



# Molecular contribution function in RESOLFT nanoscopy

LARS FRAHM,<sup>1</sup> JAN KELLER-FINDEISEN,<sup>1</sup>  PHILIPP ALT,<sup>1</sup> SEBASTIAN SCHNORRENBURG,<sup>1</sup>  MIGUEL DEL ÁLAMO RUIZ,<sup>2</sup> TIMO ASPELMEIER,<sup>2,3</sup> AXEL MUNK,<sup>2,3</sup> STEFAN JAKOBS,<sup>1,4</sup>  AND STEFAN W. HELL<sup>1,\*</sup> 

<sup>1</sup>Department of NanoBiophotonics, Max Planck Institute for Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen, Germany

<sup>2</sup>Institute for Mathematical Stochastics, University of Göttingen, Goldschmidtstraße 7, 37077 Göttingen, Germany

<sup>3</sup>Felix-Bernstein Institute for Mathematical Statistics in the Biosciences, University of Göttingen, Goldschmidtstraße 7, 37077 Göttingen, Germany

<sup>4</sup>Clinic for Neurology, University Medical Center of Göttingen, Robert-Koch-Str. 40, 37077 Göttingen, Germany

\*stefan.hell@mpibpc.mpg.de

**Abstract:** The ultimate objective of a microscope of the highest resolution is to map the molecules of interest in the sample. Traditionally, linear imaging systems are characterized by their spatial frequency transfer function, which is given, in real space, by the point spread function (PSF). By extending the concept of the PSF towards the molecular contribution function (MCF), that quantifies the average contribution of a single fluorophore to the image, a straightforward concept for counting fluorophores is obtained. Using reversible saturable optical fluorescence transitions (RESOLFT), fluorophores are effectively activated only in a small, subdiffraction-sized volume before they are read out. During readout the signal exhibits an increased variance due to the stochastic nature of prior activation, which scales quadratically with the brightness of the active fluorophores while the mean of the signal scales only linearly with it. Using a two-state Markov model for the activation, showing comparable behavior to the switching kinetics of the switchable fluorescent protein rsEGFP2, we can approximate quantitatively the MCF of RESOLFT nanoscopy allowing to count the number of fluorophores within a subdiffraction-sized region of the sample. The method is validated on measurements of tubulin structures in *Drosophila melanogaster* larvae. Modeling and estimation of the MCF is a promising approach to quantitative microscopy.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

## 1. Introduction

Labeling molecules of interest with a fluorescent marker molecule delivers excellent contrast and allows optical readout using far-field fluorescence microscopes. For a long time, the spatial resolution of optical microscopes was limited to about half the wavelength of the used light, but exploiting molecular state properties of fluorophores, super-resolution microscopy was able to bring the resolution down to a few nanometers.

Mapping the number density of fluorophores on the nanoscale, where only a few fluorophores are present in each resolved region, has the potential to significantly advance knowledge about cellular processes. Knowing the absolute numbers of molecules and stoichiometries are the basis for structural models of protein complexes and simulations of biological processes. It also allows to define thresholds on the number of molecules to be able to produce a certain effect or identifying the content of molecule clusters or small compartments on a quantitative basis. For example, the estimation of the number of constituent proteins in kinetochores reported

unexpected high numbers of proteins present [1], while the quantification of numbers of proteins used for the regeneration of flagella helped refine models for flagella assembly [2]. Cluster of Tom20 proteins in the mitochondrial outer membrane seem to have a fixed diameter suggesting an equal number of proteins in each cluster, even though the nanoscale distribution of the clusters is finely adjusted to growth conditions [3]. Further knowledge of the number of Tom20 on the nanoscale would support elucidating the regulatory mechanism of the distribution of these proteins. However, quantifying numbers of fluorophores requires a careful calibration of the contribution of each fluorophore to the microscopic image.

All fluorescence nanoscopy or super-resolution microscopy methods rely on consecutively bringing the fluorophores in bright signaling (on) and dark non-signaling (off) states. The state transitions can be elicited either in a spatially controlled or in a spatially stochastic manner. The first strategy is realized in methods called stimulated emission depletion [4,5] (STED), saturated structured-illumination microscopy [6] (SSIM) and reversible saturable optical fluorescence transition [7,8] (RESOLFT). In these methods, an intensity distribution with one or multiple intensity minima drives the molecules optically between an on- and an off-state, transferring the markers to one of these states except those in the vicinity of the minima. The light pattern is then scanned across the sample so that every position is represented in the super-resolved image. Spatially stochastic methods such as photoactivated localization microscopy [9] (PALM) or stochastic optical reconstruction microscopy [10] (STORM) switch adjacent molecules to the on-state one by one. Once activated, they are able to emit multiple fluorescence photons, which are used to identify the position of the molecule.

Spatially stochastic methods intrinsically contain information about single fluorophore contributions. However, counting the molecules is not straightforward. Molecules in the on-state may not emit enough photons to be recognized. Some molecules may not assume this state at all, while other molecules may occupy the on-state repeatedly and be counted multiple times. These effects require extensive modelling of the switching and imaging behavior to be able to count the number of fluorescent molecules in the sample [11–13].

In spatially controlled nanoscopy like STED or RESOLFT, the mean obtained signal is proportional to the number of active fluorophores. The signal from a fluorophore, which determines the molecular contribution function (MCF), is measured in detected photons per fluorophore during readout. One approach to calibrate the average signal from each fluorophore would be to use a sample with a known label density, and to compare this standard to the sample of interest. However, with the brightness per fluorophore being sensitive to the kinetics of the state transition parameters, and different optimal imaging parameters like dwell times chosen in each experiment, an intrinsic calibration of the fluorophore in the imaging process is preferable. The average signal can also be calibrated from single bleaching steps, provided that the signal from single fluorophores is detectable [14]. In the case of diffusing fluorophores, the average signal of molecules can be estimated intrinsically using the mean and the variance of the fluctuating signal [15]. For example, the mean and variance method was utilized to estimate the brightness in combined STED and fluorescence correlation spectroscopy (FCS) experiments for diffusing probes [16]. In imaging applications, photon antibunching, which is the fact that a single excited fluorophore can only emit a single photon at a time, and a stochastic model of the image formation process could estimate the number of molecules from a combination of confocal and STED recordings without any external calibration of the brightness per molecule [17].

RESOLFT microscopy based on reversibly switchable fluorescent proteins (RSFP) is able to achieve superresolution using low light levels [18–20]. Here, the diffraction limit is broken using long-lived dark states, such as the dark isomer of reversibly switchable fluorescent proteins. As these proteins emit fewer photons in their on-state isomer than the synthetic dyes used for STED microscopy, bleaching steps cannot be easily detected and the photon antibunching effect would be rather weak.

