0$ only during application of the inner model. Conversely, we suppose that a fluorophore that is not in its bright state is unaffected by the inner model. With these restrictions in place, M^{ℓ} and M^{s} can be brought in the respective parametric forms

$$M^{\ell} = \begin{pmatrix} 1 & q_{01} & \cdots & q_{0(r-1)} & 0 \\ 0 & q_{11} & \cdots & q_{1(r-1)} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & q_{(r-1)1} & \cdots & q_{(r-1)(r-1)} & 0 \\ 0 & q_{r1} & \cdots & q_{r(r-1)} & 1 \end{pmatrix}, \tag{3.4}$$

$$M^{\mathrm{s}} = \begin{pmatrix} q_{00} & 0 & \cdots & 0 & 0 \\ q_{10} & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ q_{(r-1)0} & 0 & \cdots & 1 & 0 \\ q_{r0} & 0 & \cdots & 0 & 1 \end{pmatrix},$$

where we defined transition probabilities $q_{xz} \in [0, 1]$ for $x \in \mathcal{S}$ and $z \in \mathcal{S} \setminus \{r\}$. The restrictions also imply that no photons are emitted if the exposure starts in a non-bright state, meaning that $p_{xx'}(0) = 1$ for $x' \neq 0$. In contrast, if the exposure begins in the bright state $x' = 0$, the fluorophore will produce photons and may switch to any other state $x \in \mathcal{S}$ until the end of the exposure.

An additional assumption that is required to make the model analytically tractable is that the distribution p_{x0} does not depend on the final state x if the fluorophore exits the bright state. This means that

$$p_{x0} = p_{10} \quad \text{if } x > 0. \tag{3.5}$$

REMARK 7. Condition (3.5) can be understood as assuming a common "exit state" in the inner model that is the only possibility for the fluorophore to become dark during exposure. From this state, it can then jump to all dark states of the outer model as soon as the frame ends. Note that this exit state does not have to correspond to a (single) physical state: it could cover several states with similar exit conditions (see Remark 1 in this context).

Under these constraints, the conditional distributions of the photon statistics defined in (3.2) read

$$P(y) = \begin{pmatrix} p_{00}(y) & \delta(y) & \cdots & \delta(y) \\ p_{10}(y) & \delta(y) & \cdots & \delta(y) \\ \vdots & \vdots & \ddots & \vdots \\ p_{10}(y) & \delta(y) & \cdots & \delta(y) \end{pmatrix}, \tag{3.6}$$

where δ denotes the Dirac measure with point mass 1 on $y = 0$. Fluorophore models that satisfy conditions (3.4) and (3.6) are collected in the set $\mathcal{F}^{\mathrm{s}}$ of (physical) single-fluorophore models.

Generating function

Even when exploiting the additional constraints for $\mathcal{F}^{\mathrm{s}}$, the process $Y = (Y_t)_t$ of emitted photons remains very complex. In particular, it is hard to use the HTMM for straightforward inference. Maximum likelihood estimation, for example, is impossible for real world datasets due to the prohibitive expense of calculating the likelihood, see Remark 5. We can, however, use the moment generating function of Y and the specific structure of $\mathcal{F}^{\mathrm{s}}$ to calculate expressions for the expectations $\mu_t = \mathbb{E}[Y_t]$ and the covariance Σ,

$$\Sigma_{tt'} = \mathbb{E}\big[(Y_t - \mu_t)(Y_{t'} - \mu_{t'})\big],$$

where $t, t' = 1, \ldots, T$. These moments carry relevant information and allow recovering the number of fluorophores in case of multiple molecules (see Section 5). For preparation, we first look at the moment generating matrix $G(s)$ associated to $P(y)$ and find

$$G(s) = \begin{pmatrix} G_{00}(s) & 1 & \cdots & 1 \\ G_{10}(s) & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ G_{10}(s) & 1 & \cdots & 1 \end{pmatrix}, \tag{3.7}$$

where $G_{00}(s) = \mathbb{E}[e^{sY_t} \mid X_t = X'_t = 0]$ and $G_{10}(s) = \mathbb{E}[e^{sY_t} \mid X_t = 1, X'_t = 0]$. Then, we define the auxiliary matrix

$$H(s) = \big(G(s) \circ M^{\mathrm{s}}\big) M^{\ell}, \tag{3.8}$$

where $\circ$ denotes the entry-wise (Hadamard) product. In the following, we will only consider inner models for which the expectations G_{00} and G_{10} exist and are finite in some vicinity $(-\epsilon, \epsilon)$ of zero for $\epsilon > 0$. In particular, this implies that all derivatives of H exist at $s = 0$ [14].

LEMMA 2. *The moment generating function G_Y of the process Y is*

$$G_Y(\tau) = (1, \ldots, 1) H(\tau_T) \cdots H(\tau_1) \nu \tag{3.9}$$

for $\tau = (\tau_1, \ldots, \tau_T) \in (-\epsilon, \epsilon)^T$.

The proof of this statement can be found in Appendix B.

Moments of the inner model

Before we derive the expectation and covariance of Y by differentiating equation (3.9), we introduce three parameters θ_1, θ_2, and θ_3 that describe the photon emission statistics p_{00} and p_{10} up to second order. The first parameter θ_1 describes the expected number of photons emitted during the frame if the fluorophore was in the bright state initially,

$$\begin{aligned} \theta_1 &= \mathbb{E}\big[Y_t \mid X'_t = 0\big] \\ &= \sum_{x \in \mathcal{S}} q_{x0} \mathbb{E}\big[Y_t \mid X'_t = 0, X_t = x\big], \end{aligned} \tag{3.10a}$$

where we used that $q_{x0} = M^{\mathrm{s}}_{x0} = \mathbb{P}(X_t = x \mid X'_t = 0)$ by definition. The second parameter θ_2 quantifies the contribution to the expectation θ_1 if the fluorophore not only starts the frame in the bright state but also stays there,

$$\theta_2 = \frac{q_{00}\mathbb{E}[Y_t \mid X'_t = X_t = 0]}{\theta_1}. \tag{3.10b}$$

Finally, we capture the conditioned variance of Y_t given $X'_t = 0$ via the parameter θ_3,

$$\theta_3 = \frac{\mathrm{Var}[Y_t \mid X'_t = 0]}{\theta_1^2} - \frac{1}{\theta_1}. \tag{3.10c}$$

This parameter can be viewed as the excess relative variance with respect to a Poisson distribution: if $Y_t \mid X_t = 0$ was distributed Poissonian, then $\theta_3 = 0$. A Poissonian statistics is often assumed as an approximation for the photon emission of fluorophores, but corrections may be necessary for accurate results [1].

REMARK 8. The parameters $\theta = (\theta_1, \theta_2, \theta_3)$ clearly depend on the inner model choice and are usually related to the entries of the short-time transition matrix M^{s}. For example, we show in Section 6 that θ_2 is fully determined by q_{00} in the inner model that we use for the fluorophore Alexa 647.

Expectation

In order to derive analytical expressions for the expectation and covariance, we will assume that the transition matrix $M = M^{\mathrm{s}}M^{\ell}$ is diagonalizable and has eigenvalues $\lambda_0, \dots, \lambda_r \in \mathbb{C}$. We argue that this assumption is no significant restriction, since the stochastic matrices that are not diagonalizable form a null set in a reasonable sense—see Lemma E.1 in Appendix E for details. We thus write

$$M = V\Lambda V^{-1}, \tag{3.11}$$

where $\Lambda = \mathrm{diag}(\lambda_0, \dots, \lambda_r)$ and where V is a matrix containing the eigenvectors of M as columns. Due to the absorbing nature of the bleached state $x = r$, we can assume that $\lambda_r = 1$ with eigenvector $(0, 0, \dots, 1)^{\mathrm{T}}$, which hence constitutes the last column of V.

THEOREM 1. *Assume that the Markov matrix M for a single-fluorophore model in $\mathcal{F}^{\mathrm{s}}$ is diagonalizable like in (3.11). Then the expectation value μ_t of the number Y_t of photons emitted by the fluorophore at time $t = 1, \dots, T$ is*

$$\mu_t = \theta_1 \sum_{x \in \mathcal{S}} \alpha_x \lambda_x^{t-1}, \tag{3.12}$$

where the coefficients α_x for $x \in \mathcal{S}$ are defined by

$$\alpha_x = V_{0x} \frac{\lambda_x}{q_{00}} \sum_{z \in \mathcal{S}} V^{-1}_{xz} \nu_z. \tag{3.13}$$

This statement follows from differentiating the moment generating function G_Y in (3.9) and exploiting the structural restrictions of M^{s}, M^{ℓ}, and G in (3.4) and (3.7). For a detailed proof, we refer to Appendix B.

