MICROSCOPY

Toward quantitative super-resolution microscopy: molecular maps with statistical guarantees

Katharina Proksch[1,†], **Frank Werner** ⓘ[2,†,*], **Jan Keller–Findeisen**[3,†], **Haisen Ta**[4] and **Axel Munk**[5,6]

[1]Faculty of Electrical Engineering, Mathematics and Computer Science, Universiteit Twente, Zilverling 2098, Enschede 7500, The Netherlands
[2]Institute of Mathematics, University of Würzburg, Emil-Fischer-Str. 30, Würzburg 97074, Germany
[3]Department of NanoBiophotonics, Max-Planck-Institut für multidisziplinäre Naturwissenschaften, Am Fassberg 11, Göttingen 37077, Germany
[4]Center for Hybrid Nanostructures, Universität Hamburg, Luruper Chaussee 149, Hamburg 22607, Germany
[5]Institute for Mathematical Stochastics, University of Göttingen, Goldschmidtstraße 7, Göttingen 37077, Germany
[6]Felix Bernstein Institute for Mathematical Statistics in the Bioscience, University of Göttingen, Goldschmidtstraße 7, Göttingen 37077, Germany
*To whom correspondence should be addressed. E-mail: frank.werner@uni-wuerzburg.de
†These authors contributed equally

Abstract

Quantifying the number of molecules from fluorescence microscopy measurements is an important topic in cell biology and medical research. In this work, we present a consecutive algorithm for super-resolution (stimulated emission depletion (STED)) scanning microscopy that provides molecule counts in automatically generated image segments and offers statistical guarantees in form of asymptotic confidence intervals. To this end, we first apply a multiscale scanning procedure on STED microscopy measurements of the sample to obtain a system of significant regions, each of which contains at least one molecule with prescribed uniform probability. This system of regions will typically be highly redundant and consists of rectangular building blocks. To choose an informative but non-redundant subset of more naturally shaped regions, we hybridize our system with the result of a generic segmentation algorithm. The diameter of the segments can be of the order of the resolution of the microscope. Using multiple photon coincidence measurements of the same sample in confocal mode, we are then able to estimate the brightness and number of molecules and give uniform confidence intervals on the molecule counts for each previously constructed segment. In other words, we establish a so-called *molecular map* with uniform error control. The performance of the algorithm is investigated on simulated and real data.

Key words: asymptotic normality, counting, family-wise error rate, multiplicity adjustment

Introduction

Super-resolution microscopy

In fluorescence microscopy, structures of interest inside a specimen are labeled with fluorescent markers and then imaged using visible light illumination. Only the fluorescence itself and thus the labeled structures are detected, making it possible, for example, to investigate details inside living cells with unrivaled contrast. The tremendous development of super-resolution fluorescence microscopy in recent decades has extended spatial resolution beyond the diffraction limit of conventional microscopy to the nanometer scale.

All super-resolution light microscopy concepts rely on distinguishing fluorophores locally by consecutively transferring them between a dark (non-fluorescent) and a bright (fluorescent) state using light to induce these transitions [1,2]. The transitions between these states can be performed in either a spatially controlled or stochastic manner, with the latter denoted here as single-molecule switching microscopy [3]. In both approaches, only a small subset of molecules is left in the bright state at each measurement step and the final image is assembled by repeating the experiment many times. A well-established spatially controlled method uses stimulated emission depletion (STED) [4,5]. Thereby, a red-shifted

light spot featuring a central intensity minimum is co-aligned with the excitation light spot. It induces strongly saturated stimulated emission, effectively inhibiting the fluorophores from emitting fluorescence in the periphery of a focused excitation light spot. The very small spot of effectively allowed fluorescence emission can be scanned over the sample, where each scanning position in a rectangular grid corresponds to a pixel in the final image. For example, the STED principle has been used in the past to reveal the distribution of synaptic proteins in living mice [6] or the dynamics of membrane lipids in living cells [7]. Recently, a combination of stochastic switching and excitation light patterns with at least one isolated intensity zero, called minimal photon fluxes, was used to achieve isotropic resolution on the order of a few nanometers [8].

From a statistical perspective, the recovery of spatial intensity and specimen distribution from super-resolution fluorescence microscopy images leads to sophisticated convolution models with Poisson or Binomial data distributions, which in themselves present a number of challenges (see, e.g. [9–12] and references therein).

In many biological contexts, however, it is not only the precise localization of structures that is of interest but also

---

the determination of the exact number of fluorescent markers at a given location, especially if this number can be related to the local number of proteins or other biological targets of interest. Such quantitative knowledge of target structures at the nanoscale has the potential to greatly improve the understanding of many biological processes. Knowledge of the absolute number of molecules can provide the basis for structural models of protein complexes or determine thresholds for the number of molecules required to produce a particular effect. For example, estimating the number of constituent proteins in kinetochores reported unexpectedly high numbers of proteins present [13], whereas quantifying the number of proteins used for flagellar regeneration helped refine models for flagellar assembly [14].

Toward molecule counting

In general, the mean recorded fluorescence signal in a microscope is proportional to the number of active fluorescent markers. If we denote by f_N a spatial function assigning each location (or pixel) the corresponding number of markers, and by f_p their corresponding brightness (i.e. probability to emit a photon after an excitation pulse was applied), then the observed quantity is mathematically given by a spatial convolution of the product $f_N \cdot f_p$ with the so-called point spread function (PSF) h, which is determined by the microscope (see the 'Modeling, notation and prerequisites' section for details). This already shows that the observations cannot be readily used to infer the absolute number of markers, since the brightness of a fluorophore depends on the local environment, and thus, f_p is unknown and not constant. If the microscope could guarantee that each fluorophore was perfectly separated, counting fluorophores would be a trivial task, but in almost all current applications, this is not the case. Robust statistical modeling of the contributions of the fluorophores to the acquired image data and, in particular, a method for calibrating the local brightness of molecules are therefore required. This calibration is ideally performed during the measurement itself, since the molecular brightness also depends on the microscope configuration used and the sample conditions.

In the super-resolution techniques that leave only isolated markers in the bright state (as it is, e.g. the case in single-molecule switching microscopy), counting can be performed by a careful analysis of the localization events (see e.g. [15–18]). From a statistical perspective, this is mainly based on the *temporal* Markovian dynamics of the transition between bright and dark states of the fluorophores [19,20] and requires detailed knowledge about the fluorescent state transition kinetics.

In the case of *spatially* separating super-resolution microscopy, e.g. for STED as considered in this work, a similar approach is not possible. Instead, the single molecule brightness has traditionally been estimated by observing single photon bleaching steps, which is challenging in dense samples. For the case of repeatedly activatable markers, one can also determine an excess variance, which was used to determine the molecule brightness [21,22]. Despite its practical relevance, stochastically sound methods for counting molecules from scanning microscopy images remain elusive till now.

Our approach

In this work, we will address this issue by means of a different path. To this end, we will perform counting based on the photon emission statistics, specifically the number of photons that are emitted simultaneously. This has recently emerged as a tool for intrinsically calibrating the molecular brightness and to infer on the number of molecules present. The physical effect affecting the photon emission statistics is called photon antibunching, i.e. an excited fluorophore cannot emit more than one photon during the lifetime of the excited state. First measurements showed that the number of markers can be inferred from the photon emission statistics [23]. Termed antibunching microscopy, it has been implemented later in STED mode and could, for example, count the number of internalized receptors within small vesicles in HEK293 cells [24]. This technique is able to account for locally varying molecular brightness: given that the event of observing a single photon has a probability proportional to $f_N \cdot f_p$ as argued earlier, the probability that two photons are measured simultaneously must be proportional to $(f_N \cdot f_p)^2$ minus the physically impossible contribution that both photons stem from the same molecule, which is itself proportional to $f_N \cdot f_p^2$. So, after a linear transformation, we obtain observations proportional to $f_N \cdot f_p, f_N \cdot f_p^2, \ldots$, which, in principle, allows decoupling the (local) number of markers f_N from their brightness f_p. For details, see Theorem 1 and its proof in the Supplementary Material. A scheme of an antibunching STED microscope is shown in Fig. 1. It provides simultaneous detections of multiple photons, i.e. the measurement of photon coincidences. State-of-the-art single-photon detectors (avalanche photo diodes) still feature considerable times where they are insensitive after each photon detection (on the order of 50 ns), rendering them incapable of detecting photon coincidences. This limitation is effectively overcome by parallel detection, i.e. by splitting the emitted light equally among multiple detectors.

At each scan position (the scan covers the whole sample equally), a fixed number t of excitation light pulses is applied and the statistics of detected number of photons are recorded. This can be seen as repetitions of a multinomial experiment with an additional twist that multiple photons arriving at the same detector after each excitation light pulse are only registered as a single detection event. The accessible observable is indeed only the number of active detectors for each excitation light pulse, which depends on the number of emitted photons and the number of the available detectors. In the 'Modeling, notation and prerequisites' section, we develop an explicit multinomial model that encodes the photon emission and detection probabilities and is given by convolutions and products of the local brightness, the local number of markers and the effective PSF of the microscope itself. The accuracy of the estimation of the multinomial probabilities of our model increases with increasing number t of excitation light pulses. Therefore, our theoretical considerations assume an asymptotic viewpoint as this number tends to infinity. Typical numbers of t in applications easily reach $t = 10{,}000$, such that the approximation by the asymptotic results is rather accurate. Utilizing this model, measuring multiple photon coincidences allows for inference on the (local)
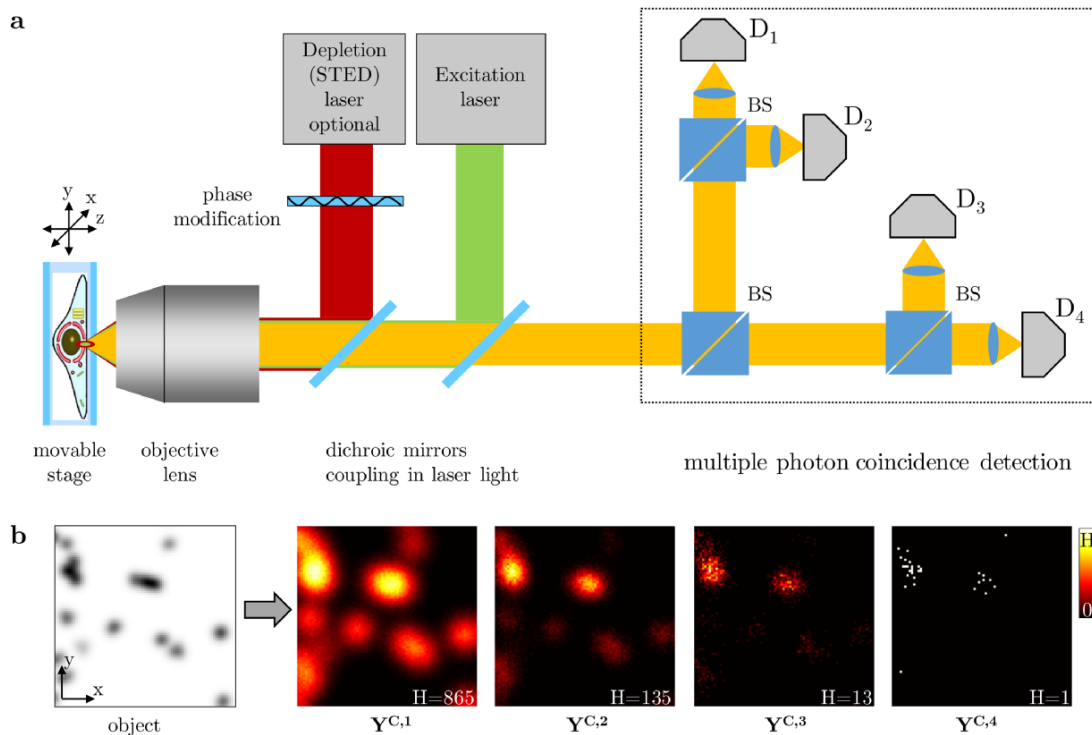
Fig. 1. Microscope scheme and example images for detecting multiple photon coincidences. (a) Schematic of a confocal/STED fluorescence light microscope capable of detecting multiple photon coincidences. Excitation, depletion and fluorescence light are combined with dichroic mirrors. A specific phase distribution is imprinted on the depletion beam in order to create a doughnut-shaped intensity distribution of depletion light at the focal point. Effectively, the spot in the sample that is allowed to fluoresce is reduced to a sub-diffraction extent. In the detection unit, the fluorescence light is split by three beam splitters (BS) in four equal intensity paths and directed on four identical detectors ($D_k, k = 1, .., 4$). Such a microscope has been built and used for antibunching microscopy in [24]. (b) Example object (molecule density) and corresponding simulated photon coincidence measurement images $\mathbf{Y}^{C,k}, k = 1, .., 4$ in confocal mode (excitation pulses per pixel $t = 3000$; see the 'Modeling, notation and prerequisites' section for details on the model and data. The scaling of the colorbar is indicated by the value H in each image.