Here, a novel method for the intrinsic estimation of the brightness of RSFPs is developed. We propose a model for the image formation process in RESOLFT microscopy. The image is explained by the convolution of the fluorophore distribution with the MCF, which describes the average contribution of a single fluorophore to the image in a linear imaging system, in an un-normalized, quantitative way and directly connects number densities with collected signal levels. Knowledge of the shape and amplitude of the MCF provides a direct method to count the number of molecules in a highly resolved image.

Based on a suitable switching model for the fluorophores, we approximate the shape of the MCF. Free model parameters are calibrated from the image data itself using a data analysis similar to the mean and variance method of [15] but applicable to immobile or only slowly changing molecule arrangements. The calibrated MCF is then applied to estimate the number density in RESOLFT images of microtubule filaments in *Drosophila melanogaster* larvae where an independent biochemical estimation of the number density is available.

## 2. Theory

### 2.1. The molecular contribution function (MCF) for counting in fluorescence imaging

A fluorescence image contains the contribution of each fluorescent molecule in the sample. In analogy to the point spread function, which is the image of a single theoretical point source, we define the molecular contribution function (MCF) as the quantitative spatial distribution of the average number of photons counted from a single fluorophore. We assume a space-invariant, linear imaging system where identical fluorophores act independently. Denoting the sampled region by  $X \subset \mathbb{R}^2$ , the measured 2D image  $Y(\vec{x})$  at each scan position  $\vec{x} \in X$  is then given by the convolution of the number density  $n(\vec{x})$  with the MCF( $\vec{x}$ )

$$Y(\vec{x}) = (n * \text{MCF})(\vec{x}) + \varepsilon(\vec{x}), \quad (1)$$

where  $*$  denotes the convolution operator  $(f * g)(\vec{x}) = \int f(\vec{s})g(\vec{x} - \vec{s}) d\vec{s}$  and  $\varepsilon(\vec{x})$  describes the measurement noise, i.e. the deviation of a recorded image from the average image. The number density is represented by

$$n(\vec{x}) = \sum_{j=1}^{n_0} \delta(\vec{x} - \vec{x}_j), \quad (2)$$

where  $\delta$  is the Dirac delta,  $n_0$  the total number of fluorophores, and  $\{\vec{x}_j\}_{j=1, \dots, n_0}$  are the positions of the fluorophores within the sampled region  $X$ .

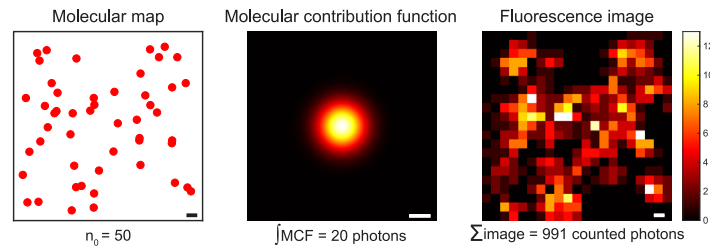
The aim of this work is to count the number of molecules in regions as small as the size of the resolution limit, i.e. the width of the central peak of the MCF. This requires the quantitative knowledge of the MCF. Note that every fluorescence microscopy technique with linear imaging conditions and independent and identically behaving fluorophores features an MCF, which lends itself to mapping the number of fluorophores. The integral over the MCF gives the average total signal  $S$  that each fluorophore contributes to the image,

$$S = \int_{\mathbb{R}^2} \text{MCF}(\vec{x}) d\vec{x}. \quad (3)$$

The estimated number of fluorophores  $\hat{n}_i$  in an isolated region  $X_i$  is then simply estimated by dividing the signal integrated over  $X_i$  by the total signal per fluorophore,

$$\hat{n}_i = \frac{1}{S} \int_{X_i} Y(\vec{x}) d\vec{x}. \quad (4)$$

A sketch of the role of the MCF in estimating the number of fluorophores is shown in Fig. 1. In the following we refine this framework by deriving analytic expressions of the MCF of RESOLFT nanoscopy for a specific switching model, and by estimating  $S$  using the mean and the variance of the measured signal.



**Fig. 1.** The molecular contribution function (MCF) (center, 2× magnified, with a total contribution of 20 photons of a single molecule the image) quantitatively relates the number and positions of molecules  $n_0 = 50$  shown in the molecular map (left) with the total number and distribution of observed photons in the image (right). The number of molecules in a region could be estimated by adding the recorded photons and dividing by the integrated MCF.

## 2.2. The MCF of RESOLFT nanoscopy

The RESOLFT principle is based on using switchable fluorophores with on- and off-states. We assume that they are independently switching and emitting fluorescence. For simplicity, we restrict ourselves to positive imaging, i.e. the MCF features a sharp, sub-diffraction sized positive signal peak.

For each scanning position  $\vec{s}$ , two conceptual steps can be differentiated. At first, fluorophores are transferred from the off-state, in which they reside initially, into the on-state only within a sharp, subdiffraction sized region centered at  $\vec{s}$ . In practice, this is achieved by first activating fluorophores with a diffraction-limited focus, and then deactivating fluorophores in the periphery using a doughnut-shaped focus with a central intensity zero. The spatial distribution of the probability of a single fluorophore to be effectively activated is named  $p_{\text{act}}(\vec{x})$ . Activated fluorophores are then read out recording their fluorescence signal. The spatial distribution of the recorded fluorescence photon signal of a single activated fluorophore during readout time is  $s_{\text{read}}(\vec{x})$ , typically proportional to the confocal point spread function (PSF). The MCF of RESOLFT nanoscopy is therefore given by the readout signal multiplied by the activation probability,