When deriving the covariance of Y later on, we will need the expectation value μ_t^0 of Y_t on the condition that the fluorophore was in the bright state at the beginning of the experiment. This corresponds to the case $\nu = (1, 0, \dots, 0)$. According to equation (B.4) in the proof of Theorem 1, we find

$$\mu_t^0 = \mathbb{E}[Y_t \mid X_0 = 0] = \frac{\theta_1}{q_{00}} (M^t)_{00}. \tag{3.14}$$

Under assumption (3.11) of diagonalizability for M, Theorem 1 lets us write

$$\mu_t^0 = \theta_1 \sum_{x \in \mathcal{S}} \alpha_x^0 \lambda_x^{t-1},$$

where the coefficients α_x^0 are given by equation (3.13) with $\nu = (1, 0, \dots, 0)$,

$$\alpha_x^0 = \frac{\lambda_x}{q_{00}} V_{0x} V^{-1}_{x0}. \tag{3.15}$$

These auxiliary coefficients can be related to α_x. If $\nu_0 = 1$, then $\alpha_x^0 = \alpha_x$ by definition. If $\nu_0 < 1$, we can plug $\nu' = \frac{1}{1-\nu_0}(0, \nu_1, \dots, \nu_r)$ in equation (3.13) and define

$$\alpha_x^1 = \frac{\lambda_x}{q_{00}} V_{0x} \sum_{z \in \mathcal{S}} V^{-1}_{xz} \nu'_z, \tag{3.16}$$

which allows us to decompose α_x as

$$\alpha_x = \nu_0 \alpha_x^0 + (1 - \nu_0)\alpha_x^1.$$

This way of splitting up the model parameters has the advantage that a simple set of constraints for α_x^0 and α_x^1 arises (see Lemma 3 below).

Spectral properties and parameter constraints

The eigenvalues λ_x and coefficients α_x in Theorem 1 can in general be complex-valued. When estimating these parameters numerically, however, it is beneficial to assume real eigenvalues and eigenvectors of M. In Appendix F, we provide some criteria that guarantee $\lambda_x \in [0, 1]$ for $r \leq 3$. In summary, (i) real and (ii) positive eigenvalues are ensured if the diagonal values of the transition matrix M are (i) diverse and (ii) large enough. Usually, both of these assumptions are physically reasonable: the diagonal values are diverse if the outer states of the fluorophore exhibit diverse live times, and they are large if the outer states are on average stable over several frames. Under the restriction $\lambda_x \in [0, 1]$ on the spectrum of M, equation (3.12) states that the expected number μ_t of emitted photons is the superposition of exponential decays with timescales determined by λ_x.

Furthermore, note that the coefficients α_x are implicitly constrained by their definition in equation (3.13). First,

$$\alpha_r = 0$$

is enforced due to $V_{0r} = 0$. This is physically expected as fluorophores in the bleached state do not emit photons. Second, summing over $x \in \mathcal{S}$ in (3.13) shows

$$\sum_{x\in\mathcal{S}} \alpha_x = \frac{1}{q_{00}} \sum_{z\in\mathcal{S}} M_{0z}\nu_z = (1, q_{01}, \ldots, q_{0(r-1)}, 0)\nu. \tag{3.17}$$

Similarly, we find the relation

$$\sum_{x\in\mathcal{S}} \frac{\alpha_x}{\lambda_x} = \frac{\nu_0}{q_{00}} \tag{3.18}$$

by dividing equation (3.13) by λ_x and again summing over $x \in \mathcal{S}$. Applying the last three equations to the coefficients α_x^0 and α_0^1, defined in (3.15) and (3.16), yields a set of simple constraints.

LEMMA 3. *We have* $\alpha_r^0 = \alpha_r^1 = 0$. *Furthermore, it holds that*

$$\sum_{x\in\mathcal{S}} \alpha_x^0 = 1 \quad \textit{and} \quad \sum_{x\in\mathcal{S}} \frac{\alpha_x^0}{\lambda_x} = \frac{1}{q_{00}}, \tag{3.19a}$$

$$0 \le q_{00} \sum_{x\in\mathcal{S}} \alpha_x^1 \le 1 \quad \textit{and} \quad \sum_{x\in\mathcal{S}} \frac{\alpha_x^1}{\lambda_x} = 0. \tag{3.19b}$$

PROOF. The first statement holds due to $V_{0r} = 0$. The relations in (3.19a) follow from equation (3.17) and (3.18) for $\nu = (1, 0, \ldots, 0)$. Constraint (3.19b) follows similarly if we take into account that ν' is a probability vector with $\nu'_0 = 0$. □

Covariance

We next look at the covariance matrix Σ of the process Y, which can be obtained from the second derivatives of the moment generating function.

THEOREM 2. *Under assumption* (3.11), *the covariance matrix* Σ *of the process* $Y = (Y_t)_{t=1}^T$ *is*

$$\Sigma_{tt} = (\theta_1(\theta_3 + 1) + 1 - \mu_t)\mu_t, \tag{3.20a}$$

$$\Sigma_{tt'} = \left[\left(\theta_2 - q_{00}\frac{1-\theta_2}{1-q_{00}}\right)\mu^0_{t-t'} + \frac{1-\theta_2}{1-q_{00}}\mu^0_{t-t'+1} - \mu_t\right]\mu_{t'}, \tag{3.20b}$$

on the diagonal and off-diagonal with $t > t'$, *respectively.*

The theorem is proven in Appendix B. Similar to Theorem 1, the proof crucially relies on the special structure of M^{s}, M^{ℓ}, and G.

REMARK 9. The expectation and covariance in equations (3.12) and (3.20) depend on the eigenvalues and eigenvectors of the matrix $M = M^{\mathrm{s}}M^{\ell}$, but not on the single transition probabilities in M^{s} and M^{ℓ} directly. From joint knowledge of μ and Σ, the parameters $\nu_0, q_{00}, \alpha_x^0, \alpha_x^1, \lambda_x, \theta_1, \theta_2$, and θ_3 are identifiable. Not all of them, however, are independent (see Lemma 3), and knowing these parameters is in general not sufficient to reconstruct the matrices M^{s} and M^{ℓ}. Plots of μ and Σ as well as simulation results for the processes X and Y are depicted in Figures 5 and 6.

REMARK 10 (variance "dip"). Figure 5(d) illustrates a characteristic property of the variance Σ_{tt} in dependence of the frame number t. Initially, the variance increases for some frames before it subsequently relaxes towards the background noise exponentially. This salient "dip" in the variance curve is also observed in experimental data [30] for large values of ν_0, that is, if most fluorophores are bright at the beginning of the experiment. It is caused by bright fluorophores getting dark during the first few frames, such that the observable distribution of photons Y_t is effectively an additive model composed of two parts: dark fluorophores with $Y_t = 0$ and bright ones with Y_t around θ_1. This split in the distribution of Y_t temporarily causes a high variance until the number of dark fluorophores eventually dominates in the long run.

Multiple fluorophores

The signal we observe in experiments is based on the fluorescent activity of an unknown number m of fluorophores. As we will typically not be able to distinguish between the contributions resulting from different fluorophores, we can only rely on the total number $Y_t^{(m)}$ of photons emitted in frame t. It is given by the sum of m single-fluorophore processes Y^k modeled via $\mathcal{F}^{\mathrm{s}}$,

$$Y_t^{(m)} = \sum_{k=1}^{m} Y_t^k. \tag{3.21}$$

We make the assumption that the contributions Y^k are independent and identically distributed (see Remark 4). Even though these assumptions are approximations—conditions like the biochemical properties of the fluorophore's neighbourhood or its spatial orientation have a certain impact—they appear to lead to a decent description for the multi-fluorophore dynamics in practical situations [30]. The set of all multi-fluorophore models that obey the i.i.d. assumptions is henceforth denoted by $\mathcal{F}$.

The expectation and covariance of $Y^{(m)}$ as sum of m i.i.d. random processes simply acquire the prefactor m with respect to the single-fluorophore expressions. Note that we will use the same symbols μ, μ^0, and Σ as for the single-fluorophore process, see (3.12), (3.14), and (3.20), to denote the respective generalizations to $m \ge 1$ fluorophores.