number of markers, as the simultaneous arrival of two photons implies that the two photons must originate from at least two different fluorophores. We will show later on that the empirical distribution of the number of active detectors depends uniquely on the number and location of molecules present. An estimation of the local molecular number and brightness by a penalized maximum likelihood-based reconstruction algorithm (without statistical guarantees) was implemented in [24]. Besides the lack of statistical guarantees for the local molecular numbers, the statistical model therein is limited to two simultaneously arriving photons at most.

Our contribution

As the goal of counting the number of fluorescent markers in a sample would greatly benefit from both being able to calculate error bounds on the numbers and considering higher-order photon coincidences, in this work, we will extend the existing approach in these directions. First of all, we derive a detailed and sound statistical model for the observations obtained from the antibunching microscope including contributions of arbitrarily high photon coincidences. One major contribution is that instead of globally estimating the number of markers inside the specimen, we construct a so-called *molecular map* with uniform error control. This is a collection of family-wise error rate (FWER) controlled, distinct segments in the specimen together with uniform confidence intervals (CIs) for the number of markers contained in each segment. The diameter

of these segments can be of the order of the resolution of the microscope. To derive this molecular map with a given total error level $\alpha \in (0, 1)$, the method performs three steps:

S1: Segmentation. We first construct a segmentation of the image space such that each segment contains at least one molecule with uniform probability at least $1 - \alpha/2$ and, at the same time, has a suitable shape. All image pixels that are not contained in any of the segments are no longer considered in steps S2 and S3.

S2: Estimation. Given M (say) segments from S1, we estimate (locally) the number of molecules, i.e. segment by segment.

S3: Confidence. Based on a central limit theorem, we construct $1 - \alpha/(2M)$ CIs for each of the M local (segment-wise) numbers of molecules.

The outcome of this procedure is a collection of distinct segments, estimated marker counts in these segments and CIs such that asymptotically (as the number t of excitation light pulses tends to infinity) the following statement is true (Theorem 4):

$$\mathbb{P} \left[\begin{array}{c} \text{Each CI contains} \\ \text{the correct number of} \\ \text{markers in its segment} \end{array} \right] \geq 1 - \alpha. \qquad (1)$$

The provided MATLAB® code implements the approach described earlier, and we investigate its performance in numerical simulations as well as real data examples in the 'Numerical study' and 'Counting molecules in a DNA origami measurement' sections.

Let us discuss some immediate issues and the rationale behind this three-step approach. In principle, the measured photon coincidences do allow for a pixel-wise estimation of the number of markers and hence also for pixel-wise confidence statements. This is also possible with the aforementioned three-step approach by considering each pixel as a separate segment. However, to obtain a uniform coverage as in (1), the corresponding CIs (e.g. as constructed in step S3) will then be unfavorably large due to the large number of pixels (in practical applications on the order of 10^5–10^6) and its corresponding multiplicity correction. To overcome this burden, we exploit the fact that in many samples, the molecules are concentrated in parts of the measurement volume and our algorithm estimates on relevant regions only. Consequently, it neglects a substantial fraction of all pixels before performing the estimation. If the segments are still chosen reasonably small (e.g. in the order of the microscope resolution rather then single pixels), local information is maintained, while at the same time, the number of confidence statements to be made is reduced significantly. Therefore, a reasonable segmentation in step S1 will strongly ease the estimation and confidence procedure afterwards. Such a segmentation will be achieved by a hybridization of an established segmentation approach from image processing such as, e.g. the watershed segmentation, with a hot spot detection procedure that provides rigorous statistical guarantees such as the Multiscale Inverse SCAnning Test (MISCAT) procedure introduced in [25], which is a method based on multiple statistical hypothesis testing. Such a hybrid approach to segmentation is new and offers a lot of flexibility in the choice of methods while providing a strong theoretical justification. Note that even though the number M of segments in step S2 is random, this does not cause problems in step S3, as it can be considered deterministic conditional on the data used for step S1. Moreover, in our asymptotic considerations, the number of pixels, which is an upper bound for M, is fixed.

The three-step approach is visualized in Fig. 2. In the example shown there, the images have $512 \times 512 = 262,144$ pixels, but M, the number of identified segments containing markers, is only 18.

We investigate the introduced methodology both in simulations and on experimental data. The former gives a precise description of the abilities and limitations, and the latter shows the high practical value of our approach.

The remainder of the paper is organized as follows: in the 'Modeling, notation and prerequisites' section, we state the statistical model for raw photon counts in antibunching microscopy and their relation to the local number of markers inside the sample. The technical derivation of this model is deferred to Section B.1 of the Supplementary Material. 'Segmentation' and 'Statistical molecule counting' sections are devoted to the three steps of our inference algorithm. Detailed numerical simulations are provided in the 'Numerical study' section. In the 'Counting molecules in a DNA origami measurement' section, we discuss the applicability on real-world data. We end with a short conclusion in the 'Discussion and outlook' section.

Modeling, notation and prerequisites

Throughout this work, we denote by $[n]$ the set $\{1,\dots,n\}$. Vectors and multi-indices will be written in bold face, where, with a slight abuse of notation, we denote the vector of component-wise ratios i_k/n_k by $\mathbf{i/n}$.

In the following, we denote by N the total (unknown) number of fluorescent markers in our specimen of interest. Without loss of generality, we can assume that all markers are contained in the unit square $[0,1]^2$, which is discretized by the grid $\mathbf{x}_{(i_1,i_2)} = \mathbf{x_i} = \left(\frac{i_1}{n}, \frac{i_2}{n}\right)$, $\mathbf{i} \in [n] \times [n]$. The jth marker has a position $\mathbf{x}_{\psi(j)}$ on this grid (where the true position is just assigned to the closest grid point), which is encoded in terms of the mapping $\psi : [N] \to [n] \times [n]$ that assigns to the j-th marker, $j \in [N]$, its position on the grid. Furthermore, the j-th marker has an individual brightness p_j. The individual brightness of a marker j is its probability to collect a photon after an excitation light pulse focused to $\mathbf{x}_{\psi(j)}$ was applied, see, e.g. [9].

The antibunching microscope shown in Fig. 1 works as follows. The specimen is scanned spatially along the grid $\mathbf{x_i}, \mathbf{i} \in [n] \times [n]$. For each grid point $\mathbf{x_i}$, the following experiment is repeated a total number of t times: A short excitation pulse with a duration much shorter than the excited state's lifetime is applied to the sample (focused on the current grid point $\mathbf{x_i}$), and afterwards the fluorescence is recorded at the detectors. Between two experiments, there is a certain waiting time, which takes into account the dead time of the detectors after recording a photon (typically around 100 ns) and the typical fluorescence lifetime of the markers (typically ≤ 10 ns). In each experiment, the number of detected photons, k, is recorded.

It is important to note that a pulse centered at the point $\mathbf{x}$ will also illuminate neighboring grid points due to diffraction, however, with a lower intensity. This implies that the probability $\mathfrak{p}_j(\mathbf{x})$ to detect an emission from the j-th marker in a single excitation pulse when scanning at position $\mathbf{x} \in [0,1]^2$ is given by

$$\mathfrak{p}_j(\mathbf{x}) = p_j \cdot h(\mathbf{x} - \mathbf{x}_{\psi(j)}). \qquad (2)$$

Here, h is the PSF of the microscope, which reaches its maximal value at 0, and therefore, a detection at the correct position is most likely. However, relation (2) shows that markers collectively contribute to the measured signal at a specific scanning position $\mathbf{x}$, even though they are located at different grid points. The shape of h depends on various experimental parameters in the microscope. As a rough guidance, h has a larger extent when using conventional (i.e. not super-resolved) microscopy (as more surrounding markers are also excited) and a smaller extent in the case of STED measurements corresponding to a smaller or larger resolution. An often used measure describing the shape of h and also of the microscope's resolution is its full width at half maximum (FWHM), cf. Kulaitis *et al.* [12] for an explanation in statistical terms. In the case of a Gaussian peak PSF h with variance σ^2, one has FWHM $= 2\sqrt{2\log 2}\sigma$.