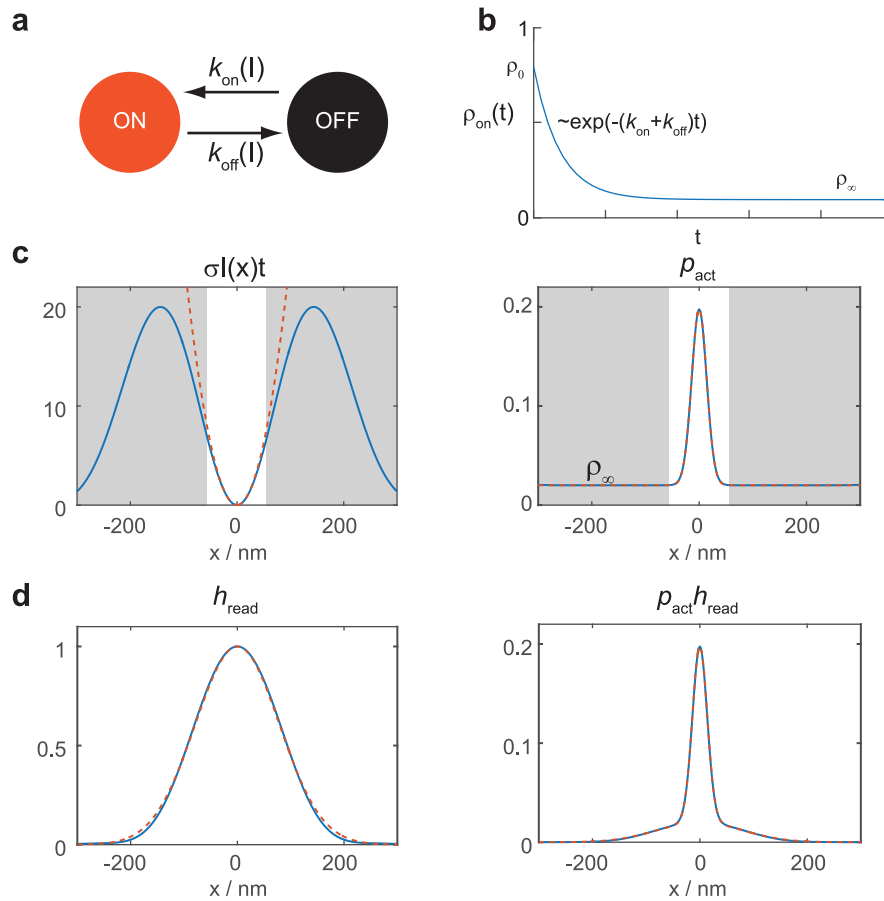
$$\text{MCF}(\vec{x}) = p_{\text{act}}(\vec{x})s_{\text{read}}(\vec{x}). \quad (5)$$

In the following, we derive expressions for the two parts of the MCF within the focal plane assuming switching kinetics represented by a two-state Markov model, strong saturation of the off-switching transition while preparing the fluorophore in the on-state, a parabolic intensity dependence of the deactivation light, and a sufficiently short readout duration.

### 2.2.1. Effective activation by saturated off-switching

The effective activation step preceding the readout depends on the switching kinetics of the fluorophore. A simple model for the switching kinetics is the two-state Markov model with light driven rates depending linearly on the applied light intensity (see Fig. 2a). The evolution of the state populations for this model is laid out in Appendix B.

The effective activation of a molecule in the on-state in RESOLFT nanoscopy is typically the result of two sub-steps [18]. First, a diffraction-limited spot of activation light is applied. The wavelength of the activation light is chosen so that the on-switching cross section is larger than the off-switching cross section. Consequently, fluorophores are switched on in the focal region. Then, fluorophores in the periphery of the focus are switched off using a doughnut-shaped focal spot of deactivation light centered at the scan position. For the wavelength of the deactivation



**Fig. 2.** Image formation in RESOLFT nanoscopy. **a:** Two-state switching model with a fluorescent on-state and a non-fluorescent off-state. The on- and off-switching rates depend on the laser intensity. **b:** Time evolution of the on-state population for the model in (a) and for a given total switching rate  $k_{\text{on}} + k_{\text{off}}$ . **c:** Left: Profile of the doughnut-shaped deactivation light (calculation using vectorial diffraction theory,  $\lambda = 491$  nm, NA=1.4(oil), blue) and corresponding parabolic approximation around the focal center (red, dashed). Right: Population of the on-state after application of the on- and off-switching light (blue) and with spatially constant on-switching and parabolic off-switching approximations (red, dashed). The saturation level ( $\sigma It$ ) of the off-switching is quite high (equals 20 on the rim of the doughnut distribution). **d:** Left: Confocal readout PSF (calculation using vectorial diffraction theory,  $\lambda_{\text{excitation}} = 491$  nm,  $\lambda_{\text{detection}} = 525$  nm, NA=1.4(oil), blue) and Gaussian peak approximation of equal (red, dashed) Right: Resulting shape of the molecular contribution function ( $\text{MCF}/b = \rho_{\text{act}} h_{\text{read}}$ ) with calculated illumination and detection distributions using diffraction theory (blue) and with approximations (red, dashed).

light, the off-switching cross-section is larger than the on-switching cross section. To achieve a sub-diffraction sized region, the off-switching by the deactivation light is saturated.

For the two-state switching model, the effective population in the on-state after applying the activation and the deactivation light can be calculated by applying Eq. (25) (see Appendix B) two times one after another with suitable switching light cross sections, spatially dependent irradiation doses and the initial condition of a subsequent switching step equating the final state of the previous switching cycle.

We introduce some approximations that hold for sufficiently strong saturation of the deactivation light, allowing us to derive a particularly simple, radially symmetric expression for  $p_{\text{act}}(\vec{x})$  within the focal plane.

For strong saturation, the final state is independent of the initial state as all fluorophores switch to the equilibrium level. The equilibrium level is reached everywhere except close to the focal center. Close to the central zero of the doughnut-shaped spot of deactivation light, the intensity profile can be well approximated by a radially symmetric 2D parabola  $I(\vec{x}) \approx I_{\text{max}}\alpha|\vec{x}|^2$  with a pattern steepness parameter  $\alpha$ , a maximal intensity on the rim of the doughnut distribution  $I_{\text{max}}$  and  $\vec{x}$  being a vector in the focal plane with the focal center as origin. For strong saturation, the width of the central peak of activated fluorophores will be far below the diffraction-limit, with a maximal on-state level  $p_{\text{on}}$  at the center. The effective activation probability in the focal plane, using the Gaussian peak approximation, is then,

$$p_{\text{act}}(\vec{x}) = p_{\infty} + (p_{\text{on}} - p_{\infty}) \exp(-4 \ln(2)|\vec{x}|^2 w_{\text{RESOLFT}}^{-2}), \quad (6)$$

with the equilibrium level  $p_{\infty}$ . This distribution describes a 2D Gaussian peak function with a full width at half maximum (FWHM)  $w_{\text{RESOLFT}}$  and a constant offset. The FWHM of this peak follows the well-known inverse square root relation on the applied maximal irradiation dose  $I_{\text{max}}t$  [8],

$$w_{\text{RESOLFT}} = \sqrt{\frac{4 \ln(2)}{\sigma \lambda / (hc) \alpha I_{\text{max}} t}},$$

with the energy of a photon of the deactivation light  $\lambda/(hc)$  and the corresponding total switching cross-section  $\sigma$  (see also Appendix B). From Eq. (6), it becomes apparent that the sharp, central spot of activated fluorophores is contrasted against a constant background of fluorophores in the equilibrium state.