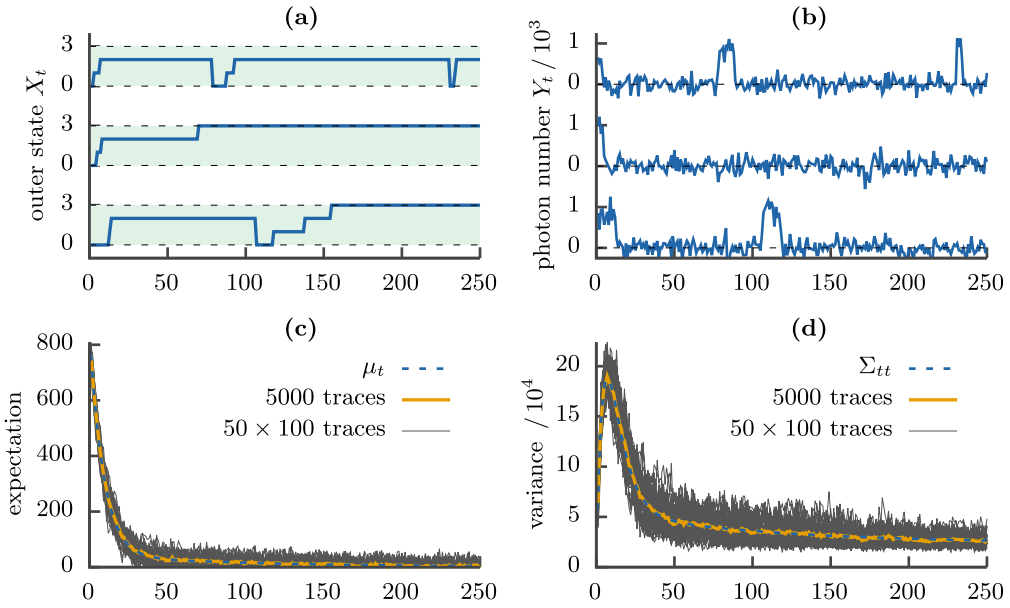

FIG. 5. *Simulation results of the single-fluorophore HTMM. (a) shows three exemplary paths of the fluorophore in the outer state space* $\mathcal{S} = \{0, \dots, r\}$ *for* $r = 3$. *One can see that two of the three fluorophores have already bleached in the first* 250 *time steps. (b) shows three fluorescence traces* $y = (y_t)_{t=1}^{250}$ *corresponding to the paths in (a). To obtain more realistic traces, Gaussian white noise with mean* 0 *and standard deviation* $\theta_1/5$ *was added to each observation* y_t. *(c)–(d) show the theoretical expectation* μ_t *and variance* Σ_{tt} *of* Y_t *compared to their empirical estimates for* 5000 *and* 100 *simulated traces. In all simulations, we use the inner model described in Section 6. The parameters for the inner and outer model are chosen such that the resulting traces roughly resemble the experimental data in* [30].

THEOREM 3. *The expectation* μ *and covariance* Σ *of a multi-fluorophore process* $Y^{(m)}$ *in* $\mathcal{F}$ *are*

$$\begin{aligned}\mu_t &= m\theta_1 \sum_{x\in\mathcal{S}} \alpha_x \lambda_x^{t-1} \\ &= m\theta_1 \sum_{x=0}^{r-1} (\nu_0 \alpha_x^0 + (1-\nu_0)\alpha_x^1)\lambda_x^{t-1},\end{aligned} \tag{3.22a}$$

and

$$\Sigma_{tt} = \frac{1}{m}\big(m\theta_1(\theta_3+1) + m - \mu_t\big)\mu_t, \tag{3.22b}$$

$$\begin{aligned}\Sigma_{tt'} = \frac{1}{m}\Bigg[&\bigg(\theta_2 - q_{00}\frac{1-\theta_2}{1-q_{00}}\bigg)\mu_{t-t'}^0 \\ &+ \frac{1-\theta_2}{1-q_{00}}\mu_{t-t'+1}^0 - \mu_t\Bigg]\mu_{t'},\end{aligned} \tag{3.22c}$$

where

$$\mu_t^0 = m\theta_1 \sum_{x\in\mathcal{S}} \alpha_x^0 \lambda_x^{t-1}, \tag{3.22d}$$

for $t, t' = 1, \dots, T$ *with* $t > t'$. *The coefficients* α_x, α_x^0, *and* α_x^1 *are given by equations* (3.13), (3.15), *and* (3.16), *respectively.*

Parameterization

At this point, it is instructive to think about the parameterization of the multi-fluorophore model class $\mathcal{F}$ (see also Table A.1 on page 2). The full model for the photon emission process $Y_t^{(m)}$ depends on the fluorophore number m, on all transition probabilities $Q = (q_{xz})$, on the initial state ν, and on an unspecified number of parameters that come with a concrete choice of the inner model. If we only want to describe the first two moments μ and Σ, however, several parameters become hidden and are not required to be estimated for our purposes.

According to equations (3.22a)–(3.22d), we only need the fluorophore number m, the fraction of initially bright fluorophores ν_0, the probability q_{00} for a fluorophore to stay bright during the exposure, and the eigenvalues $\lambda = (\lambda_x)_{x\in\mathcal{S}}$ as well as coefficients $\alpha^0 = (\alpha_x^0)_{x\in\mathcal{S}}$ and $\alpha^1 = (\alpha_x^1)_{x\in\mathcal{S}}$ for the multi-exponential decay in order to calculate the expectation and covariance. The parameters m and ν_0 contribute one degree of freedom each, while q_{00} is fixed by λ and α^0 due to equation (3.19a). To specify λ, assuming it is real, we need r free components since $\lambda_r = 1$ is determined through the bleached state. Similarly, $\alpha_r^0 = \alpha_r^1 = 0$. Due to the constraints (3.19) in Lemma 3, the parameters α^0 and α^1 hence contribute $r-1$ free components each. This makes a total number of $3r$ independent parameters, compared to $r^2 + (r-1)$ degrees of freedom needed to specify all components of Q and ν.

The three parameters $\theta = (\theta_1, \theta_2, \theta_3)$ are sufficient to specify the effect of the inner model in the second-order description. Still, specific knowledge of the inner model is necessary, since the relation of θ to other parameters is unclear otherwise. For example, an inner model with a

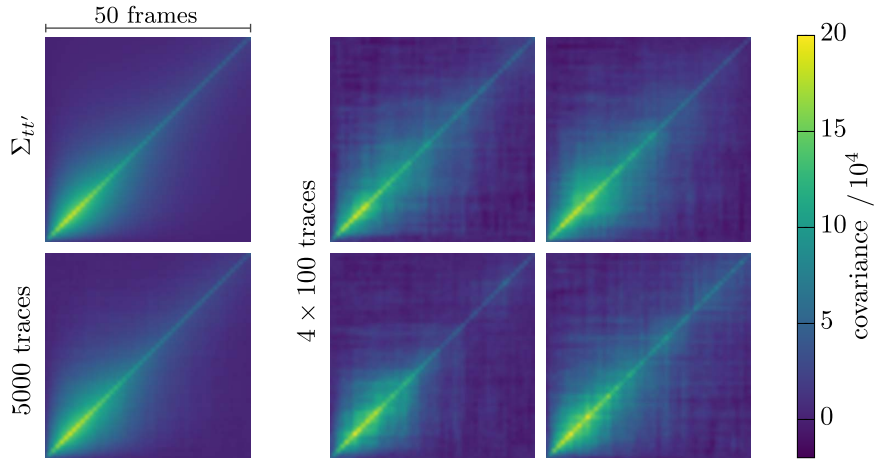

FIG. 6. *Real and empirical covariance of the HTMM. Shown are the covariances* $\Sigma_{tt'}$ *for* $t, t' = 1, \ldots, 50$ (*upper left image*) *as well as their empirical counterparts for* 5000 *traces* (*lower left image*) *and* 100 *traces* (*four images on the right*). *The same simulated traces as in Figure* 5(*c*)–(*d*) *are used.*

Poissonian photon statistics $Y_t \mid X'_t = 0$ enforces $\theta_3 = 0$, which evidently reduces the number of free parameters. Similarly, θ_2 is not a free parameter for the inner model that we employ to describe the fluorophore Alexa 647 in Section 6; it is completely determined by q_{00}.

A setting we want to emphasize is the one where $\nu_0 = 1$, that is, where each fluorophore is bright at the beginning of the experiment. This can be enforced by the experimental setup, like in the super-resolution scheme applied in [30]. Then, the $r-1$ parameters that are needed to specify the coefficients α^1 drop out of the formulae for μ and Σ, which makes this choice particularly beneficial.