As a consequence of (2), we obtain by superposition a convolution model

$$g(\mathbf{x}) = \sum_{j=1}^N \mathfrak{p}_j(\mathbf{x}) = \int_{[0,1]^2} f_N(\mathbf{y}) f_p(\mathbf{y}) h(\mathbf{x} - \mathbf{y}) \, d\mathbf{y}, \qquad (3)$$
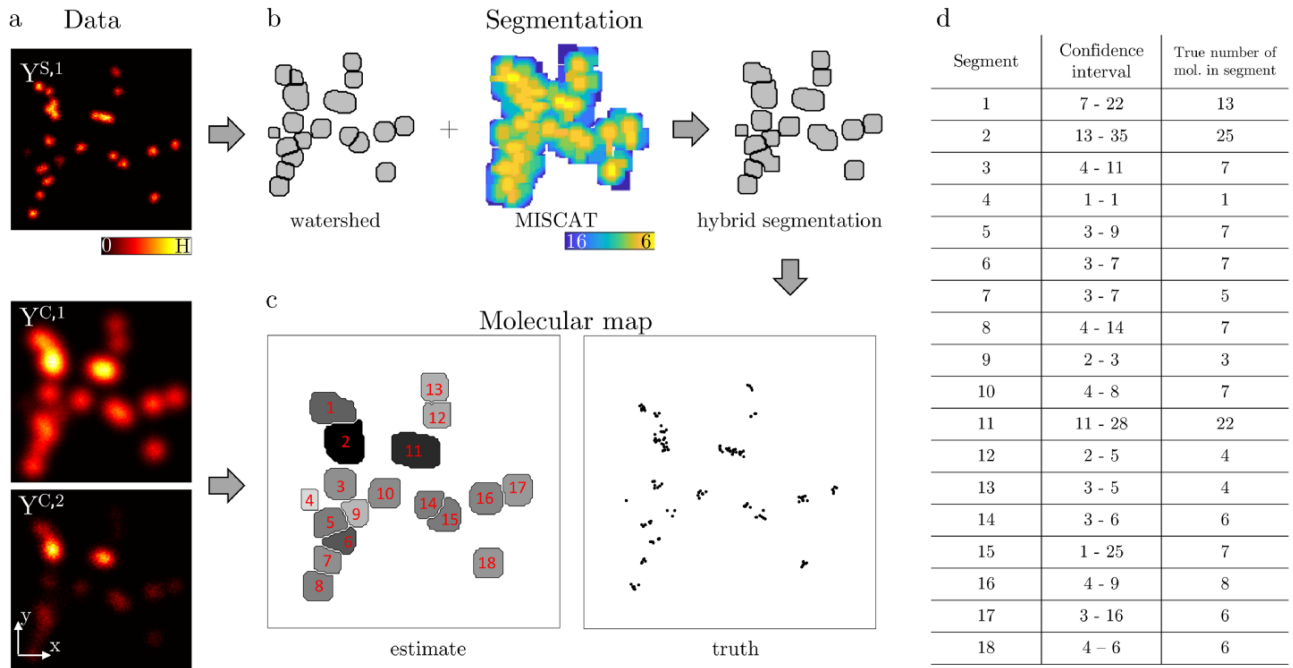
Fig. 2. Three-step analysis workflow. (a) Simulated one-photon STED image with high spatial resolution (top left) and one- and two-photon detection images in confocal mode (center left and bottom left) for the example object depicted in Fig. 1b. (b) The STED image Y_1^S is used for a combined watershed and MISCAT segmentation that results in an effective hybrid segmentation, where each segment contains at least one molecule with uniform probability. Color scale for MISCAT is inverse to the box area, and smaller boxes are drawn on top of larger boxes. (c) Using the one- and two-photon statistics obtained in confocal mode $Y^{C,1}$, $Y^{C,2}$ and the hybrid segmentation in (b), the number of molecules and confidence bounds can be obtained for each segment. The segments are enumerated, and the fill color (on a gray scale) indicates the estimated local molecule density within a segment. The true location of markers is shown on the right. (d) Segment number, confidence bounds ($\alpha = 0.1$) and true number of molecules for each of the 18 obtained segments.

where $g(\mathbf{x})$ denotes the probability to detect at least one photon when scanning at $\mathbf{x} \in [0,1]^2$, $f_N : [0,1]^d \to \mathbb{R}$ denotes the spatial number of markers, and $f_p : [0,1]^2 \to \mathbb{R}$ their brightness at $\mathbf{x} \in [0,1]^2$. Note that formally f_N is a (finite) sum of dirac measures (located at the molecule positions) rather than a function, and hence, the integral in (3) is in fact just a sum. The same applies to all integrals involving f_N. As an alternative, one could model f_N as a density as well, which underlies all classical (continuous) convolution models in microscopy. Using the notation in (3), we stay in line with these models, and essentially, all subsequent derivations work in both models.

The convolution model in (3) is most commonly used for standard fluorescence microscopes to recover the product $f_N \cdot f_p$, see e.g. [9]. However, our modified microscope sketched in Fig. 1 does not only measure the total number of photons when scanning at $\mathbf{x_i}$, $\mathbf{i} \in [n] \times [n]$ but furthermore measures the number of coincidences, i.e. k detected photons at the same time for $0 \leq k \leq m_d$, where m_d denotes the number of parallel detector units present in the microscope (Fig. 1). This allows us to finally decouple f_N and f_p. The data for the current grid point $\mathbf{x_i}$ consequently consist of integers, with $Y_\mathbf{i}^k$ being the number of k-photon events at $\mathbf{x_i}$, $0 \leq k \leq m_d$. This immediately implies that our (ideal) data will consist of (pixel-wise) multinomial observations, i.e.

$$\mathbf{Y_i} = \left(Y_\mathbf{i}^0, ..., Y_\mathbf{i}^{m_d}\right) \sim \mathcal{M}\left(t, E_0(\mathbf{x_i}), ..., E_{m_d}(\mathbf{x_i})\right), \quad \mathbf{i} \in [n] \times [n], \tag{4}$$

where $\mathcal{M}$ denotes the multinomial distribution with a total number of t experiments per image pixel and the numbers

$E_k(\mathbf{x_i})$ denote the probabilities to observe k photons at position $\mathbf{i}$. Note, that in the notation of (3), we have $g(\mathbf{x_i}) = \sum_{k \geq 1} E_k(\mathbf{x_i})$, and due to multinomial distribution, it holds $t = \sum_{k=0}^{m_d} Y_i^k$ for each pixel $\mathbf{i} \in [n] \times [n]$.

Throughout this work, we assume that the observations for different grid points $\mathbf{x_i}$ are independent, which means that we assume that the observations $\mathbf{Y_i}, \mathbf{i} \in [n]^2$ are independent, multinomially distributed random variables. This is a reasonable assumption as long as markers do not come so close to each other that they interact. Note that model (4) does not incorporate external noise sources such as detector read-out errors. However, such statistical thinning as it might be caused by loss of photons in the detectors can be included by redefining the brightness p_j of the individual molecules, cf. [11]. In conclusion, the model (4) can be considered as highly accurate if the number of repetitions per pixel, t, is of the order of several thousands, which corresponds to typical experimental conditions.

A detailed model for the probabilities $E_k, k = 0, ..., m_d$, in dependence on the unknown quantities of interest such as the local number of markers f_N and their brightness f_p, is developed in the Supplementary Material. To describe it briefly, let us introduce the function $\mathbf{s} : [0,1]^2 \to \mathbb{R}^{m_d}$, defined by

$$\mathbf{s}(\mathbf{x_i}) = \left(\sum_{j=1}^{N}(\mathfrak{p}_j(\mathbf{x_i}))^k\right)_{k=1,...,m_d}$$

$$= \left(\int_{[0,1]^2} f_N(\mathbf{x}) f_p(\mathbf{x})^k h(\mathbf{x} - \mathbf{x_j})^k \, d\mathbf{x}\right)_{k=1,...,m_d}. \tag{5}$$

Then, **s** can be related to the probabilities $\mathbf{E} = (E_k)_{k=0,...,m_d}$ – which can themselves be estimated from the available observations—as stated in the following theorem.

Theorem 1. (Statistical model for the antibunching microscope) Let m_d denote the number of detectors, N the total number of fluorescent markers in the specimen and $\kappa = \kappa(\mathbf{x})$ the number of emitted photons in one excitation pulse at position $\mathbf{x}$. Assume that $N > m_d$. Let $\mathbf{E}(\mathbf{x_i}) := (E_0(\mathbf{x_i}),...,E_{m_d}(\mathbf{x_i}))$ denote the multinomial probabilities defined in model (4). Assume that, for any $\mathbf{x} \in [0,1]^2$, there exists a $\gamma = \gamma(\mathbf{x}) \in (0,1)$ such that for any $\kappa > m_d$ it holds that

$$\mathbb{P}\left(\begin{array}{c} \text{more than } \kappa(\mathbf{x}) \text{ photons} \\ \text{emitted at position } \mathbf{x} \\ \text{during one pulse} \end{array} \right) \leq \gamma(\mathbf{x})^{\kappa}. \qquad (6)$$

Then, there exists an explicitly known differentiable, invertible map $\mathbf{T} : [0,1]^{m_d} \to \mathbb{R}^{m_d}$ such that

$$\mathbf{E}(\mathbf{x_i}) = \mathbf{T}(\mathbf{s}(\mathbf{x_i})) + O\left(\gamma^{m_d+1}\right) \quad \text{as } \gamma \to 0 \qquad (7)$$

for all $\mathbf{i} \in [n] \times [n]$.

For an explicit formula for $\mathbf{T}$, see the proof of Theorem 1. Note that assumption (6) will be satisfied as soon as the local number of molecules around a position $\mathbf{x}$ and/or their brightness is not too large compared to the number of detectors m_d. As a consequence, for sufficiently large m_d, we can neglect the remainder term in (7) and obtain (approximately) a non-linear forward model

$$\mathbf{E} = (E_k)_{0 \leq k \leq m_d} = F\left(f_N, f_p\right), \qquad (8)$$

with F given by the concatenation of $\mathbf{T}$ and the mapping $(f_N, f_p) \mapsto \mathbf{s}$ (recall (5)), from which we aim to determine local information on the spatial number f_N (and the spatial brightness f_p) of molecules from empirical measurements **Y** as in (4). This can be seen as a nonlinear inverse problem, which is also ill-posed due to convolution with the PSF h. Note that the model (8) does not take background contributions into account, e.g. from out-of-focus planes. This is for simplicity mostly, as the following considerations can immediately be adjusted to a spatially varying background intensity λ, which, however, leads to even more technical results. In principle, this would allow us to determine not only f_N and f_p from the available data but also λ. In practice, this corresponds to a highly underdetermined problem, which is why typically the background intensity is pre-estimated and then used to correct the data. To keep the theory concise, we have therefore decided to neglect all background contributions in the main document for the sake of simplicity. A more refined model including the estimation of λ is presented in the implementation (cf. 'Counting molecules in a DNA origami measurement' section).