The deviation between the used approximations and calculations of focal intensity profiles based on diffraction-theory is shown in Fig. 2. As already mentioned, the deviation is negligible at strong saturation levels.

### 2.2.2. Signal during readout

We assume that the readout duration is chosen sufficiently small so that during the readout, no additional switching takes place. The signal during readout  $s_{\text{read}}(\vec{x})$  is assumed to be collected in a confocal excitation and detection scheme. It can then be described by its shape distribution  $h_{\text{read}}(\vec{x})$ , normalized so that  $h_{\text{read}}(0) = 1$ , multiplied by a brightness factor  $b$ ,

$$s_{\text{read}}(\vec{x}) = b h_{\text{read}}(\vec{x}). \quad (7)$$

The focal readout brightness  $b$  is the average signal of an activated fluorophore at the focal center during readout. It summarizes optical and spectral properties of the fluorophores, as well as the detector integration time, and is regarded to be constant and common to all fluorophores within the image region. The spatial constancy of  $b$  is an assumption, which needs to be verified in real measurements by means of statistical testing (see the Discussion and Appendix F.3).

For confocal or STED imaging conditions during readout, the shape of the signal distribution within the focal plane is often well approximated by another 2D Gaussian peak function [21]

$$h_{\text{read}}(\vec{x}) = \exp(-4 \ln(2)|\vec{x}|^2 w_{\text{read}}^{-2}), \quad (8)$$

with  $w_{\text{read}}$  the FWHM of the 2D Gaussian peak describing the signal collection.

### 2.3. Estimation of the focal readout brightness $b$ from signal mean and variance

To obtain information about the statistical behavior of the signal, we model the signal as the result of a two-stage stochastic process, where the first stage is the activation and the second stage is the photon emission by activated fluorophores. The activation is modeled as a Bernoulli process with a spatially dependent activation probability distribution.

The states of the fluorophores are represented by Bernoulli random variables  $A_j$  with the effective activation probability  $p_{\text{act}}(\vec{x}_j - \vec{s})$  for a single fluorophore at position  $\vec{x}_j$  and scan position  $\vec{s}$ . The fluorophore indexed  $j$  is active ( $A_j(\vec{s}) = 1$ ) or inactive ( $A_j(\vec{s}) = 0$ ) depending on the value of the random variable

$$A_j(\vec{s}) \sim \text{Bernoulli}(p_{\text{act}}(\vec{x}_j - \vec{s})). \quad (9)$$

A fluorophore that remains in the on-state will on average yield  $bh_{\text{read}}(\vec{x}_j - \vec{s})$  detected photons, which we model by a Poisson distribution with mean value  $bh_{\text{read}}(\vec{x}_j - \vec{s})$ . In the off-state, it is modeled as not yielding any detectable photons. Therefore, the random variable describing the signal of a fluorophore  $B_j(\vec{s})$  depends on the activation state  $A_j(\vec{s})$ :

$$B_j(\vec{s}) \sim \begin{cases} \text{Poisson}(bh_{\text{read}}(\vec{x}_j - \vec{s})), & A_j(\vec{s}) = 1 \\ 0, & A_j(\vec{s}) = 0 \end{cases} \quad (10)$$

The mean and the variance of the signal of a single fluorophore are given by

$$\begin{aligned} \mathbb{E}[B_j(\vec{s})] &= bh_{\text{read}}(\vec{x}_j - \vec{s})p_{\text{act}}(\vec{x}_j - \vec{s}) \\ \text{Var}[B_j(\vec{s})] &= \mathbb{E}[B_j(\vec{s})^2] - \mathbb{E}[B_j(\vec{s})]^2 \\ &= \mathbb{E}[B_j(\vec{s})] + b^2h_{\text{read}}^2(\vec{x}_j - \vec{s})p_{\text{act}}(\vec{x}_j - \vec{s})(1 - p_{\text{act}}(\vec{x}_j - \vec{s})) \end{aligned} \quad (11)$$

Here we used that activation and readout are independent processes. Additionally, we assume that each fluorophore switches its state and emits independently, so that the mean and the variance of the total signal measured at each scan position are simply the sum of the mean and the variance of each individual fluorophore. The total signal detected at position  $\vec{s}$  is then given by

$$Y(\vec{s}) = \sum_j B_j(\vec{s}) + D(\vec{s}), \quad (12)$$

where  $D(\vec{s}) \sim \text{Poisson}(d(\vec{s}))$  represents background noise with a mean value of  $d(\vec{s})$ . We assume that the background contribution  $D(\vec{s})$  is independent of the emission from the fluorophores  $B_j(\vec{s})$  and also uncorrelated with the background signal at other positions.

Furthermore, the mean  $m(\vec{s}) = \mathbb{E}[Y(\vec{s})]$  and the variance  $v(\vec{s}) = \text{Var}[Y(\vec{s})]$  of the total collected signal  $Y(\vec{s})$  at each scan position  $\vec{s}$  can conveniently be expressed as convolutions with the density of fluorophores,

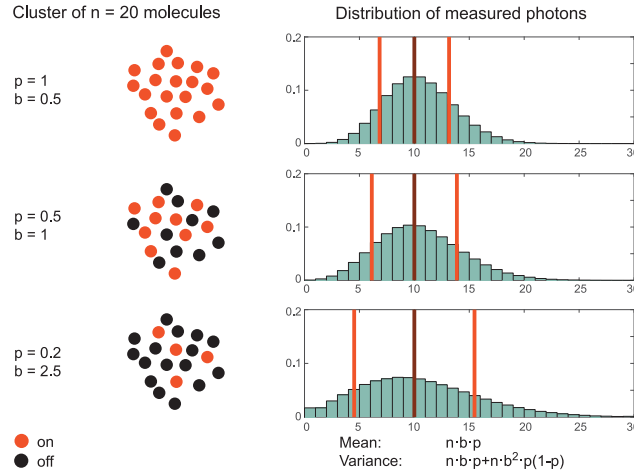
$$\begin{aligned} m(\vec{s}) &= b(n * h_{\text{read}}p_{\text{act}})(\vec{s}) + d(\vec{s}), \\ v(\vec{s}) &= m(\vec{s}) + b^2(n * h_{\text{read}}^2p_{\text{act}}(1 - p_{\text{act}}))(\vec{s}). \end{aligned} \quad (13)$$

The mean is the convolution of the number density with the MCF including an additional background while the variance exceeds the mean by an additional term with a convolution kernel  $b^2h_{\text{read}}^2(\vec{x})p_{\text{act}}(\vec{x})(1 - p_{\text{act}}(\vec{x}))$  which is similar but not identical to the MCF.