4. IMAGE ACQUISITION

In the previous section, we introduced an elaborate statistical model $\mathcal{F}$ for the number of photons that are generated by a set of m fluorophores during a series of exposures in super-resolution microscopy. We next look at the image acquisition procedure and discuss the relation between the photon emission process $Y^{(m)}$ and the final time trace $\widetilde{Y}^{(m)}$ captured by the CCD camera. Fortunately, most processing steps subsequent to the emission of photons—like thinning in the microscope or amplification through the CCD camera—can be included into the model by modifying the photon statistics p_{00} and p_{10}, see (3.6). Consequently, merely the parameters $\theta = (\theta_1, \theta_2, \theta_3)$ will be affected in our second-order description, and equations (3.22a–d) for the expectation μ and the covariance Σ will remain intact: we just need to substitute θ by suitable transformed parameters θ'.[1]

It might thus seem superfluous to explicitly model any further steps in the microscope and camera, since we will typically estimate θ from the data anyway. However, there are several reasons why it is important to understand how the original parameter θ is transformed to θ'. First, these transformations could alter the constraints placed on parameters by the inner model (like $\theta_3 = 0$ if $Y_t \mid X'_t = 0$ is Poissonian) by possibly introducing new parameters (such that θ'_3 could be a free parameter again, e.g., due to an unknown variance of the amplification for the specific camera model). Second, the relation between θ and θ' could be interesting in its own right, because θ contains immediate information about the actual physics of the fluorophore, while θ' merges this information with further details of the experimental setup. This additional degree of insight could also be helpful for Bayesian inference approaches, where prior knowledge about the parameters is taken into account.

Thinning

In Section 2, we mentioned that only a certain fraction of emitted photons hits the detector interface and is registered at some CCD pixel. Many photons will fail to reach the optical pathway or will be absorbed by the equipment (like lenses or mirrors). The probability that an emitted photon triggers a photoelectron in a specified region of interest R of the camera was denoted by p_{d}, such that

$$Y'_t \sim \mathrm{Bin}(Y_t, p_{\mathrm{d}}) \tag{4.1}$$

models the thinned photon number for a single fluorophore. The parameter transformation from θ to θ' that accompanies this thinning process can be established by

[1] This is not entirely accurate. Camera noise contributions that do not depend on the fluorophore and its state of activity, which we called ϵ_t in equation (2.1) of Section 2, cannot be modeled that way and have to be considered separately. Their inclusion in the inner model would require a signal $Y_t > 0$ even for fluorophores in dark states $X'_t > 0$, which we explicitly prohibited during our derivations of μ and Σ in the previous section.

plugging Y_t' in equations (3.10a-c) defining the inner parameters (see Appendix B for the respective calculations).

LEMMA 4. *The moment parameters θ' of the thinned photon counts Y_t' in (4.1) are given by*

$$\theta_1' = p_{\mathrm{d}}\theta_1, \qquad \theta_2' = \theta_2, \qquad \theta_3' = \theta_3. \tag{4.2}$$

Therefore, only θ_1 is affected by binomial thinning while θ_2 and θ_3 are left unaffected.

Signal amplification

When photoelectrons are read out in an EMCCD camera, an electron amplifying system that consists of a cascade of electron multipliers (EM) is triggered. Each stage of this cascade has a certain probability of generating extra electrons, and the succession of many stages results in a stochastic signal amplification of the incident photons. In Appendix C, we introduce a statistical model for this (noisy) signal amplification in the camera and derive the transformation rules

$$\theta_1' = a\theta_1, \qquad \theta_2' = \theta_2, \qquad \theta_3' = \theta_3 + \frac{f^2-1}{\theta_1}, \tag{4.3}$$

where $a > 0$ and $f^2 > 0$ are camera specific parameters, namely the overall amplification factor and the excess noise factor.

Remarks on the full model

The preceding considerations show how the modeling steps of thinning and signal amplification transform the moment parameters θ. As discussed in Appendix C, however, there are other contributions to the full model for the observed time trace, introduced as $\widetilde{Y}^{(m)}$ in equation (C.6), which cannot be incorporated in the description by merely modifying θ. These include additional noise in each frame with (possibly time dependent) variance σ_t^2, which results from background photons or thermal noise, and an offset $o > 0$ in the single pixels of the sensor.

Some of the emerging parameters are known, like f^2, or can be estimated independently from the fluorophore model, like σ_t^2, a, or o. The detection probability p_{d}, however, cannot be separated statistically from the expected number of photons θ_1. For this reason, we drop p_{d} from the final model formulation and simply write θ_1 for what ought to be called $p_{\mathrm{d}}\theta_1$ in the following. We also assume preprocessed time traces, where the camera data $\widetilde{Y}^{(m)}$ has been corrected by the offset o and divided by the total amplification a (see (C.7) in Appendix C). The mean μ and the covariance Σ of the resulting normalized process are then given by

$$\mu_t = m\theta_1 \sum_{x=0}^{r-1} \big(\nu_0\alpha_x^0 + (1-\nu_0)\alpha_x^1\big)\lambda_x^{t-1}, \tag{4.4a}$$

$$\Sigma_{tt} = \frac{1}{m}\big(m\theta_1(\theta_3+1) + mf^2 - \mu_t\big)\mu_t + \sigma_t^2/a^2, \tag{4.4b}$$

$$\Sigma_{tt'} = \frac{1}{m}\Big[\Big(\theta_2 - q_{00}\frac{1-\theta_2}{1-q_{00}}\Big)\mu_{t-t'}^0 + \frac{1-\theta_2}{1-q_{00}}\mu_{t-t'+1}^0 - \mu_t\Big]\mu_{t'}, \tag{4.4c}$$

where

$$\mu_t^0 = m\theta_1 \sum_{x=0}^{r-1} \alpha_x^0\lambda_x^{t-1}, \tag{4.4d}$$

for $t, t' = 1, \ldots, T$ with $t > t'$. This can easily be derived from Theorem 3 and definition (C.6) of the process $\widetilde{Y}^{(m)}$, using equations (4.2) and (4.3).

5. ESTIMATION

In the previous sections, we have developed a statistical model for the time series of the observable fluorescence generated by m fluorophores. We now address the central goal of this article: estimating m with our model. To this end, let

$$y = (y_t)_{t=1}^T \tag{5.1}$$

be a realization of the process $\widetilde{Y}^{(m)}$ that models the observable fluorescence during the measurement process. In practice, y is obtained from a series of microscopy images (frames) by summing the camera output values over some fixed region of interest. See [30] for details on necessary or beneficial preprocessing steps.

The fluorophore number m will not be the only unknown parameter of $\widetilde{Y}^{(m)}$. Indeed, several (or even all) of the parameters $\gamma = (m, q_{00}, \nu_0, \alpha^0, \alpha^1, \lambda, \theta)$ that describe the first two moments of $\widetilde{Y}^{(m)}$ (see Table A.1 in Appendix A) are usually not known precisely, since the properties of the fluorophore heavily depend on the fluorophore type itself and on details of the experimental setting. The preferable choice is therefore to jointly estimate all values in γ, respecting the constraints that are inherent to the model.[2]

Pseudo log-likelihood

The process $\widetilde{Y}^{(m)}$ has a complicated non-Gaussian and non-stationary structure with long term correlations. Furthermore, it is essentially impossible to evaluate the likelihood function numerically for a given set of parameters as it consists of too many terms (see Remark 5). This makes direct likelihood-based methods to estimate the model parameters γ unsuitable. To overcome this difficulty, we approximate $\widetilde{Y}^{(m)}$ by a Gaussian process with known parametric form of the expectation $\mu = \mu(\gamma)$ and

[2] See the end of Section 3 for a general discussion of these constraints, and Section 6 for a discussion in context of the fluorophore Alexa 647.

covariance $\Sigma = \Sigma(\gamma)$, see equations (4.4a-d). This leads to the *pseudo log-likelihood*

$$\tilde{l}_y(\gamma) = -\frac{1}{2}\big[(y-\mu)\Sigma^{-1}(y-\mu) + \log\det\Sigma\big], \tag{5.2}$$

where we neglect an additive constant that would belong to the full log-likelihood of the Gaussian process. We estimate the model parameters γ by finding a set of values $\hat{\gamma}$ that maximize this expression,

$$\hat{\gamma} = \operatorname*{arg\,max}_{\gamma\in\Gamma} \tilde{l}_y(\gamma). \tag{5.3}$$

While this approach significantly simplifies the estimation compared to direct treatment of $\widetilde{Y}^{(m)}$, equation (5.3) still represents a non-convex optimization problem over a parameter space Γ that obeys several (non-linear) constraints. As such, there is neither a closed theory nor a canonical method for numerical treatment available.

In the general case, where all parameters in γ need to be estimated and no additional constraints can be posed, Γ will be a manifold of dimension $3r+3$ (see the discussion at the end of Section 3). For specific choices of the inner model, there could be fewer free parameters in θ, reducing the dimension of Γ. If ν_0 is known and not equal to 1 or 0, there is 1 parameter less. If it is known and equal to 1 or 0, then there are even r free parameters less, because α^1 respectively α^0 drop out of the expressions in (4.4). In case of Alexa 647, with a model of $r=3$ dark states and an additional constraint on θ_2, see Section 6, we are thus confronted with an 8, 10, or 11-dimensional parameter space Γ.