In practice, we are able to obtain the measurements **Y** in different imaging modes. As shown in Fig. 1, the microscope has an additional STED laser, which can either be turned on or off, yielding different effective PSFs h. This allows us to image the specimen once via classical confocal microscopy (where the STED laser is turned off) and once via super-resolution STED microscopy. In total, this means that we collect and analyze both STED data $\mathbf{Y}^S = \left(\mathbf{Y}^{S,0},...,\mathbf{Y}^{S,m_d}\right)$ consisting of matrices containing the pixel-wise k-photon counts $\mathbf{Y}^{S,k} = \left(Y_{\mathbf{i}}^{S,k}\right)_{\mathbf{i} \in [n] \times [n]}$ as well as confocal data $\mathbf{Y}^C = \left(\mathbf{Y}^{C,0},...,\mathbf{Y}^{C,m_d}\right)$ consisting of similar matrices $\mathbf{Y}^{C,k} = \left(Y_{\mathbf{i}}^{C,k}\right)_{\mathbf{i} \in [n] \times [n]}$. The properties, advantages and disadvantages of both imaging modes will be discussed in more detail later on.

Segmentation

If we tried to construct pixel-wise CIs for the local number of molecules everywhere in the image (comparable to (1)), we would not obtain meaningful results due to the large number of pixels, as the necessary multiplicity adjustments would inflate the results. Therefore, it is important to first select regions of interest (RoIs) in the image, on which we afterwards analyze the local number of molecules. The aim of this section is to introduce a both intuitive and statistically rigorous data-driven segmentation of the given image into 'active' regions (i.e. containing molecules) and 'inactive' regions (i.e. containing no or only a very few molecules). To make this useful, we aim to select a system of RoIs such that

(**R1**) all interesting clusters of molecules are contained in one of the RoIs,
(**R2**) with high probability, each RoI contains at least one molecule,
(**R3**) the RoIs do not intersect (i.e. form a valid segmentation),
(**R4**) the RoIs are reasonably small and
(**R5**) the RoIs have suitable shapes.

A standard approach to this problem would be the usage of a standard data-driven segmentation algorithm on the STED data $\mathbf{Y}^{S,1}$ (offering a much better resolution compared to a confocal image), which—using suitable tuning parameters—hopefully yields reasonable RoIs satisfying (R1) and (R3)–(R5). However, strong statistical guarantees such as (R2) do typically not hold. On the other hand, systems of sets satisfying (R2) are often overlapping and thus (possibly highly) redundant, i.e. violate (R3), and furthermore do not satisfy (R5). Our approach is therefore to profit from the strengths of two different approaches, creating a hybrid version that inherits the positive aspects of both ingredients. A theoretical guarantee of property (R1) is very difficult and is not provided by our method. This can still be justified by a liberal choice of the selection method, e.g. in terms of a smaller probability in (R2). However, even with 90% confidence in (R2), our simulations show very good coverage properties of our final segmentation.

Hybridization

Suppose that the set $\widetilde{\mathcal{B}}$ is a (possibly complex and highly redundant) system of sets satisfying (R2). The hybrid procedure is based on the important yet simple observation that (R2) remains valid if sets from $\widetilde{\mathcal{B}}$ are removed or enlarged. Therefore, to reduce the complexity (and redundancy) of $\widetilde{\mathcal{B}}$, we first neglect all sets that completely contain smaller sets. This might cause a loss of information, but on the other hand, we are mostly interested in smaller sets as they contain the highest spatial information. This step yields a system $\mathcal{B} \subset \widetilde{\mathcal{B}}$ of sets still obeying (R2).
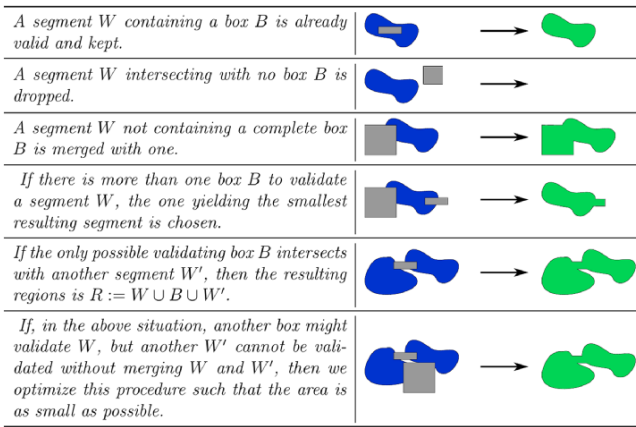
Fig. 3. Illustration of the hybridization algorithm: boxes B (gray), segments W to be validated (blue) and final segments (green).

In the following, we describe how the system $\mathcal{B}$ can be hybridized with a segmentation $\mathcal{W}$ (i.e. a system of disjoint connected subsets of $[n] \times [n]$) such that on the one hand, the segmentation property is obeyed, and on the other hand, (R2) is kept valid. There is a lot of freedom in choosing the segmentation algorithm, which allows the user to apply any method of choice and hence to generate a system of RoIs consisting of more naturally shaped segments. Possible examples include the famous Watershed segmentation algorithm (see later for a brief description), k-means clustering or more recent AI-based techniques. Xie *et al.* [26] provide a comprehensive review on spatial hot-spot detection methodology.

In the hybridization step, we try to validate each segment $W \in \mathcal{W}$ using one (or more) of the sets $B \in \mathcal{B}$. Thus, let $W \in \mathcal{W}$. If there exists a $B \in \mathcal{B}$ such that $B \subset W$, then W is already valid (in the sense of (R2)). If no such B exists, then we merge W with one of the intersecting sets to generate an enlarged valid segment. This step does, however, cause complications, as the chosen set $\hat{B} \in \mathcal{B}$ might intersect with other segments in $\mathcal{W}$. This issue is resolved as follows:

1. We generate a set of *validation sets* $\mathcal{B}_W := \{B \in \mathcal{B} \mid W \cap B \neq \emptyset\}$ for each $W \in \mathcal{W}$.
2. If $\mathcal{B}_W = \emptyset$, then W has to be dropped.
3. If $\mathcal{B}_W \neq \emptyset$, we check for each $B \in \mathcal{B}_W$ if $B \in \mathcal{B}_{W'}$ for another $W' \in \mathcal{W}$. This yields a list $\{B \in \mathcal{B}_W \mid \nexists W' \in \mathcal{W}, W' \neq W \text{ s.th. } B \in \mathcal{B}_{W'}\}$ of candidates for validating W. Every box B in this list of candidates can now be used only to validate W (and not to validate any other segment).
4. From the list, we choose the one $\hat{B}$ yielding the smallest new valid segment $R = \hat{B} \cup W$.
5. If no such B is found, we merge W with another segment $W' \in \mathcal{W}$, which is again done such that the new resulting segment is as small as possible.

The process described earlier is repeated iteratively until all segments have been validated or dropped, i.e. until (R3) is satisfied. The final, hybridized segmentation is then given as a set of regions R denoted by $\widehat{\mathcal{ROI}}$. We illustrate this procedure in Fig. 3.

The next theorem guarantees that the hybrid selection $\widehat{\mathcal{ROI}}$ provides a valid segmentation in the sense of (R2) and (R3).

Theorem 2. Let $\widehat{\mathcal{ROI}}$ be the regions of interest arising from the aforementioned hybridization algorithm of two systems of subsets $\mathcal{B}, \mathcal{W} \subset 2^{[n] \times [n]}$. Then the following holds true:

1. Let $\alpha \in (0, 1)$. If the system of sets $\mathcal{B}$ obeys property (R2) in the sense of (11), then also $\widehat{\mathcal{ROI}}$ does.
2. If $\mathcal{W}$ is a segmentation, i.e. if the sets in $\mathcal{W}$ are pairwise disjoint, then also $\widehat{\mathcal{ROI}}$ is a segmentation.

Clearly, the question whether (R1) and (R4)–(R5) are also satisfied depends on the choice of $\widetilde{\mathcal{B}}$ and $\mathcal{W}$. If those are generated suitably, then this will be the case as discussed in the following section.

Practical implementation

In the following, we will focus on two specific methods to generate the sets $\widetilde{\mathcal{B}}$ and $\mathcal{W}$, which we used in our implementation. To generate $\widetilde{\mathcal{B}}$, we employ the MISCAT procedure introduced in [25], which we will briefly describe in the following for convenience (see also [27] for a comprehensive review of the MISCAT procedure). We start our segmentation by determining rectangular regions of interest via a multiple testing approach. The rectangular regions will be referred to as boxes, denoted by $B_{\mathbf{x},\mathbf{h}}$. The subscript $\mathbf{x}$ denotes the position of the upper left corner of the box within the grid of pixels, while the subscript $\mathbf{h} = (h_1, h_2)$ denotes the side lengths of the box. For a given $\mathbf{h}$, we consider all $B_{\mathbf{x},\mathbf{h}} \subset \{\mathbf{x}_i \mid \mathbf{i} \in [n] \times [n], \mathbf{i} + \mathbf{h} \in [n] \times [n]\}$ and furthermore we use several different *scales* $\mathbf{h}$ ranging from small to large, so that all together, we consider a highly redundant system of boxes that guarantees very good detection properties of the MISCAT method. To test whether a box $B_{\mathbf{x},\mathbf{h}}$ contains markers, we design pairs of functions $\varphi_{\mathbf{x},\mathbf{h}}, \Phi_{\mathbf{x},\mathbf{h}}$ such that

$$\langle f * h, \Phi_{\mathbf{x},\mathbf{h}} \rangle = \langle f, \varphi_{\mathbf{x},\mathbf{h}} \rangle, \tag{9}$$

with the fluorescence intensity $f = f_N f_p$, where $*$ denotes spatial convolution. The function pairs $\varphi_{\mathbf{x},\mathbf{h}}, \Phi_{\mathbf{x},\mathbf{h}}$ will in practice be generated by a kernel somewhat similar to Wavelets, such that they form a multiscale system that adapts to the PSF h. This ensures optimal detection power of MISCAT by fine-tuning the kernel, see [25]. Given equation (9) and having in mind that $\mathbf{Y}^{S,1} = (Y_i^{S,1})_{i \in [n] \times [n]}$ consists of the most highly resolved measurements available, $\langle \mathbf{Y}^{S,1}, \Phi_{\mathbf{x},\mathbf{h}} \rangle$ can serve as a *local test statistic* for testing whether $f_{|_{B_{\mathbf{x},\mathbf{h}}}} \not\equiv 0$. The local test statistics are then combined by taking the maximum after subtracting a scale-dependent penalization (to ensure equal contribution of the differently sized boxes). This yields a test statistic for a multiple test over all considered boxes with a controlled FWER as a consequence of Theorem 4 in [25]. It is furthermore shown in Theorem 1 of [25] that quantiles of this maximum (or scan) statistic can be simulated using a suitable Gaussian approximation, which then gives rise to local (taking a suitable scale penalization into account) critical values $c_{\mathbf{h},\alpha}$. The outcome of this first step is then a set $\widetilde{\mathcal{B}}$ of candidate RoIs given by

$$\widetilde{\mathcal{B}} = \{B_{\mathbf{x},\mathbf{h}} \mid \langle \mathbf{Y}^{S,1}, \Phi_{B_{\mathbf{x},\mathbf{h}}} \rangle \geq c_{\mathbf{h},\alpha}\}.$$

Note that the resulting candidate RoIs are boxes of different sizes $\mathbf{h}$. The MISCAT procedure at level α is constructed

in such a way that

$$\mathbb{P}\left[\text{There is a pair } (\mathbf{x}, \mathbf{h}) : B_{\mathbf{x}, \mathbf{h}} \in \widetilde{\mathcal{B}} \mid \text{no markers at all}\right] \leq \alpha. \tag{10}$$

Furthermore, as the distribution of the maxima of the statistics $\left\langle \mathbf{Y}^{S,1}, \Phi_{B_{\mathbf{x}, \mathbf{h}}} \right\rangle$ does not depend on the actual positions of the N markers in the image, we have *subset pivotality*, cf. [28]. As a consequence, the FWER control (10) for MISCAT implies that with high probability, *all selected boxes contain a marker*:

$$\inf_{C \in \mathcal{C}_N} \mathbb{P}\left[f_{|_{B_{\mathbf{x}, \mathbf{h}}}} \not\equiv 0 \text{ for all } B_{\mathbf{x}, \mathbf{h}} \in \widetilde{\mathcal{B}} \mid \text{configuration } C\right] \geq 1 - \alpha, \tag{11}$$

where $\mathcal{C}_N$ is the set of all possible configurations (distributions) of N markers in the image of interest. This is a FWER control over the selected system of boxes in the strong sense, which in general—more precisely without subset pivotality—does not follow from the weaker FWER control (10).