This additional variance, which is visualized in Fig. 3, can be used to estimate the focal readout brightness.

Note that for  $p_{\text{act}}(\vec{x})$  equal to 1 we recover the situation of non-switchable fluorophores, the excess variance term vanishes and a purely Poisson distributed signal is regained.

For the case of photoswitchable fluorophores, the variance of the collected signal is augmented due to the stochastic nature of the activation in the preparation step. The excess variance or



**Fig. 3.** Concept for counting in RESOLFT nanoscopy. Left: Depiction of a cluster of  $n_0 = 20$  molecules activated with different activation probabilities. Right: Photon counting histogram of the number of photons measured for different activation probabilities, while keeping the average number of photons per fluorophore constant. The variance of the distribution depends on the molecular parameters non-linearly. Mean and variance allow to calibrate the brightness per fluorophore.

overdispersion term is proportional to the square of the focal readout brightness  $b$ , unlike the mean signal, which scales linearly with it. This relation can be used to estimate  $b$  directly from integrated mean and variance of the image data using a Method of Moments estimator (MME), to be derived in the following.

To estimate the mean and variance of the signal at position  $\vec{s}$ , we require  $N \geq 2$  statistically independent images  $Y_j(\vec{s})$  ( $j = 1, \dots, N$ ) of the same structure of interest. The estimated mean  $\hat{m}(\vec{s})$  and the estimated variance  $\hat{v}(\vec{s})$  are obtained from the data,

$$\hat{m}(\vec{s}) = \frac{1}{N} \sum_j^N Y_j(\vec{s}),$$

$$\hat{v}(\vec{s}) = \frac{1}{N-1} \sum_j^N (Y_j(\vec{s}) - \hat{m}(\vec{s}))^2.$$
(14)

The pixel-wise estimation will still be influenced largely by the image noise. Also, the average signal and variance depends on the number density in the region of interest. The influence of the in general unknown and non-constant number density is overcome by integrating the mean and variance over the sampled region  $X \subset \mathbb{R}^2$ . Due to the convolution theorem, only the total number of molecules  $n_X$  within the region appears in the resulting expression, such that it is possible to estimate the focal readout brightness within this region. The integrated mean  $M = \int_X m(\vec{s}) d\vec{s}$  and integrated variance  $V = \int_X v(\vec{s}) d\vec{s}$  of the model in Eq. (13) gives

$$M = bn_X H_1 + d_{\text{total}},$$

$$V = M + b^2 n_X H_2,$$
(15)

with the integrated background signal  $d_{\text{total}} = \int_X d(\vec{s}) d\vec{s}$ , and  $H_1 = \int_{\mathbb{R}^2} (h_{\text{read}} p_{\text{act}})(\vec{s}) d\vec{s}$  and  $H_2 = \int_{\mathbb{R}^2} (h_{\text{read}}^2 p_{\text{act}}(1 - p_{\text{act}}))(\vec{s}) d\vec{s}$  being integrals of products of activation probability



distribution  $p_{\text{act}}$  and readout PSF  $h_{\text{read}}$ . Here, we assumed that the convolution kernels decay fast enough so that contributions of the fluorophores in the region  $X$  are mostly confined to the signal within the region.

The integrated mean  $M$  and integrated variance  $V$  can be estimated from the data by summing the estimated pixel mean and variances weighted with the area  $\Delta_p$  of a pixel in the image,

$$\begin{aligned}\hat{M} &= \sum_{\vec{s} \in X} \hat{m}(\vec{s}) \Delta_p, \\ \hat{V} &= \sum_{\vec{s} \in X} \hat{v}(\vec{s}) \Delta_p.\end{aligned}\quad (16)$$

Equating the integrated model mean and variances in Eq. (15) with the integrated empirical mean and variances in Eq. (16), and solving for the focal readout brightness  $b$ , results in the MME  $\hat{b}$  for the brightness,

$$\hat{b} = \frac{H_1}{H_2} \frac{\hat{V} - \hat{M}}{\hat{M} - d_{\text{total}}}. \quad (17)$$

An expression for the variance of  $\hat{b}$  in the limit of many fluorophores of known locations is given in Appendix F. For the special case of a single, isolated cluster of fluorophores, an analytic expression for the error in the estimation of  $b$  are derived in Appendix C.

Essentially, the relative error of  $\hat{b}$  depends mainly on the number of measurements, and is roughly independent of the number of molecules in the image.

### 3. Experimental Results

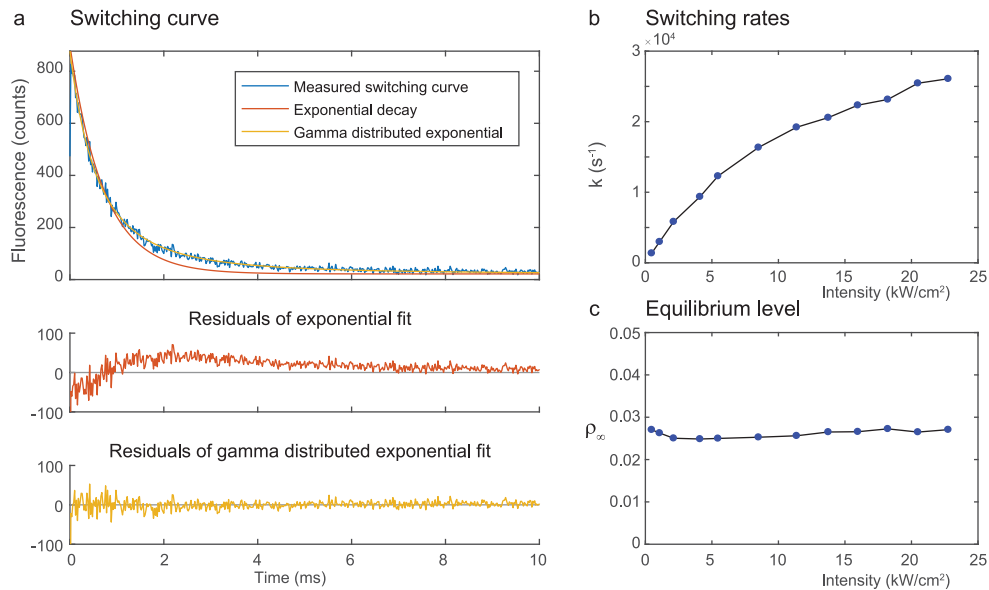
#### 3.1. Measurements of rsEGFP2 switching kinetics