Numerical procedure

Finding a numerical solution of the optimization problem (5.3) poses several challenges. First, the high dimensionality of Γ in combination with both equality and inequality constraints forces one to apply very general optimization schemes (like (quasi) Newton methods, primal-dual-splitting, or nonlinear conjugate gradient methods). Some schemes rely on gradient information about $\tilde{l}_y(\gamma)$ while others are gradient-free. All of them, however, work in a local fashion and thus crucially rely on the choice of suitable initial parameters γ_{init}. Indeed, optimizing over all parameters of γ simultaneously was empirically found to depend sensitively on the initial values and did not always converge to the global maximum. Instead, approaches where different components of γ were held fixed at times—and partial optimizations with methods like the simplex-search algorithm by Nelder and Mead [37] were conducted sequentially—turned out to be more successful in practice.

To find suitable initial parameters, different methods can be applied. One option is to first employ a multi exponential fit of the expectation value μ_t. This will yield first guesses for λ and for the product $m\theta_1\alpha$. However, this fit may be of poor quality if the number r of dark states is large. The value of θ_1 may furthermore be guessed from late segments in the time traces, where with high probability at most one fluorophore is active due to bleaching. In case of the experimental data analyzed in [30], we eventually found a set of initial parameters that worked well on a range of different image series in experimentally similar conditions.

Estimation results

In order to demonstrate that the proposed pseudo log-likelihood approach works in principle, we apply it to estimate γ for simulated traces. We use the same model choices like in Figures 5 and 6 and consider the case $\nu_0 = 1$. The true parameter values in this setting are given by

$$\theta_1 \approx 767, \qquad \theta_2 \approx 0.95,$$
$$\theta_3 \approx 0.056, \qquad q_{00} \approx 0.9,$$

as well as

$$\lambda \approx (0.99, 0.89, 0.86), \qquad \alpha^0 \approx (0.05, 1.23, -0.28).$$

Including the molecule number m, there are 8 degrees of freedom in total. We contrast two different choices of initial parameters: the true parameters $\gamma^1_{\text{init}} = \gamma$, and an arbitrary selection γ^2_{init} determined by

$$\begin{aligned} &\theta_1 = 1, \qquad \theta_3 = 0.1, \\ &\lambda = (0.99, 0.9, 0.8), \quad \text{and} \quad m\theta_1\alpha^0 = (1,1,1), \end{aligned} \tag{5.4}$$

which in large part yields parameters very different from the true ones.

The estimation results for $m \in \{10, 100, 1000\}$ under the observation of $T \in \{100, 500\}$ time points of the simulated process $Y^{(m)}$ are depicted in Figure 7. One can see that the estimates are largely reasonable, especially when m is large or when the initial parameters are set to the oracle choice γ^1_{init}. On the other hand, the estimator evidently struggles for small values of m and poor initial parameters γ^2_{init}, where it even fails to provide the right magnitude of m in about one fourth of the repetitions. Generally, providing longer fluorescence traces ($T = 500$ instead of 100) improves the estimator notably—even though it also results in a more pronounced bias under γ^2_{init} for $m = 100$.

The estimation of other parameters, like the average photon number θ_1, the excess relative variance θ_3, or the probability q_{00} that the fluorophore stays in the bright state during exposure, performed similarly to the estimation of m. However, we often observed that the estimator severely struggles to guess the values of α and λ correctly, particularly for small values of m. This is not surprising, since the coefficients and exponents for the

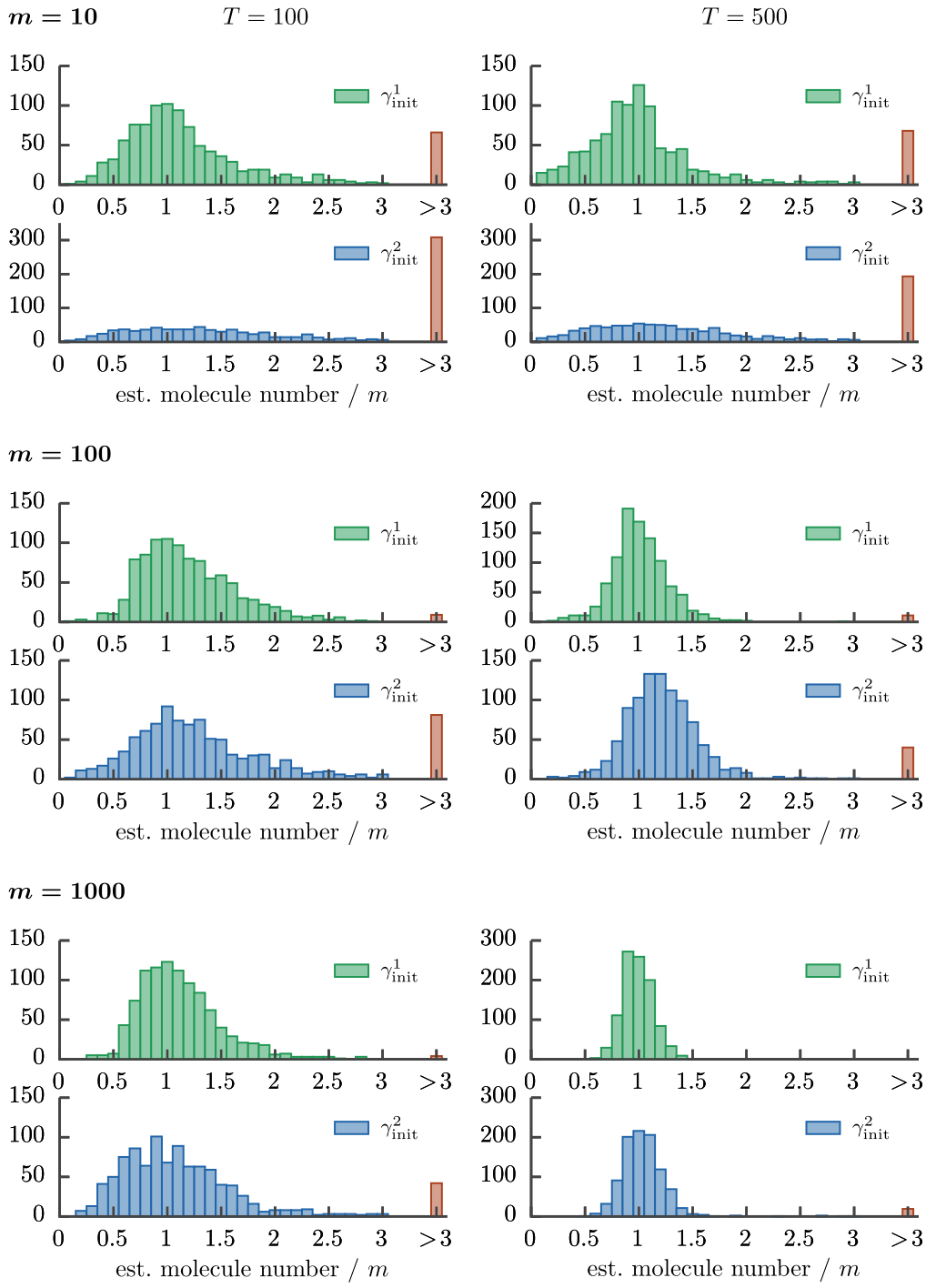

FIG. 7. *Estimated molecule number for simulated fluorescence traces of length* $T = 100$ (*left column*) *as well as* $T = 500$ (*right column*). *For each* $m \in \{10, 100, 1000\}$ *and initial parameters* γ^1_{init} (*true values, green*) *or* γ^2_{init} (*"wrong" values defined by* (5.4), *blue*), *we show histograms obtained from the empirical distribution of the pseudo log-likelihood estimator for* 1000 *independent repetitions. Since the estimator sometimes failed to converge to reasonable solutions* (*usually resulting in exploding values for* m), *we collect estimates* >3 m *in the red bar on the right of each graph.*

multi-exponential decay $\sum_{x\in\mathcal{S}} \alpha_x \lambda_x^{t-1}$ are hard to identify even under generous conditions; let alone from few noisy molecules.

We want to emphasize that the estimation scheme introduced here is preliminary and can be improved in several ways. First, there are various (local) optimization algorithms that typically perform superior to the basic Nelder-Mead method we applied. For example, making use of gradients of the pseudo log-likelihood, which can be calculated explicitly, could greatly improve the performance and runtime of the estimation process. Furthermore, our results indicate that prior knowledge—in form of suitable initial parameters or prior distributions on the parameters—have beneficial effects on the estimation re-

sults. Indeed, many of the failed estimations arise due to convergence to minima far from the true parameters. The adoption of Bayesian methods therefore appears promising to yield more stable and reliable estimators.