For the generation of the segmentation $\mathcal{W}$, we use here the well-known watershed algorithm, whose name is in reference to a geological watershed, which separates adjacent drainage basins. The algorithm can be applied to gray scale images, which are interpreted as a topographic map, with the brightness of each point representing its height. The algorithm finds the lines that run along the tops of ridges (see [29] for more details).

As a corollary of Theorem 2, we obtain for the particular choice of MISCAT at level α and the watershed segmentation $\mathcal{W}$ with any choice of tuning parameter asymptotically

$$\inf_{C \in \mathcal{C}_N} \mathbb{P}\left[\begin{array}{l} \text{Each } \widehat{R}_i \in \widehat{\mathcal{ROI}} \text{ contains} \\ \geq 1 \text{marker} \mid \text{config. } C \end{array}\right] \geq 1 - \alpha.$$

This means that the so-obtained hybrid segmentation inherits the strong control of the FWER from the MISCAT procedure.

An illustration of the performance of the segmentation step is provided in Fig. 2 and in Fig. 1 in the Supplementary Material.

Statistical molecule counting

In this section, we discuss how to estimate the local number of molecules inside a region R from the data. Furthermore, we provide a central limit theorem for our estimator, which immediately allows us to construct (asymptotic) CIs quantifying the precision of our counting method. Later on, we will apply the methodology presented here to all regions in the previously discussed hybrid segmentation $\widehat{\mathcal{ROI}}$.

For the derivation of our estimator, suppose for a moment that all the N molecules of the specimen are clustered at a single point, this is $\psi(j) = \mathbf{k} \in [n] \times [n]$ for all $1 \leq j \leq N$, with some fixed $\mathbf{k} \in [n] \times [n]$. In this case, it follows immediately from the definition that

$$\sum_{\mathbf{i} \in [n] \times [n]} s_l(\mathbf{x_i}) = \sum_{\mathbf{i} \in [n] \times [n]} \sum_{j=1}^{N} \mathfrak{p}_j(\mathbf{x_i})^l = \sum_{\mathbf{i} \in [n] \times [n]} \sum_{j=1}^{N} p_j^l \cdot h(\mathbf{x_i} - \mathbf{x_k})^l.$$

In the following, we assume that the brightness f_p is a function of the location only and not a function of the individual molecules and we further assume a certain regularity of the function f_p such that it can be assumed constant within sufficiently small regions. As the brightness depends mostly on local parameters such as the pH, this is a realistic restriction of our original model where each marker was allowed to have an individual brightness. We obtain

$$\sum_{\mathbf{i} \in [n] \times [n]} s_l(\mathbf{x_i}) = p^l N \sum_{\mathbf{i} \in [n] \times [n]} h(\mathbf{x_i} - \mathbf{x_k})^l.$$

In all practically relevant examples, the PSF h is rapidly decaying and can hence be considered as periodic, which justifies the additional assumption that

$$H_l := \sum_{\mathbf{i} \in [n] \times [n]} h(\mathbf{x_i} - \mathbf{x_k})^l \tag{12}$$

is independent of the location $\mathbf{k} \in [n] \times [n]$. Consequently, it holds

$$N = \frac{H_2}{H_1^2} \frac{\left(\sum_{\mathbf{i} \in [n] \times [n]} s_1(\mathbf{x_i})\right)^2}{\sum_{\mathbf{i} \in [n] \times [n]} s_2(\mathbf{x_i})}. \tag{13}$$

In the situation of Theorem 1, the quantities $\mathbf{s}(\mathbf{x_i})$ can be immediately estimated, as then

$$\mathbf{s}(\mathbf{x_i}) = (\mathbf{T}^{-1}(\mathbf{D}(\mathbf{x_i})) + O\left(\gamma(\mathbf{x_i})^{m_d+1}\right). \tag{14}$$

This leads to a natural plug-in estimator for the total number of molecules, where $\mathbf{D}(\mathbf{x_i})$ is estimated by the relative frequencies of detecting $0, 1, \ldots, m_d$ photons at position $\mathbf{x_i}$.

Let us now return to the general situation that we want to estimate

$$N_R = \#\{\text{markers in region } R\} = \int_R f_N(x) \, dx$$

for some region $R \subset [0,1]^2$. For this, we pose the following physical assumption:

Assumption 3. The brightness within the region R is approximately constant if R is sufficiently small and the functions H_l defined in (12) do not depend on $\mathbf{k}$.

This assumption is (approximately) valid for many experimental settings as long as R is not too large, as the brightness depends mostly on local conditions such as temperature, pH value and so on. Under Assumption 3, it seems now natural to replace the sum over all locations $\mathbf{x_i}$ in (13) by the sum over all $\mathbf{x_i} \in R_\varepsilon$ with a slightly enlarged segment $R_\varepsilon = \{x \in [0,1] \mid \text{dist}(x, R) \leq \varepsilon\}$. The rationale behind is that the rapid decay of the PSF h ensures that molecules inside R will only or at least mostly contribute to those $\mathbf{x_i} \in R$. In practice, we choose ε of the order of the FWHM and enlarge only as long as R_ε does not intersect with any other enlarged segment. Let us introduce $\mathcal{I}_{R_\varepsilon} := \{\mathbf{i} \in [n] \times [n] \mid \mathbf{x_i} \in R_\varepsilon\}$. Then, the aforementioned considerations give rise to the plug-in estimator

$$\widehat{N}_R = \frac{H_2}{H_1^2} \frac{\left(\sum_{\mathbf{i} \in \mathcal{I}_{R_\varepsilon}} \widehat{s}_1(\mathbf{x_i})\right)^2}{\sum_{\mathbf{i} \in \mathcal{I}_{R_\varepsilon}} \widehat{s}_2(\mathbf{x_i})} \quad \text{with}$$

$$\widehat{\mathbf{s}}(\mathbf{x_i}) = \mathbf{T}^{-1}\left(Y_{\mathbf{i}}^{C,0}, \ldots, Y_{\mathbf{i}}^{C,m_d}\right)/t.$$

Note that the inversion of $\mathbf{T}$ might introduce uncertainties (and actually does, cf. Fig. 4 in our numerical simulations),
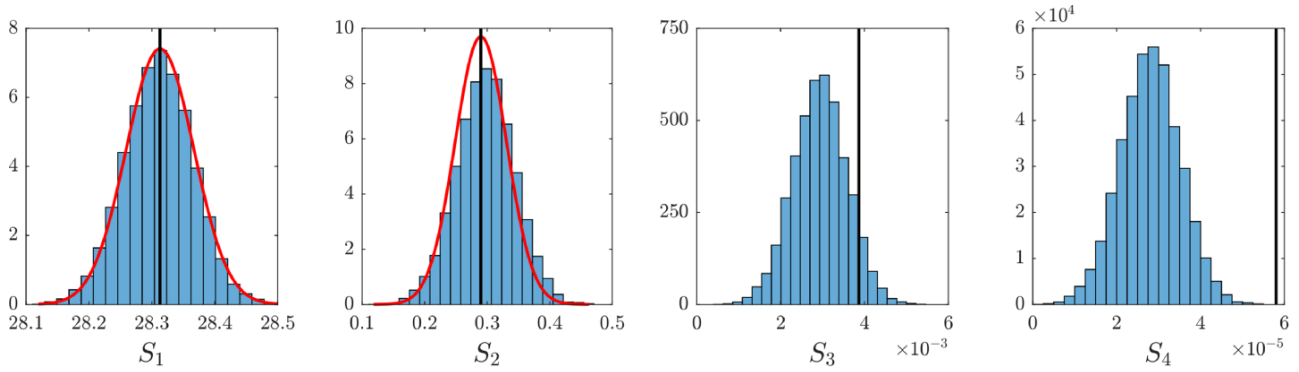
Fig. 4. Estimation of $S_l := \sum_{\mathbf{i} \in [n] \times [n]} s_l(\mathbf{x_i})$ for simulations of single, isolated cluster of $N = 20$ molecules with brightness $p = 0.02$ and a Gaussian PSF of FWHM=4px and $t = 10^4$ pulses per pixel. Histograms of estimated S_i for many repetitions. Vertical lines (black) present the true mean values. Red curves represent normal distributions with true variances around the true means.

and as the noise level of $\hat{s}_k(\mathbf{x_i})$ increases geometrically with k in view of the f_p^k-dependency, we decided to use only $\hat{s}_1(\mathbf{x_i})$ and $\hat{s}_2(\mathbf{x_i})$ to infer on N_R. Nevertheless, it is in principle possible to improve the estimate for N_R based on higher-order contributions $s_k(\mathbf{x_i})$, $k \geq 2$ along the aforementioned considerations.

The following central limit theorem for the estimator $\widehat{N}_R$ is based on asymptotics for $t \to \infty$ subject to $m_d = m_d(t) \to \infty$. Practically, it is clearly unrealistic to build microscopes with more and more detectors, but for asymptotic considerations, such an assumption is unavoidable, as for $t \to \infty$, arbitrarily high photon coincidences will occur and need to be recorded appropriately. This has already been observed in the model bias (7), and thus by $m_d = m_d(t) \to \infty$, we can assume that t is large enough such that the model bias is irrelevant. Since the influence of the size of m_d is of practical interest, finite sample bounds of Berry-Esseen type with respect to t as well as m_d are additionally provided. Practically, we will show in our simulation study that $m_d = 4$ for suffices for up to $N = 40$ molecules, $m_d = 6$ for up to $N = 100$ molecules and $m_d = 8$ even for around $N = 150$ molecules per segment. So only for extremely dense objects, an improvement of the experimental setup might be necessary.