In the theoretical part, the influence of the switching kinetics model on the shape of the MCF was derived. For a two-state switching model, the width of the central peak of the MCF is determined by the switching rate while the peak is offset by the equilibrium off-switching level. To compare the used switching model with switching of real fluorophores, we observed the switching kinetics of rsEGFP2, a widely used RSFP in RESOLFT nanoscopy [19]. The measurements were conducted on a custom built confocal microscope with additional widefield illumination paths for excitation with a 491 nm CW laser and activation with a 375 nm CW laser, as described before (Fig. S2 in [22]). The signal was detected with a point-like detector (APD) collecting fluorescence at  $525 \pm 25$  nm. In this configuration, averaging effects of the observed kinetics due to an inhomogeneous illumination intensity distribution within the detected volume can be excluded.

The sample was prepared by mounting living *E. coli* cells overexpressing rsEGFP2 on a coverslip in an agarose gel. To measure the ensemble state of the photoswitchable proteins, a stable reference was prepared for each measurement by saturating the activation using a long exposure of 1.6 ms of UV light, up to 160 times longer than the UV pulse used for activation during imaging. The intensity dependence of the switching rate and the switching background was examined using widefield excitation light exposure of constant light dose (exposure time  $\times$  intensity) using intensities up to 22 kW/cm<sup>2</sup>.

Due to the saturated activation, we assume that all molecules start in the on-state. The signal at any given time is then proportional to the population of rsEGFP2 in the on-state at that time. Results are shown in Fig. 4. As the signal shows systematic deviations from a single exponential decay especially for intermediate times, we devise a decay model assuming a Gamma distribution for the switching rate  $k$  with  $\alpha$  the shape parameter and  $\beta$  the rate parameter,

$$s(t) = A \int \frac{\beta^\alpha k^{\alpha-1}}{\Gamma(\alpha)} \exp(-\beta k) \exp(-kt) dk + b = A \frac{\beta^\alpha}{(\beta + t)^\alpha} + b, \quad (18)$$



**Fig. 4.** Switching kinetics measurements for rsEGFP2. **a:** Single switching curve with an off-switching light intensity of 500 W/cm<sup>2</sup>. Fits for exponential decay (red) and gamma distributed exponential decay (yellow) with residuals shown below. **b:** Measured switching rate  $k$  (inverse of the time where the signal drops to  $1/e$  of the initial value) in dependence of the off-switching light intensity. **c:** Equilibrium off-switching level (average level of the fluorescence in **a** at long times) in dependence of the off-switching light intensity.

where the amplitude of the decay is given by  $A$  and the signal decays to a background level  $b$ . The total switching rate  $k$  can be directly measured as the inverse of the time  $\tau = \beta(\exp(1/\alpha) - 1)$  until the signal drops to  $1/e$  of the initial value. The ratio of the equilibrium signal after complete switching and the initial signal is used to estimate the equilibrium population  $p_\infty$ .

The switching rate depends linearly on the intensity for low intensities, up to a few kW/cm<sup>2</sup>. These are typical values for light intensities around the center of the intensity zero, which determine the resolution enhancement of the RESOLFT process.

The equilibrium population does not depend on the intensity used, and is around 2.5% for all intensities.

### 3.2. Estimation of the MCF for RESOLFT images of rsEGFP2- $\alpha$ -tubulin

We recorded RESOLFT images of *Drosophila melanogaster* larvae body wall muscles expressing ubiquitously rsEGFP2 fused to  $\alpha$ -tubulin as described in [23]. Tubulin is known to form helices with a diameter of approximately 25 nm, each turn comprising 13 dimers of  $\alpha$ - and  $\beta$ -tubulin and spaced 8 nm apart [24]. The total density of  $\alpha$ -tubulin along a single filament can thus be estimated to be  $\approx 1625$  per  $\mu\text{m}$ .

To analyze the ratio of labeled to non-labeled  $\alpha$ -tubulin subunits in body wall muscles of *Drosophila melanogaster*, L3-larvae were dissected to isolate body wall muscles, which were subsequently used as a sample for Western blot analysis as described in Appendix A.2.

The ratio of  $\alpha$ -tubulin to rsEGFP2- $\alpha$ -tubulin in the muscle tissue is estimated to be  $\sim 1:6$  (see Fig. 6). Therefore, we expect the number density of rsEGFP2 molecules along a microtubule fiber to be  $\sim 230$  per  $\mu\text{m}$ .

### 3.2.1. The shape of the MCF

The MCF of RESOLFT nanoscopy is the product of the effective activation probability  $p_{\text{act}}(\vec{s})$  and the readout signal  $s_{\text{read}}(\vec{s})$ , as described in Section 2.2. For a two-state Markov model with a linear dependence of the rates on the applied light intensity, the MCF in the focal plane can be approximated as the superposition of two Gaussian peaks differing only in amplitude and width. Using the expressions in (6) and (7) gives

$$\text{MCF}(r) = b \left( p_{\infty} \exp(-4 \ln(2) r^2 w_{\text{read}}^{-2}) + (p_{\text{on}} - p_{\infty}) \exp(-4 \ln(2) r^2 w_{\text{eff}}^{-2}) \right), \quad (19)$$

where the width of the sharp RESOLFT peak  $w_{\text{eff}}$  is given by

$$w_{\text{eff}} = \left( w_{\text{read}}^{-2} + w_{\text{RESOLFT}}^{-2} \right)^{-1/2}. \quad (20)$$

By taking into account prior knowledge about the imaged object, the shape of the MCF may be estimated from image data itself [25]. The structure in the image of microtubule fibers is known to consist of homogeneously labeled filaments with a width well below the resolution of the RESOLFT microscope. Therefore, we model the density of molecules in the image as assembly of lines. For sufficiently sparse filament structures and a high enough SNR, the filaments can be localized without quantitative knowledge of the MCF simply by detecting lines in the image based on the scale space representation of local line profiles [26]. The resulting object model is considered proportional to the number density (see Fig. 5(b)).