6. CASE STUDY: ALEXA 647

In the following, we specialize our general model for the fluorophore Alexa 647, which was used in the experimental work of [30]. Fluorophores of the Alexa series are popular in diverse areas of biomedical research. Due to their properties, like high photostability and brightness, they are amongst the most common choices for super resolution microscopy markers for in vitro cell experiments. Alexa 647 dyes can be used to label a wide variety of molecules, for example, DNA [30] or proteins like IgG antibodies, streptavidin, or transferrin [4]. They have their absorption maximum at 650nm and their emission maximum at 671nm.

Markov model

We focus on the inner and outer models that are illustrated in Figure 3. According to the statistical fluorophore model $\mathcal{F}^{\mathrm{s}}$ established in Section 3, we need several components in order to describe the full fluorophore behavior: the inner and outer transition matrices M^{s} and M^{ℓ}, represented by the values $Q = (q_{xz})$, and the photon statistics $p_{00}(y)$ and $p_{10}(y)$. The outer model depicted in Figure 3 has $r = 3$ dark states, and the respective transition matrix is given by

$$M^{\ell} = \begin{pmatrix} 1 & q_{01} & q_{02} & 0 \\ 0 & q_{11} & q_{12} & 0 \\ 0 & q_{21} & q_{22} & 0 \\ 0 & q_{31} & q_{32} & 1 \end{pmatrix}.$$

The inner model, in contrast, is given through a Markov chain with inner states S_0, S_1, and D_3. We furthermore include an "exit" state E (see Remark 7) in our description, which indicates the transition to one of the dark states of the outer model. This is the only possibility for the fluorophore to leave the bright state during exposure. The short-time matrix M^{s} for $r = 3$ is

$$M^{\mathrm{s}} = \begin{pmatrix} q_{00} & 0 & 0 & 0 \\ q_{10} & 1 & 0 & 0 \\ q_{20} & 0 & 1 & 0 \\ q_{30} & 0 & 0 & 1 \end{pmatrix}.$$

Photons are emitted in this inner model if the fluorophore makes the transition from the excited singlet state S_1 to the ground state S_0.

Note that this model for Alexa 647 is not an exact representation of the quantum mechanical state diagram of the fluorophore, which would require the inclusion of a high number of states that each possess vibrational and rotational substates. However, the full details of the quantum mechanical energy landscape are unknown for most fluorophores, and, according to Remark 1, they also do not have to be known, since states that live on similar timescales can be identified. In fact, it appears from the empirical study in [30] that the description provided above captures the essential features of the observed behavior of Alexa 647. Furthermore, this Markov model—or slight modifications thereof—should be appropriate to model other fluorophores, too.

Photon statistics

We now derive the photon statistics p_{00} and p_{10} for a single exposure. Our derivation is not based on a rigorous treatment of the inner Markov chain in Figure 3 but on a reasonable approximation. For convenience, we will use the symbol Y to refer to a random variable that has distribution $Y_t \mid X'_t = 0$ in the following. This means that Y will *not* refer to the single fluorophore process $(Y_t)_{t=1}^{T}$ for the duration of this section.

Let us call a maximal uninterrupted sequence of transitions between S_0 and S_1 a "burst." A burst is ended by a transition to D_3. This leads to a geometric distribution: do the loop $\mathrm{S}_0 \to \mathrm{S}_1 \to \mathrm{S}_0$ until failure $\mathrm{S}_1 \to \mathrm{D}_3$. Calling the probability of failure $p \in (0, 1)$, it is clear that the number of loops, and hence the number of photons in this burst, has a geometric distribution with parameter $1 - p$. During each exposure interval, there will be a number of B bursts such that the total number of photons Y is a sum of B independent geometrically distributed random variables. This leads to a negative binomial distribution with parameters B and $1 - p$,

$$Y \sim \mathrm{NegBin}(B, 1 - p). \tag{6.1}$$

The number of bursts B is a random quantity. To determine its distribution, first consider the case that occurs when the bright state can never be left. Then the distribution of B would be approximately Poissonian: dividing the exposure interval into many small intervals, each much longer than a typical burst but much smaller than the exposure time,[3] there is a small probability for a burst in each interval and a large number of intervals such that one is in the Poisson limit of the binomial distribution. When taking into account the transition to the exit state E, there is a nonzero probability for a "failure" (i.e., exiting) before each burst. Hence, every burst can be viewed as one successful trial, and the bursts continue until either the exit state or the end of the exposure time is reached. Therefore, the number of bursts is given by the minimum of a Poissonian and a geometric random variable that are independent of each other,

$$\begin{aligned} B &= \min(Z, Q), \\ Z &\sim \mathrm{Poisson}(\mu), \\ Q &\sim \mathrm{Geom}(1 - q), \end{aligned} \tag{6.2}$$

[3] This is possible due to the large difference in transition rates between $\mathrm{S}_0 \to \mathrm{S}_1$ and $\mathrm{D}_3 \to \mathrm{S}_0$.

with parameters $\mu > 0$ and $0 < q < 1$. The transition to the exit state happens if and only if a failure happened before the end of the exposure time. This is the case when $Z > Q$. The following result, proven in Appendix B, provides a connection between the parameters q and μ of the photon distribution and the parameter q_{00} of the transition matrix M^{s}.

LEMMA 5. *The Alexa* 647 *fluorophore stays in the bright state during exposure with probability*

$$q_{00} = \mathbb{P}(Z \le Q) = e^{-(1-q)\mu}. \tag{6.3}$$

In Appendix B, we also derive the moment generating function G of Y. It can be expressed via the moment generating function G^B of the number B of bursts.

LEMMA 6. *The moment generating function of Y is given by*

$$G(\xi) = G^B\left(\log\left(\frac{1-p}{1-pe^{\xi}}\right)\right), \tag{6.4}$$

here

$$\begin{aligned} G^B(\xi) &= e^{-(1-q)\mu} e^{-(1-e^{\xi})q\mu} \\ &\quad + \left(1 - e^{(1-q)\mu}\right)\frac{1-q}{1-qe^{\xi}} \\ &\quad \times \frac{1 - e^{-(1-qe^{\xi})\mu}}{1 - e^{(1-q)\mu}} \\ &= q_{00} G^B_{00}(\xi) + (1-q_{00}) G^B_{10}(\xi). \end{aligned} \tag{6.5}$$

The functions G^B_{00} and G^B_{10} denote the generating functions of B conditioned on $Z \le Q$ and $Z > Q$, respectively.

Similar results hold for the conditional generating functions G_{00} and G_{10}, which are expressible by G^B_{00} and G^B_{10} in a likewise fashion. Therefore, the photon statistics p_{00} and p_{10} are completely specified in terms of the inner parameters p, q, and μ by equations (6.4) and (6.5).

Moments

In order to understand the effects of our inner model for the second-order description of the fluorophore, we look at the parameters θ_1, θ_2, and θ_3 as introduced in (3.10). By differentiating G, one can show the following (see Appendix B).

LEMMA 7. *In the Alexa* 647 *inner model, the moment parameters are given by*:

$$\theta_1 = \frac{p}{1-p}\frac{q}{1-q}(1-q_{00}), \tag{6.6a}$$

$$\theta_2 = -\frac{q_{00}\log q_{00}}{1-q_{00}}, \tag{6.6b}$$

$$\theta_3 = \frac{2}{1-q_{00}}\left(\frac{1-q}{q} - \theta_2 + 1\right) - 1. \tag{6.6c}$$

Surprisingly, according to equations (6.6b) and (6.3), the parameter θ_2 only depends on the internal parameters via $q_{00} = e^{-(1-q)\mu}$. This consideration shows that even though the inner model in Figure 3 depends on three independent parameters (p, q, and μ), only two free parameters, namely θ_1 and θ_3, remain in the second-order description. This must be taken into account when formulating and conducting the optimization routine for the pseudo log-likelihood (5.2) used to estimate the model parameters.

Invariance under thinning

There is another remarkable feature of this inner model choice that deserves to be highlighted. In Section 4, we derived how thinning—the independent loss of photons with a certain probability—affects the parameters $\theta = (\theta_1, \theta_2, \theta_3)$, and we concluded that θ_1 is transformed to $\theta_1' = p_{\mathrm{d}}\theta_1$ while θ_2 and θ_3 are left untouched. Interestingly, we can make a much stronger statement for our model of Alexa 647. We refer to Appendix B for the respective calculations.