Denote by $\Pi^d_{(k_1,\dots,k_l)} : \mathbb{R}^d \to \mathbb{R}^l$ the projection of a vector in $\mathbb{R}^d$ onto the vector of its coordinates $k_1,\dots,k_l$.

Theorem 4. Let $R \subset [0,1]^2$ and let $m_d = m_d(t) \to \infty$, such that $m_d \leq 5 \log(t)$. Suppose that the region R contains at least one fluorescent marker, that Assumption 3 is satisfied and that any marker in R has individual brightness smaller than 0.5 (Remark 5 (v)). Let

$$\Sigma_{\mathbf{E}}(\mathbf{x_i}) = \left(\begin{cases} E_j(\mathbf{x_i})(1 - E_j(\mathbf{x_i})) & j = k \\ -E_j(\mathbf{x_i})E_k(\mathbf{x_i}) & j \neq k \end{cases} \right)^{m_d}_{j,k=1},$$

and consider an arbitrary but fixed ordering $\mathbf{x_{i_1}}, \dots, \mathbf{x_{i_{|R|}}}$ of the points $\mathbf{x_i} \in R$. Furthermore, define the block diagonal matrix Σ_R by $\text{diag}\left(\Sigma_{\mathbf{E}}(\mathbf{x_{i_j}}) \right)_{j=1,\dots,|R|}$

and let $\Psi : \mathbb{R}^{m_d |R|} \to \mathbb{R}$ be given by

$$\Psi(\mathbf{y})$$
$$:= \frac{H_2}{H_1^2} \cdot \frac{\left\langle \mathbf{1}_{m_d|R|}, \left(\Pi_1^{m_d} \mathbf{T}^{-1}(y_{jm_d+1}, \dots, y_{(j+1)m_d})^T \right)_{j=1,\dots,|R|} \right\rangle^2}{\left\langle \mathbf{1}_{m_d|R|}, \left(\Pi_2^{m_d} \mathbf{T}^{-1}(y_{jm_d+1}, \dots, y_{(j+1)m_d})^T \right)_{j=1,\dots,|R|} \right\rangle}.$$

Assume that the Hessian matrix of Ψ, Hess_Ψ, exists and is bound in a neighborhood of $\mathcal{E}$ and that

$$\sigma_R^2 = \nabla \Psi(\mathcal{E})^T \Sigma_R \nabla \Psi(\mathcal{E}) > 0, \tag{15}$$

where $\mathcal{E} = (\mathbf{E}(\mathbf{x_{i_j}}))_{j=1,\dots,|R|}$. Let Z be a random variable following a centered normal distribution with variance σ_R^2. If t is sufficiently large, there exists a constant $C > 0$ such that

$$\sup_{s \in \mathbb{R}} \left| \mathbb{P}\left(\sqrt{t}\left(\widehat{N}_R - N_R \right) \leq s \right) - \mathbb{P}\left(Z \leq s \right) \right|$$
$$\leq C \frac{\log(t)^{\frac{2}{3}}}{t^{\frac{1}{6}}}, \qquad \text{as} \qquad t \to \infty. \tag{16}$$

Remark 5.

(i) In particular, the aforementioned theorem ensures that

$$\sqrt{t}\left(\widehat{N}_R - N_R \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_R^2), \qquad \text{as} \qquad t \to \infty,$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution.

(ii) Recall that the asymptotic considerations are with respect to the number t of laser pulses. As t tends to infinity, the number of pixels remains fixed. Therefore, the collection of regions R considered does not change asymptotically and contains only finitely many elements. Therefore, naturally, (16) holds uniformly in R.

(iii) We let $m_d \to \infty$ as $t \to \infty$, as for a fixed number m_d, the model bias (6) may otherwise asymptotically dominate the quantity $\sqrt{t}(\widehat{N}_R - N_R)$. The upper bound on the speed of convergence of m_d is needed in the proof, as with m_d, the number of events of the multinomial distribution tends to infinity. However, this restriction
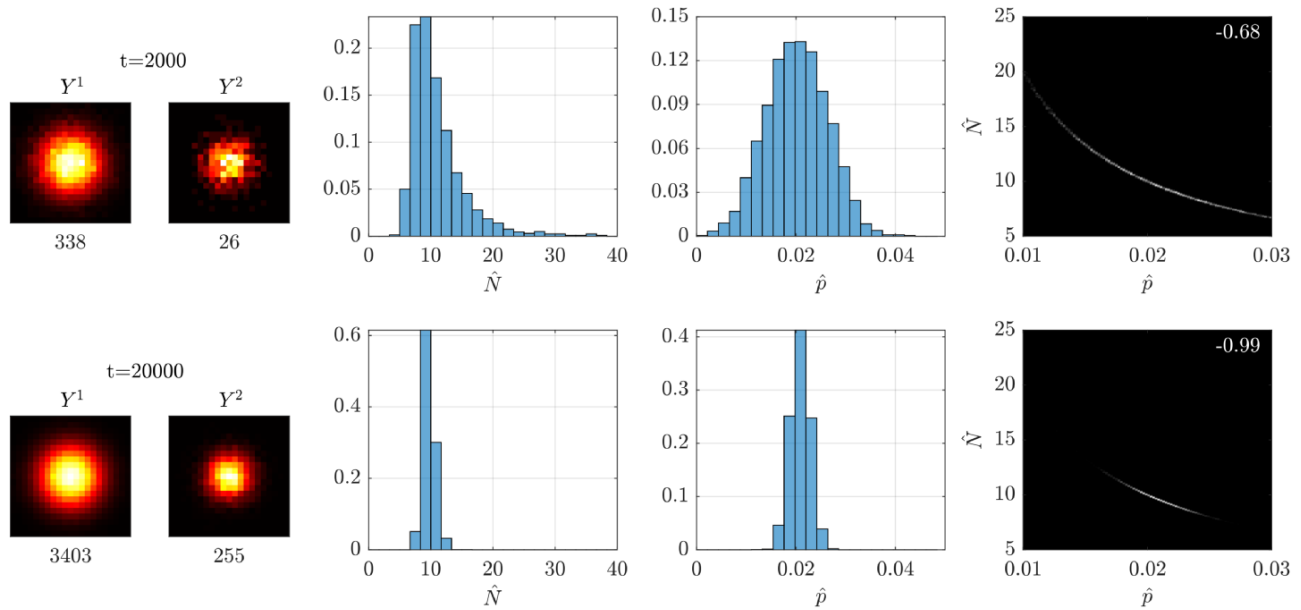
Fig. 5. Histograms of estimated N and p for simulations of single, isolated cluster of $N = 10$ molecules with brightness $p = 0.02$ and a Gaussian PSF of FWHM=4px and $t = 10^3$ and $t = 10^4$ pulses per pixel, respectively.

is irrelevant in practice, where the number of detectors that can be realized is limited.

(iv) In this work, the focus is on the estimation of local numbers of molecules, N_R. However, similar to (13), we immediately obtain an expression for p_R, the brightness in a region R, as $p_R = H_1 \sum_{\mathbf{i} \in \mathcal{J}_R} s_2(\mathbf{x_i}) / (H_2 \sum_{\mathbf{i} \in \mathcal{J}_R} s_1(\mathbf{x_i}))$, giving rise to a plug-in estimator $\hat{p}_R$. The theoretical properties of $\hat{p}_R$ can be analyzed using the same techniques as applied in the analysis of $\widehat{N}_R$.

(v) We assumed that any marker in R has individual brightness smaller than 0.5. This assumption is needed for technical reasons, but it is not restrictive, as typical brightness values are of the order of 0.02, as also used in our simulations. A brightness of 0.5 would entail that a marker emits a photon on average once every two pulses, which is unrealistic.

(vi) Given that $m_d \to \infty$, the model bias vanishes asymptotically. Our numerical simulations show that as few as six detectors suffice to deal with local numbers of molecules of about 100 (Fig. 5). The experimental setup we used (as shown in Fig. 1) exploits $m_d = 4$ detectors, and we find from Fig. 5 that it is expected to be accurate to treat local molecule numbers of around 30 – 40.

Theorem 4 gives rise to asymptotic CIs for N_R: $\left[\widehat{N}_R - \frac{\hat{\sigma}_R}{\sqrt{t}}, \widehat{N}_R + \frac{\hat{\sigma}_R}{\sqrt{t}} \right]$, where $\hat{\sigma}_R$ is the plug-in estimator for the asymptotic standard deviation σ_R defined in (15), that is, $\hat{\sigma}_R = \nabla \Psi(\hat{\mathcal{E}})^T \Sigma_R \nabla \Psi(\hat{\mathcal{E}})$, with $\hat{\mathcal{E}} = 1/t \cdot (Y_{\mathbf{i}}^{C,0}, \ldots, Y_{\mathbf{i}}^{C,m_d})_{\mathbf{i} \in R}$.

Numerical study

In this section, we will investigate the finite sample properties of our estimation procedure and the overall algorithm. The complete MATLAB® code including all examples discussed in the manuscript at hand is attached.

As a first step, we assumed a certain number of independent and identically behaving molecules with a fixed brightness p to be located all at the same position and to be imaged with the antibunching microscope operating only in confocal mode. The recording of the molecule sample is performed by scanning on a square lattice–like grid relative to the molecules' position and the grid spacing is here defined relative to the resolution of the microscope, i.e. with a FWHM of the PSF given in scanning grid pixel sizes. For a single cluster, segmentation of the data was omitted and the whole measurement area was assumed to belong to the single segment representing the molecule cluster. In Fig. 4, we depict histograms of the empirical distributions of $S_l := \sum_{\mathbf{i} \in [n] \times [n]} s_l(\mathbf{x_i})$ together with the theoretical expectations of S_l. It can be seen that for typical experimental conditions, the first two orders of S_l are reasonably well distributed around their expectations and can be used for retrieving the number and brightness of the molecules. For higher orders, systematic deviations are visible in the distributions of S_l, most probably a side effect of the ill-conditioned inversion of $\mathbf{T}$ under these circumstances.