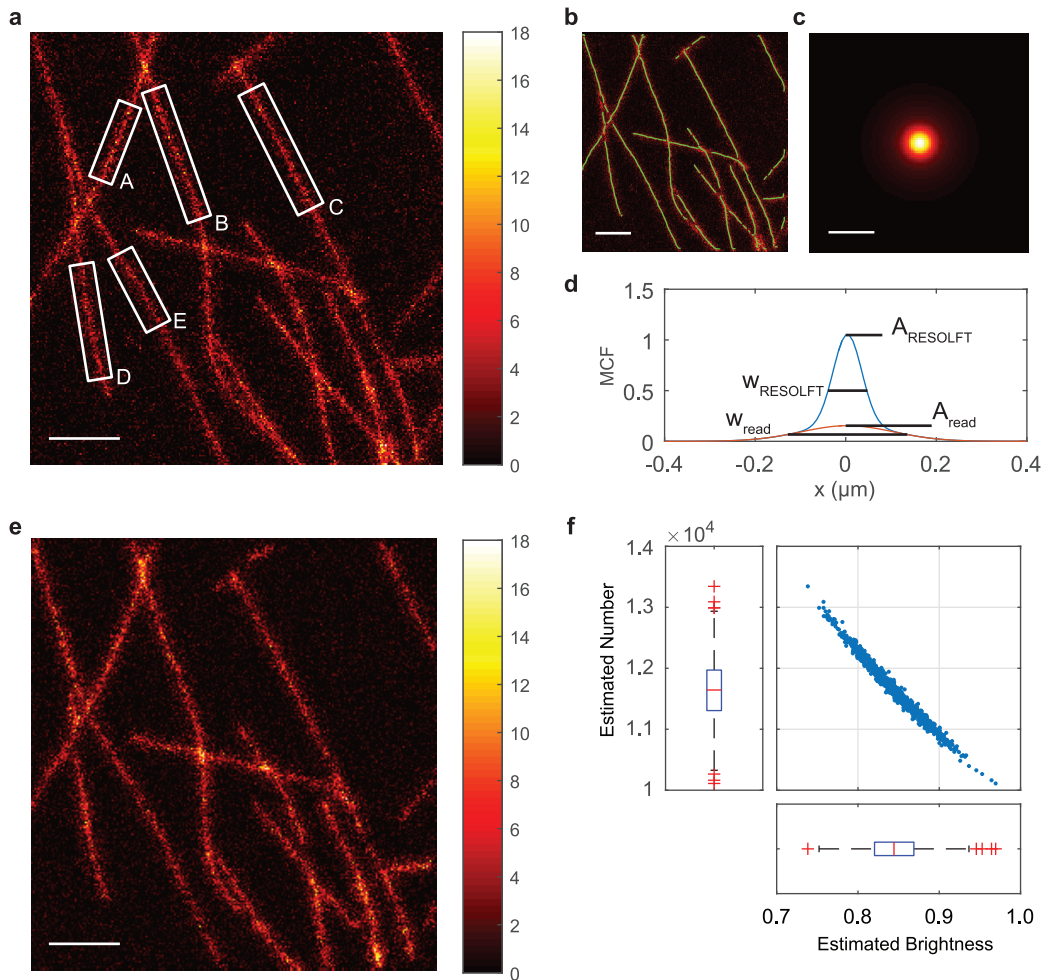
Given the object model, the shape of the MCF and the constant background level can be estimated by fitting a superposition of two centered Gaussian peaks convolved with the estimated object to the image data using a maximum likelihood approach for Poisson noise (see Fig. 5(c,d)).

From the widths of the fitted Gaussian peaks, the confocal readout FWHM  $w_{\text{read}}$  was estimated to be 235 nm and for the sharp, RESOLFT peak FWHM  $w_{\text{RESOLFT}}$  a value of 73 nm was obtained. As the number density was determined only up to a multiplicative constant in the structure estimation, the absolute amplitude of the MCF remains undetermined. However, the ratio of the amplitudes of the two Gaussian peaks can be used to estimate the activation probability at the focal center  $p_{\text{on}}$ , as the equilibrium population  $p_{\infty} = 2.5\%$  is known from the kinetics measurements and assumed to be reached in the periphery of the focal spot (i.e. where the off-switching transition is strongly saturated).

With a ratio of the amplitudes of the sharp peak to the background peak of  $\sim 6.9$ , we estimate that about 17% of the rsEGFP2 molecules at the focal center were effectively activated before readout. Using these values, the ratio of the integrated values of the convolution kernels  $H_1/H_2$  (see Eq. (17)) can be estimated to be  $\sim 1.7$ .

### 3.2.2. Inference on the focal readout brightness $b$

To quantitatively estimate the MCF of the RESOLFT image, the multiplicative constant, given by the focal readout brightness  $b$ , has to be estimated. As was laid out in Section 2.3, this can be achieved by measuring the integrated mean and the variance of the image signal. For the estimation of the variance, at least two images of the same structure are required. In practice, this can also be achieved by acquiring one image with half the desired pixel size in the fast scanning direction. Splitting the image in two along the fast scanning direction results in two independent images each with square pixels and only a slight relative shift of half a pixel length in the splitting direction (here 12 nm). While this allows to obtain multiple images of the same structure uninfluenced by drift or photobleaching, it can introduce a bias in the empirical variance in Eq. (14). As shown in Fig. 8 in the Appendix, the magnitude of this bias depends on the sampling of the image and the orientation of structures in the image. We use an estimator for



**Fig. 5.** Measurement of rsEGFP2- $\alpha$ -tubulin and estimation of the shape of the MCF. **a:** RESOLFT 2D image of dissected body wall muscles of wandering third instar *Drosophila melanogaster* larvae expressing rsEGFP2- $\alpha$ -tubulin. **b:** Line detection of the structures in the image in **a**. **c:** Result of the PSF estimation (superposition of two Gaussian peaks). **d:** Calibration of the activation probability (setting the equilibrium level to the one observed for rsEGFP in kinetics measurements). **e:** Simulated image resembling the image of **a** by taking the estimated structure from **b** and uniformly drawing a certain number of molecules from it. With the estimated MCF in **c**, noisy images can be simulated. **f:** Histograms of estimated readout brightness and estimated total fluorophore number for 1000 simulations. The true readout brightness and number of molecules in the image were chosen to equal the estimated readout brightness and number determined from the measurement. Scale bars  $1\mu\text{m}$  (a,b,e),  $200\text{nm}$  (c).

the variance with significantly reduced bias, derived in Appendix D (see also [27] for a detailed analysis of such difference-based variance estimators):

$$\hat{v}_{\text{shift}}(\vec{s}) = \frac{2}{3} (Y_1(\vec{s}) - Y_{2'}(\vec{s}))^2, \quad (21)$$

where  $Y_{2'}(\vec{s})$  is a split image that has neighboring pixels averaged along the split direction.