LEMMA 8. *The parametric family of the distribution Y for Alexa* 647 *is left invariant by thinning. More precisely, the thinned process $Y' \sim \mathrm{Bin}(Y, p_{\mathrm{d}})$ obeys equations* (6.1) *and* (6.2) *defining Y if the parameter p is replaced by*

$$p' = \frac{pp_{\mathrm{d}}}{1 - p + pp_{\mathrm{d}}},$$

while q and μ remain the same.

7. OUTLOOK

Optical nanoscopy has evolved into a scientific junction point that drives cutting-edge research in disciplines as diverse as optics, biochemistry, and statistics. The work we presented in this article contributes to this development by providing a new way to statistically model the temporal activity pattern of fluorophores, which form the basis for fluorescence super-resolution microscopy. More than that, however, our work is meant to help expose the wide spectrum of worthwhile statistical, computational, and mathematical questions that are raised by current developments in this area.

One point of immediate interest is a better understanding of the photon emission process and its approximation by a Gaussian process. Besides heuristic hints—like the practical success for the purpose of estimating the molecule number—we have not yet found analytical guarantees of how well the Gaussian approximation captures the essential properties of the original HTMM investigated in this article, or if other methods of estimation could offer improvements. Furthermore, we lack statistical results for the estimation process via maximum likelihood estimation of the pseudo log-likelihood. In this light,

the derivation of (asymptotic) statements and confidence bounds for Gaussian processes that are constrained like in our case pose an interesting challenge. This kind of insights would also help address another prevalent question: to what extent is it possible to infer the HTMM model parameters in situations where not all fluorophores are bright in the beginning of the experiment, meaning $\nu_0 < 1$? Conclusive results in this regard are still pending. Reliable inference in this scenario, however, could be an important step towards more flexibility in the design of quantitative super-resolution experiments. A different aspect that is emphasized by our observations in Section 5, but that we only touched upon in passing, is the beneficial effect of prior knowledge for the estimation performance. This underscores the crucial role that Bayesian estimation methods routinely play for the statistical modeling of biological system. For a simplified, binomial model to count fluorophores, the advantage of Bayesian approaches was recently demonstrated in [45].

Another set of problems that our work calls attention to, especially [30], is the modeling of dependencies in biomolecular systems. Even though the description of complex molecules via Markov chains has been proven to be highly successful, the appropriate statistical description of interactions between them, and the resulting influence on measured data in experiments, is largely unresolved. Future research may focus on this question and investigate dependency mechanisms for hidden Markov models, like the proposed HTMM, which are simple enough for analytical investigation but are still able to capture dependency structures suggested by experimental observations.

ACKNOWLEDGMENTS

This work was in part supported by the German Science Foundation (DFG) through grant CRC 755 "Nanoscale Photonic Imaging", projects A4 and A6, RTG 2088 "Discovering Structure in Complex Data", and Germany's Excellence Strategy EXC 2067/1-390729940. We are furthermore grateful to Mira Jürgens for computational assistance and proofreading.

SUPPLEMENTARY MATERIAL

Supplement to "Statistical Molecule Counting in Super-Resolution Fluorescence Microscopy: Towards Quantitative Nanoscopy" (DOI: 10.1214/19-STS753SUPP; .pdf). Contains a table that summarizes the notation, omitted proofs, a brief discussion of CCD cameras, as well as auxiliary results on the lumpability of Markov chains and Eigenvalues of stochastic matrices.

REFERENCES

[1] ASPELMEIER, T., EGNER, A. and MUNK, A. (2015). Modern statistical challenges in high-resolution fluorescence microscopy. *Annu. Rev. Stat. Appl.* **2** 163–202.

[2] BAKSHI, S., SIRYAPORN, A., GOULIAN, M. and WEISSHAAR, J. C. (2012). Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Mol. Microbiol.* **85** 21–38. https://doi.org/10.1111/j.1365-2958.2012.08081.x

[3] BALZAROTTI, F., EILERS, Y., GWOSCH, K. C., GYNNÅ, A. H., WESTPHAL, V., STEFANI, F. D., ELF, J. and HELL, S. W. (2017). Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. *Science* **355** 606–612.

[4] BERLIER, J. E., ROTHE, A., BULLER, G., BRADFORD, J., GRAY, D. R., FILANOSKI, B. J., TELFORD, W. G., YUE, S., LIU, J. et al. (2003). Quantitative comparison of long-wavelength Alexa Fluor dyes to Cy dyes: Fluorescence of the dyes and their bioconjugates. *J. Histochem. Cytochem.* **51** 1699–1712.

[5] BERNING, S., WILLIG, K. I., STEFFENS, H., DIBAJ, P. and HELL, S. W. (2012). Nanoscopy in a living mouse brain. *Science* **335** 551–551.

[6] 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.

[7] BORN, M. and WOLF, E. (1999). *Principles of Optics*, 7th ed. Cambridge Univ. Press, Cambridge.

[8] BRAKEMANN, T., STIEL, A. C., WEBER, G., ANDRESEN, M., TESTA, I., GROTJOHANN, T., LEUTENEGGER, M., PLESSMANN, U., URLAUB, H. et al. (2011). A reversibly photoswitchable GFP-like protein with fluorescence excitation decoupled from switching. *Nat. Biotechnol.* **29** 942–947.

[9] CHEN, C., ZONG, S., WANG, Z., LU, J., ZHU, D., ZHANG, Y. and CUI, Y. (2016). Imaging and intracellular tracking of cancer-derived exosomes using single-molecule localization-based super-resolution microscope. *ACS Appl. Mater. Interfaces* **8** 25825–25833.

[10] CHOJNACKI, J., STAUDT, T., GLASS, B., BINGEN, P., ENGELHARDT, J., ANDERS, M., SCHNEIDER, J., MÜLLER, B., HELL, S. W. et al. (2012). Maturation-dependent HIV-1 surface protein redistribution revealed by fluorescence nanoscopy. *Science* **338** 524–528.

[11] D'ESTE, E., KAMIN, D., GÖTTFERT, F., EL-HADY, A. and HELL, S. W. (2015). STED nanoscopy reveals the ubiquity of subcortical cytoskeleton periodicity in living neurons. *Cell Rep.* **10** 1246–1251.

[12] EGNER, A., GEISLER, C., VON MIDDENDORFF, C., BOCK, H., WENZEL, D., MEDDA, R., ANDRESEN, M., STIEL, A. C., JAKOBS, S. et al. (2007). Fluorescence nanoscopy in whole cells by asynchronous localization of photoswitching emitters. *Biophys. J.* **93** 3285–3290.

[13] EILERS, Y., TA, H., GWOSCH, K. C., BALZAROTTI, F. and HELL, S. W. (2018). MINFLUX monitors rapid molecular jumps with superior spatiotemporal resolution. *Proc. Natl. Acad. Sci. USA* **115** 6117–6122.

[14] FELLER, W. (2008). *An Introduction to Probability Theory and Its Applications*, **2**. Wiley, New York, NY.

[15] FÖLLING, J., BOSSI, M., BOCK, H., MEDDA, R., WURM, C. A., HEIN, B., JAKOBS, S., EGGELING, C. and HELL, S. W. (2008). Fluorescence nanoscopy by ground-state depletion and single-molecule return. *Nat. Methods* **5** 943–945.

[16] FRAHM, L., KELLER-FINDEISEN, J., ALT, P., SCHNORRENBERG, S., RUIZ, M. D. Á., ASPELMEIER, T., MUNK, A., JAKOBS, S. and HELL, S. W. (2019). Molecular contribution function in RESOLFT nanoscopy. *Opt. Express* **27** 21956–21987. https://doi.org/10.1364/OE.27.021956

[17] GOODMAN, J. W. (1996). *Introduction to Fourier Optics*, 2nd ed. McGraw-Hill, New York.

[18] GROTJOHANN, T., TESTA, I., LEUTENEGGER, M., BOCK, H., URBAN, N. T., LAVOIE-CARDINAL, F., WILLIG, K. I., EGGELING, C., JAKOBS, S. et al. (2011). Diffraction-unlimited all-optical imaging and writing with a photochromic GFP. *Nature* **478** 204–208.

[19] HARREMOËS, P., JOHNSON, O. and KONTOYIANNIS, I. (2010). Thinning, entropy, and the law of thin numbers. *IEEE Trans. Inform. Theory* **56** 4228–4244. MR2807322 https://doi.org/10.1109/TIT.2010.2053893

[20] HARTMANN, A., HUCKEMANN, S., DANNEMANN, J., LAITENBERGER, O., GEISLER, C., EGNER, A. and MUNK, A. (2016). Drift estimation in sparse sequential dynamic imaging, with application to nanoscale fluoresence microscopy. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 563–587. MR3506793 https://doi.org/10.1111/rssb.12128

[21] HELL, S. W. (2008). Microscopy and its focal switch. *Nat. Methods* **6** 24–32.