This first simulation shows that it is possible to estimate the number of molecules in a single, isolated cluster from S_1 and S_2. This is further illustrated in Fig. 5, where we investigate the final estimator for the total number of molecules N (and the corresponding one for the common brightness p as discussed in Remark 4.3 (iv) in the main document) in a similar setting. Again, we use a single cluster with $N = 10$ molecules with common brightness $p = 0.02$ and omit the segmentation step by assuming that the whole measurement area was assumed to belong to the single segment representing the molecule cluster. Repetition of the simulation results in distributions for $\hat{N}$ and $\hat{p}$, which are depicted as histograms. It should be noted that the estimated molecule numbers and brightnesses are correlated and the joint distribution is concentrated along a hyperbola in the number–brightness plane (Fig. 5c), i.e. the product of the molecule number and its brightness is quite well estimated compared to the knowledge about
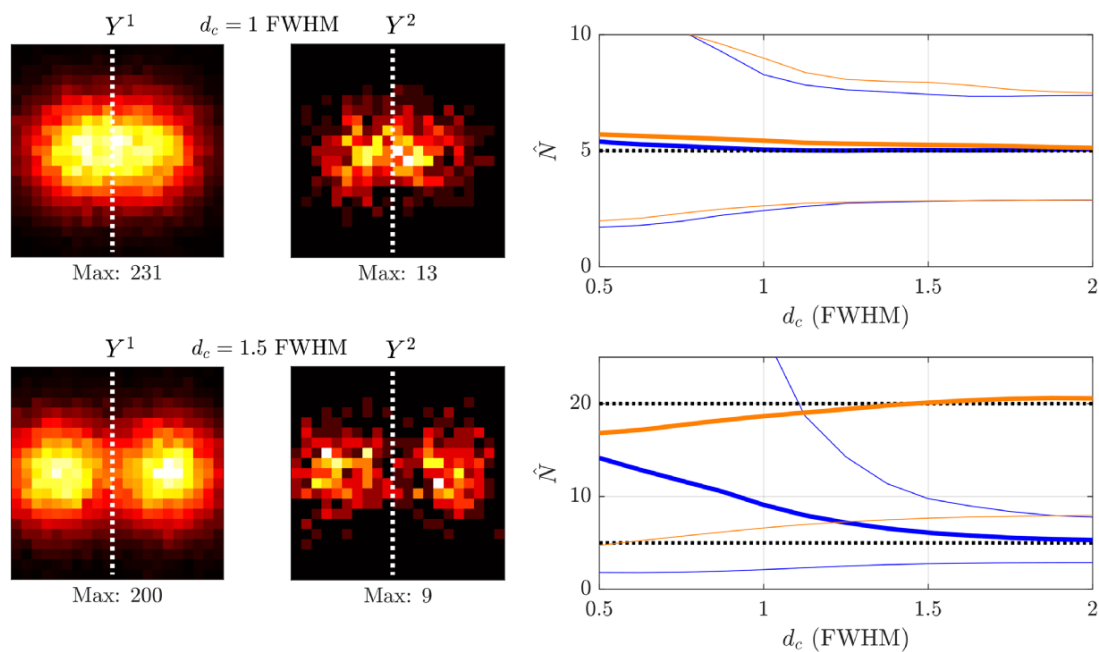
Fig. 6. Estimated N for simulations of two clusters of molecules with a defined distance d_c (given in multiples of the FWHM) with brightness $p = 0.02$, $t = 3000$ and a Gaussian PSF of FWHM=4px. (left) Example images of two clusters with five molecules each. The dotted white line represents the segment border. (right) Mean estimated number of molecules for each cluster and mean lower and upper bounds of the CIs (thin lines) in dependence of the distance d_c between the clusters. The black dotted line represents the true number of molecules in each cluster. Top: two clusters with $N_1 = N_2 = 5$ and bottom: $N_1 = 5, N_2 = 20$.

the single factors. For larger numbers of illumination pulses, the distributions of the estimated numbers and brightnesses become more concentrated, more symmetric and less biased, indicating that the employed estimators converge to the true underlying parameters. The precision is mainly limited by the number of repetitions, i.e. light pulses, that the sample can be illuminated with at every scan position. A few thousand excitation pulses per scan position have been reported to be possible without significant observable photobleaching of the molecules.

To study the influence of boundary effects, i.e. molecules residing close to the boundary of an image segment, we simulated two clusters of molecules with a defined distance that is on the order of the FWHM of the PSF. In that way, the images of the two clusters are partly overlapping. Two segments were created such that they fully covered the whole simulated image space and the border between the two segments was localized in the middle between the clusters. The analysis in each segment will be compromised by photons that are located in the respective other segment. The results depicted in Fig. 6 show that for a reasonable cluster distance (of approximately the FWHM of the PSF), the systematic boundary effects tend to become relatively small. It should be noted that our hybrid segmentation naturally tends to avoid large amounts of signal close to segment borders.

Another source of systematic bias is very dense accumulations of molecules. Therefore, we performed simulations in which we increased the number of detectors. We found that the range of molecules that can be estimated without visible bias strongly increases already for a number of detectors that is increased only moderately. Four parallel detectors, which are currently used experimentally, allow us to count up to 50 molecules per diffraction limited sample volume without

large bias with our method. With eight parallel detectors, this limit could be lifted to approximately 200, cf. the left panel in Fig. 7.

The overall performance of our algorithm is tested in a simulated arrangement of markers consisting of several clusters of molecules with varying brightness, cf. Fig. 2. It shows that the estimation of the local number of molecules is accurate in all segments and that the FWER control (11) is in fact kept. It must be said that some of the constructed CIs seem to be rather large (indicating a small estimated value $\hat{p}$ in that region), which is, however, unavoidable when asking for strong error controls such as (11). In addition, in Fig. 2, the true values of N_R are located on the boundary of the CIs in multiple cases, indicating that any method with smaller CIs, which are centered at $\widehat{N}_R$, would be susceptible to violations of the property (1).

Counting molecules in a DNA origami measurement

To establish the validity of the suggested CIs on real data, we used DNA origami sheets, which is an artificial structure that allows us to attach a relatively well-defined number of molecules at defined positions within a diffraction-limited volume [30]. The designed DNA origami sheet contained up to 24 fluorophores arranged in two lines of up to 12 fluorophore binding sites, and the distance between the two lines was ~70 nm, which is not resolvable in confocal mode of the microscope. Due to imperfections in the folding efficiency of DNA, the expected number of fluorophores per DNA origami was only ~19 [24]. Confocal microscopy cannot spatially resolve the single lines of fluorophores, but it allows us to obtain a
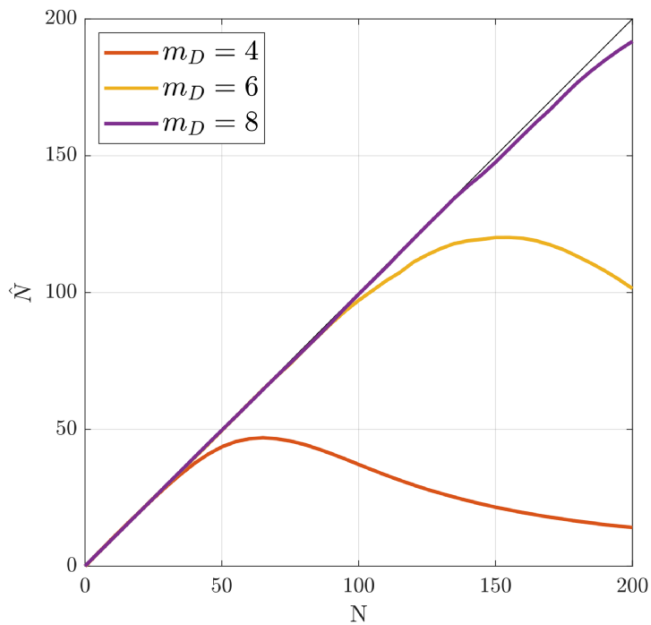
Fig. 7. Median of the estimated number of molecules for a simulation of a single cluster ($p = 0.02$, $t = 2 \cdot 10^4$) in dependence of the number of detectors used m_d. The visible bias is strongly reduced for larger m_d. All internal orders equal m_d. Graphs are slightly smoothed.

sufficient statistic on one and two photon detection events. A subsequent STED recording with improved lateral resolution (five-fold over confocal microscopy) resolved the molecular distribution within a single DNA origami sheet. Further experimental details are laid out in [24]. Here, we re-analyzed the recorded data to yield estimates and CIs on the number of fluorophores in either resolved DNA origami sheets or even single lines in these sheets (Fig. 8). The division of the data into suitable segments using our hybrid MISCAT and watershed segmentation approach was performed on the high-resolution STED data, while the molecular number and brightness were estimated from the less well-resolved but much brighter one- and two-photon confocal images. Before counting, the background intensity is estimated by a smoothing procedure from the data and then included in our refined model. The result is a segmentation map of the image area shown in Fig. 8b where an estimated number of molecules as well as a CI on the number of molecules are assigned to each segment. As expected, the estimated number of molecules is always within the calculated confidence bounds. The width of the confidence bounds (K) is about as large as the estimated number of molecules.

Discussion and outlook
Summary
In this work, a comprehensive physical and statistical modeling of the coincidence photon statistics encountered in antibunching microscopy has been conducted. The coincidence signals recorded from parallel detectors in a high-resolution fluorescence microscope intricately depend on the number of fluorescent molecules at a specific location as well as their brightness. Under the mild assumption that molecules locally exhibit the same brightness, and under neglecting all background contributions, a rigorous statistical model of the problem of estimating the number of molecules within small

regions of the sample has been developed. Based on this, an estimation procedure of the number of molecules was presented and was shown to converge to the true molecule counts for sufficiently long measurement times. Furthermore, CIs for the local number of molecules have been constructed. The estimation procedure requires choosing regions of the sample in which molecule numbers are to be estimated. A hybrid segmentation has been introduced, which combines a natural segmentation approach with guarantees that all segments contain molecules with high probability. In a numerical study, the convergence of the estimator has been affirmed. With increasing number t of excitation pulses, the obtained segment sizes as well as the size of the CIs has been shown to shrink.

Future work
In previous works, the number of molecules (as well as their brightness) was estimated as a continuous density from a global fit of the data with a coincidence statistics model up to second order [24]. The model presented here includes arbitrarily high photon coincidence orders. However, in this work, only the lowest two orders of $\hat{s}_k(\mathbf{x}_i)$ were used in the estimation of the local number of molecules. The optimal choice of the used number of orders depends on the accuracy of the estimation of each $\hat{s}_k$ and may be subject to further studies. An important issue here is the quality of the available data (i.e. the possible number of excitation pulses t), as otherwise the higher-order photon coincidences will not be able to provide additional information.

Segmentation-based estimation of local molecular counts was essential for the generation of confidence information. The importance of choosing an appropriate sample segmentation, which provides statistically sound guarantees (Theorem 2), for generating useful confidence information is a key finding of this work. A very large number of localized segments would result in overly large CIs (due to multiplicity), while a small number of extended segments would not result in meaningful local information. A balanced approach as suggested in this work results in useful molecule counting information. A multitude of different segmentations are imaginable, but the hybrid segmentation used in this work appears to be working reasonably well. Further work might deliver more insights into how segmentations of the sample can maximize the usefulness of the counting information in these cases.

The enlarged segments used for estimation are in principle required to contain all the collected fluorescence from all molecules within the segment and none from molecules outside the segment (segments should contain a molecule-free border with a width of at least the FWHM of the PSF). Clearly, this is not always the case in practice and is often difficult to achieve in densely labeled samples unless one restricts oneself to rather large segments, which would not result in useful local information. This will in general result in an uncontrollable bias in the molecule number estimation. In the case of multiple closely spaced segments, lost photon contributions from molecules within a segment can be partially compensated by contributions from molecules in adjacent segments. In our examples (Figs. 2 and 6), the estimated CIs were not compromised. A further careful analysis of segment border effects might result in additional contributions to the CI sizes.

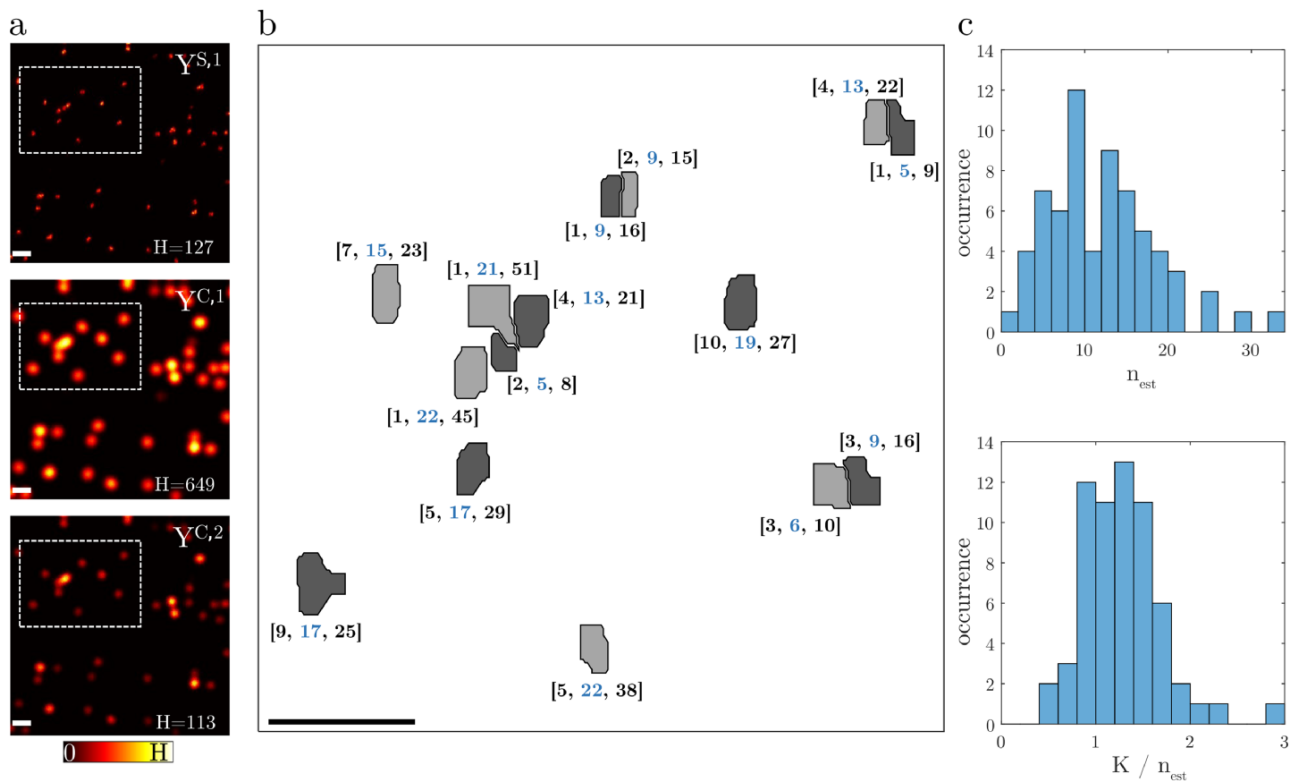The analysis performed in this study also assumed that molecules remain intact during the whole measurement.

Fig. 8. Application to recordings of DNA origami sheet structures labelled with ATTO 647N molecules (for experimental details, see [24]). The structures were immobilized at a low concentration on a glass surface and measured using confocal and STED microscopy. Each DNA origami can accommodate up to 24 fluorophores (12 in a line). (a) STED (top) and confocal (number of pulses per pixel $t = 2170$, pixel size 10 nm) one-photon (center) and two-photon (bottom) detection images. (b) Segmentation, estimated number of fluorophores in each segment and CIs for the data in the dotted rectangle. (c) Histogram of the estimated number of fluorophores in each segment of the recorded data (top) and histogram of the width of the CI (K) divided by the estimated number of fluorophores (bottom). Scale bars, 500 nm.

Indeed, in every microscopy experiment, fluorophores will eventually lose their ability to fluoresce (photobleach). To minimize estimation biases, we restricted the duration of the experiment to the characteristic time that will leave a strong majority of fluorophores intact. Should, nevertheless, some molecules photobleach during an experiment, we expect the number estimation to effectively approximate the mean number of unbleached molecules during the measurement. Photobleaching could be included in a more refined statistical model.

As discussed in the simulations, a too large number of molecules (or more precisely a too large product of local number of molecules and their brightness) in comparison to the number m_d of detectors used yield a systematic under-estimation of the number of molecules. This is due to the fact that the bias introduced by the model from Theorem 1 is no longer negligible. Our current experimental setup shown in Fig. 1 using four detectors is able to count approximately 50 molecules at a single diffraction limited spot accurately. Increasing the number of detectors modestly would increase this limit significantly as shown in Fig. 7. Further advances in detector technology currently being invented (being able to count single photons without significant dead times and therefore not limiting the observable coincidence photon order) are expected to diminish if not solve the problem completely while at the same time simplifying the microscope setup further.

All given examples in this work have been executed in two dimensions only, which corresponds to an often used imaging mode in scanning microscopy. However, the extension to 3D is straightforward and does not require any particular effort.

Supplemental data

Supplementary data are available at *Microscopy* online.

Funding

Acknowledgements

References

1. Hell S W (2007) Far-field optical nanoscopy. *Science* 316: 1153–1158.
2. Sahl S J, Hell S W, and Jakobs S (2017) Fluorescence nanoscopy in cell biology. *Nat. Rev. Mol. Cell Bio.* 18: 685–701.

3. 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.

4. 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.

5. Klar T A, Jakobs S, Dyba M, Egner A, and Hell SW (2000) Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proc. Natl. Acad. Sci. USA* 97: 8206–8210.

6. Masch J- M, Steffens H, Fischer J, Engelhardt J, Hubrich J, Keller-Findeisen J, D'Este E, Urban N T, Grant S G N, Sahl S J, Kamin D and Hell S W (2018) Robust nanoscopy of a synaptic protein in living mice by organic-fluorophore labeling. *Proceedings of the National Academy of Sciences*, Vol.115: pp. E8047–E8056.

7. Eggeling C, Ringemann C, Medda R, Schwarzmann G, Sandhoff K, Polyakova S, Belov V N, Hein B, von Middendorff C, Schönle A, and Hell S W (2009) Direct observation of the nanoscale dynamics of membrane lipids in a living cell. *Nature* 457: 1159–1162.

8. Balzarotti F, Eilers Y, Gwosch K C, Gynnå A H, Westphal V, Stefani F D, Elf J, and Hell S W (2016) Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. *Science* 355: 606–612.

9. Aspelmeier T, Egner A, and Munk A (2015) Modern statistical challenges in high-resolution fluorescence microscopy. *Annu. Rev. Stat. Appl.* 2: 163–202.

10. Hohage T, and Werner F (2016) Inverse problems with Poisson data: statistical regularization theory, applications and algorithms. *Inverse Problems* 32: 093001, 56.

11. Munk A, Staudt T, and Werner F (2020a) *Statistical Foundations of Nanoscale Photonic Imaging*, pp 125–143 (Springer International Publishing, Cham).

12. Kulaitis G, Munk A, and Werner F (2021) What is resolution? A statistical minimax testing perspective on superresolution microscopy. *Ann. Statist.* 49: 2292–2312.

13. Coffman V C, Wu P, Parthun M P, and Wu J- Q (2011) Cenp-a exceeds microtubule attachment sites in centromere clusters of both budding and fission yeast. *J. Cell Biol.* 185: 563–572.

14. Engel B D, Ludington W B, and Marshall W F (2009) Intraflagellar transport particle size scales inversely with flagellar length: revisiting the balance-point length control model. *J. Cell Biol.* 187: 81–89.

15. 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.

16. 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.

17. Hummer G, Fricke F, and Heilemann M (2016) Model-independent counting of molecules in single-molecule localization microscopy. *Mol. Biol. Cell* 27: 3637–3644.

18. Staudt T, Aspelmeier T, Laitenberger O, Geisler C, Egner A, and Munk A (2020) Statistical molecule counting in super-resolution fluorescence microscopy: towards quantitative nanoscopy. *Statist. Sci.* 35: 92–111.

19. 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. *The Annals of Applied Statistics* 13: 1397 – 1429.

20. Gabitto M I, Marie-Nelly H, Pakman A, Pataki A, Darzacq X, and Jordan M I (2021) A Bayesian nonparametric approach to super-resolution single-molecule localization. *The Annals of Applied Statistics* 15: 1742 – 1766.

21. Park M, Kim H- H, Kim D, and Song N W (2005) Counting the number of fluorophores labeled in biomolecules by observing the fluorescence-intensity transient of a single molecule. *Bull. Chem. Soc. Jpn.* 78: 1612–1618.

22. Frahm L, Keller-Findeisen J, Alt P, Schnorrenberg S, del Álamo Ruiz M, Aspelmeier T, Munk A, Jakobs S, and Hell S W (2019) Molecular contribution function in resolft nanoscopy. *Opt. Express* 27: 21956–21987.

23. Ta H, Kiel A, Wahl M, and Herten D- P (2010) Experimental approach to extend the range for counting fluorescent molecules based on photon-antibunching. *Phys. Chem. Chem. Phys.* 12: 10295–10300.

24. 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.

25. Proksch K, Werner F, and Munk A (2018) Multiscale scanning in inverse problems. *Ann. Statist.* 46: 3569–3602.

26. Xie Y, Shekhar S, and Li Y (2022) Statistically-robust clustering techniques for mapping spatial hotspots: a survey. *ACM Comput. Surv.* 55: 1–38.

27. Munk A, Proksch K, Li H, and Werner F (2020b) *Photonic Imaging With Statistical Guarantees: From Multiscale Testing to Multiscale Estimation.*, pp. 283–312 (Springer International Publishing, Cham).

28. Westfall P H, and Young S S (1993) *Resampling-Based Multiple testing: Examples and Methods for p-Value Adjustment*, *Vol. 279* (John Wiley & Sons, New York).

29. Meyer F (1994) Topographic distance and watershed lines. *Signal Process.* 38: 113–125.

30. Schmied J J, Gietl A, Holzmeister P, Forthmann C, and Steinhauer C (2012) Thorben Dammeyer, and Philip Tinnefeld. Fluorescence and super-resolution standards based on DNA origami. *Nat. Methods* 9: 1133–1134.