[22] HELL, S. W., SAHL, S. J., BATES, M., ZHUANG, X., HEINTZMANN, R., BOOTH, M. J., BEWERSDORF, J., SHTENGEL, G., HESS, H. et al. (2015). The 2015 super-resolution microscopy roadmap. *J. Phys. D, Appl. Phys.* **48** 443001.

[23] HELL, S. W. and WICHMANN, J. (1994). Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19** 780–782.

[24] HESS, S. T., GIRIRAJAN, T. P. K. and MASON, M. D. (2006). Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* **91** 4258–4272.

[25] HIRSCH, M., WAREHAM, R. J., MARTIN-FERNANDEZ, M. L., HOBSON, M. P. and ROLFE, D. J. (2013). A stochastic model for electron multiplication charge-coupled devices—from theory to practice. *PLoS ONE* **8** e53671. https://doi.org/10.1371/journal.pone.0053671

[26] HOFMANN, M., EGGELING, C., JAKOBS, S. and HELL, S. W. (2005). Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins. *Proc. Natl. Acad. Sci. USA* **102** 17565–17569.

[27] 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

[28] KOENIG, M., BERLAGE, C., REISCH, P., OELSNER, C., KOBERLING, F., TA, H. and ERDMANN, R. (2019). Molecular counting by photon statistics in confocal fluorescence imaging. *Biophys. J.* **116** 134a–135a.

[29] KOMIS, G., MISTRIK, M., ŠAMAJOVÁ, O., OVEČKA, M., BARTEK, J. and ŠAMAJ, J. (2015). Superresolution live imaging of plant cells using structured illumination microscopy. *Nat. Protoc.* **10** 1248–1263.

[30] LAITENBERGER, O., ASPELMEIER, T., GEISLER, C., MUNK, A. and EGNER, A. (2019). Towards unbiased molecule counting in superresolution fluorescence microscopy. Preprint.

[31] LAPLANTE, C., HUANG, F., TEBBS, I. R., BEWERSDORF, J. and POLLARD, T. D. (2016). Molecular organization of cytokinesis nodes and contractile rings by super-resolution fluorescence microscopy of live fission yeast. *Proc. Natl. Acad. Sci. USA* **113** E5876–E5885.

[32] 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.

[33] LIN, Y., LONG, J. J., HUANG, F., DUIM, W. C., KIRSCHBAUM, S., ZHANG, Y., SCHROEDER, L. K., REBANE, A. A., VELASCO, M. G. M. et al. (2015). Quantifying and optimizing single-molecule switching nanoscopy at high speeds. *PLoS ONE* **10** e0128135.

[34] MAGLIONE, M. and SIGRIST, S. J. (2013). Seeing the forest tree by tree: Super-resolution light microscopy meets the neurosciences. *Nat. Neurosci.* **16** 790–797.

[35] MESSINA, T. C., KIM, H., GIURLEO, J. T. and TALAGA, D. S. (2006). Hidden Markov model analysis of multichromophore photobleaching. *J. Phys. Chem. B* **110** 16366–16376.

[36] MURANYI, W., MALKUSCH, S., MÜLLER, B., HEILEMANN, M. and KRÄUSSLICH, H.-G. (2013). Super-resolution microscopy reveals specific recruitment of HIV-1 envelope proteins to viral assembly sites dependent on the envelope C-terminal tail. *PLoS Pathog.* **9** e1003198. https://doi.org/10.1371/journal.ppat.1003198

[37] NELDER, J. A. and MEAD, R. (1965). A simplex method for function minimization. *Comput. J.* **7** 308–313. MR3363409 https://doi.org/10.1093/comjnl/7.4.308

[38] PAGEON, S. V., CORDOBA, S.-P., OWEN, D. M., ROTHERY, S. M., OSZMIANA, A. and DAVIS, D. M. (2013). Superresolution microscopy reveals nanometer-scale reorganization of inhibitory natural killer cell receptors upon activation of NKG2D. *Sci. Signal.* **6** ra62. https://doi.org/10.1126/scisignal.2003947

[39] PATEL, L., GUSTAFSSON, N., LIN, Y., OBER, R., HENRIQUES, R. and COHEN, E. (2019). A hidden Markov model approach to characterizing the photo-switching behavior of fluorophores. *Ann. Appl. Stat.* **13** 1397–1429. MR4019144 https://doi.org/10.1214/19-AOAS1240

[40] PRESCHER, J., BAUMGÄRTEL, V., IVANCHENKO, S., TORRANO, A. A., BRÄUCHLE, C., MÜLLER, B. and LAMB, D. C. (2015). Super-resolution imaging of ESCRT-proteins at HIV-1 assembly sites. *PLoS Pathog.* **11** e1004677. https://doi.org/10.1371/journal.ppat.1004677

[41] ROBBINS, M. S. and HADWEN, B. J. (2003). The noise performance of electron multiplying charge-coupled devices. *IEEE Trans. Electron Devices* **50** 1227–1232.

[42] 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.

[43] RUST, M. J., BATES, M. and ZHUANG, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3** 793–796.

[44] SAHL, S. J., HELL, S. W. and JAKOBS, S. (2017). Fluorescence nanoscopy in cell biology. *Nat. Rev., Mol. Cell Biol.* **18** 685–701. https://doi.org/10.1038/nrm.2017.71

[45] SCHNEIDER, L. F., SCHMIDT-HIEBER, J., STAUDT, T., KRAJINA, A., ASPELMEIER, T. and MUNK, A. (2018). Posterior consistency for n in the binomial (n, p) problem with both parameters unknown—with applications to quantitative nanoscopy. ArXiv Preprint. Available at arXiv:1809.02443.

[46] SCHNORRENBERG, S., GROTJOHANN, T., VORBRÜGGEN, G., HERZIG, A., HELL, S. W. and JAKOBS, S. (2016). In vivo super-resolution RESOLFT microscopy of Drosophila melanogaster. *eLife* **5**. https://doi.org/10.7554/eLife.15567

[47] SHARMA, S., SANTISKULVONG, C., BENTOLILA, L. A., RAO, J., DORIGO, O. and GIMZEWSKI, J. K. (2012). Correlative nanomechanical profiling with super-resolution F-actin imaging

reveals novel insights into mechanisms of cisplatin resistance in ovarian cancer cells. *Nanomedicine* **8** 757–766.

[48] STAUDT, T., ASPELMEIER, T., LAITENBERGER, O., GEISLER, C., EGNER, A. and MUNK, A. (2020). Supplement to "Statistical Molecule Counting in Super-Resolution Fluorescence Microscopy: Towards Quantitative Nanoscopy." https://doi.org/10.1214/19-STS753SUPP.

[49] SYDOR, A. M., CZYMMEK, K. J., PUCHNER, E. M. and MENNELLA, V. (2015). Super-resolution microscopy: From single molecules to supramolecular assemblies. *Trends Cell Biol*. **25** 730–748. https://doi.org/10.1016/j.tcb.2015.10.004

[50] TA, H., KELLER, J., HALTMEIER, M., SAKA, S. K., SCHMIED, J., OPAZO, F., TINNEFELD, P., MUNK, A. and HELL, S. W. (2015). Mapping molecules in scanning far-field fluorescence nanoscopy. *Nat*. *Commun*. **6** 7977. https://doi.org/10.1038/ncomms8977

[51] TSEKOURAS, K., CUSTER, T. C., JASHNSAZ, H., WALTER, N. G. and PRESSÉ, S. (2016). A novel method to accurately locate and count large numbers of steps by photobleaching. *Mol*. *Biol*. *Cell* **27** 3601–3615. https://doi.org/10.1091/mbc.E16-06-0404

[52] VAN DE LINDE, S., LÖSCHBERGER, A., KLEIN, T., HEIDBREDER, M., WOLTER, S., HEILEMANN, M. and SAUER, M. (2011). Direct stochastic optical reconstruction microscopy with standard fluorescent probes. *Nat*. *Protoc*. **6** 991–1009. https://doi.org/10.1038/nprot.2011.336

[53] VOGELSANG, J., STEINHAUER, C., FORTHMANN, C., STEIN, I. H., PERSON-SKEGRO, B., CORDES, T. and TINNEFELD, P. (2010). Make them blink: Probes for super-resolution microscopy. *ChemPhysChem* **11** 2475–2490.

[54] WILLIAMSON, D. J., OWEN, D. M., ROSSY, J., MAGENAU, A., WEHRMANN, M., GOODING, J. J. and GAUS, K. (2011). Pre-existing clusters of the adaptor Lat do not participate in early T cell signaling events. *Nat*. *Immunol*. **12** 655–662.