Applying the method of moments estimator for the focal readout brightness in Eq. (17) on the whole image shown in Fig. 5(a) yields a value of  $\hat{b}$  of  $\sim 0.9$  photon counts per activated rsEGFP2 molecule. Analyzing the time resolved readout signal showed that during readout the effective change in the on-state population was less than 10%. In Fig. 9, the readout brightness is estimated locally where only the data of a small region is used in each case. The estimated values vary between 0.6 and 1.1, but no systematic dependence on the location of the region is visible.

### 3.2.3. Counting the number of molecules along filaments

The total contribution  $S$  of each fluorophore to the RESOLFT image can be estimated by integrating Eq. (19) using the estimated shape and focal readout brightness. The average value of  $S = 3.58$  photons per rsEGFP2 molecule can be used to quantify the number of fluorophores in a region in the image using Eq. (4). The results of applying the estimator to the regions marked in Fig. 5 are shown in Table 1. The average number of rsEGFP2 molecules per  $\mu\text{m}$  along a microtubule filament is estimated to be  $\sim 180$ , which is similar but below the expected number density.

**Table 1. Number estimation of molecules in the regions A-E of Fig. 5(a).**

Region	number	length in nm	density per $\mu\text{m}$
A	268	1324	203
B	405	2065	196
C	379	2147	176
D	293	1713	171
E	237	1378	172
A-E	1582	8627	183

### 3.3. Simulation of number and brightness analysis for a known MCF shape

To approximate the error of this estimation, we conducted a Monte Carlo simulation implementing the model for the image formation process. Placing a certain number of molecules uniformly on the identified structure and using the obtained effective RESOLFT PSF, images were simulated (Fig. 5(e)). For each scanning position, the activation of the molecules was simulated by drawing binomial random numbers according to Eq. (9). From the aggregated mean signal from all active molecules, Poisson distributed random numbers were generated to represent the measured image. On the whole image, first the focal readout brightness was estimated, then the total number of fluorophores in the image was counted.

The estimated total number and focal readout brightness values of the simulated images are shown in Fig. 5(f). As the total number of detected photon counts is the product of the total number and the signal per fluorophore, there is a strong reciprocal correlation between these values. However, the relative deviation from their mean is well below 10%. Moreover, in Appendix F we derive an analytic expression for the variance of the estimator of  $\hat{b}$ . In this simulation setting, it predicts a value 4.9% for the relative standard deviation of the estimator from its mean, and the value 4.1% for the standard deviation.

To test if variations of the brightness over the field of view can be detected using our method, we simulated 1000 images for different brightness values from the set  $B = \{0.05, 0.1, 0.15, \dots, 1.95, 2.0\}$  and compared estimated brightness values from one half of the image simulated with brightness  $b_1 \in B$  with estimated brightness values from the other half of the image simulated with brightness  $b_2 \in B$  by applying the 2-sample Welch test [28, p. 447] (see Appendix F.3 for details). As shown in Fig. 10, differences in brightness between both halves of the image of 0.5 can be detected with

about 80% probability except when either  $b_1$  or  $b_2$  are very small. Note that this figure depends on the distribution of fluorophores in the image, and is thus not universal.

#### 4. Discussion & outlook

We derived a rigorous framework for counting molecules in fluorescence microscopy based on the molecular contribution function representing the average contribution of a single fluorophore to the image. The image can be represented as the number density of fluorophores convolved with the MCF.

The MCF was calibrated for the case of RESOLFT nanoscopy by first estimating the shape of the MCF from the image data using a model for the structure of fluorophores in the sample, and then quantifying the focal brightness using the mean and the variance of the measured fluorescence signal. The excess variance term, caused by the stochastic nature of the fluorophore activation in RESOLFT, scales quadratically with the focal brightness per fluorophore. This enables the quantitative estimation of the MCF for RESOLFT nanoscopy from multiple images of the same structure.

The estimation of the shape of the MCF, as well as the focal activation probability requires an accurate model of the switching kinetics of the used reversibly switchable fluorophore. Reversibly switchable fluorescent proteins have been shown to undergo switching by a combination of processes like cis-trans isomerization, protonation, and short scale interactions in the protein [29]. Our kinetics measurements of rsEGFP2 suggest that a two-state Markov model for the switching between a fluorescent on-state and a non-fluorescent off-state can be applied to estimate the on-state population of rsEGFP2 with sufficient accuracy. The observed uniform width of the  $\alpha$ -tubulin is a strong indication that the resolution and the switching kinetics remain the same across the entire field of view.

This model does not take into account the saturation of the switching rate seen for higher intensities and experimental deviations from the exponential decay predicted in the Markov model. By modeling the switching kinetics more accurately, the shape of the MCF and the activation probability may be estimated more precisely. Any relative error in the estimation of the activation level  $p_{\text{on}}$  will also directly be present in the estimated number of fluorophores.

To quantify the focal brightness per fluorophore, the mean and the variance of the signal were calculated for a stochastic model with a binomial activation of the fluorophores and a Poisson model for the detected signal from each fluorophore. In this model, the variance depends quadratically on the focal brightness, in contrast to the mean signal, which has a linear dependence. This allows inferring on the focal brightness of the fluorophores intrinsically from the image data using a Method of Moments estimator.

The main difference to earlier work about using fluorescence fluctuations to count molecules ([15]) is that our method counts molecules in at least temporarily immobile bright fractions of the sample, while these earlier methods quantified mobile fractions of the sample.

We applied our framework to RESOLFT recordings of rsEGFP2 fused to  $\alpha$ -tubulin in *Drosophila melanogaster* larval muscle tissue. The estimated number of rsEGFP2 molecules per  $\mu\text{m}$  microtubule filament in *Drosophila melanogaster* larval muscle tissue is similar but below the value derived by quantitative western blot analysis for dissected muscle tissue. This might hint at a less efficient incorporation of rsEGFP2 fused to  $\alpha$ -tubulin.

The focal activation level was estimated to be only  $\sim 17\%$ , which limits the emission from the sharp, central peak. The reason for this may be a non-zero intensity at the center of the used doughnut-shaped off-switching light distribution or an incomplete activation in the on-switching step. However, due to increased photobleaching, it may not be desirable to activate a larger fraction of molecules at each scan position.

We assumed that active fluorophores emit fluorescence independently. While interaction of multiple close-by fluorophores has been reported, the average distance of active fluorophores at

the focal center during readout for our measurement in *Drosophila melanogaster* is about 30 nm, a quenching interaction seem unlikely.

The focal brightness is calibrated assuming that it does not change over an image region. This assumption is essential for the method, and we validate it on the image scale by means of statistical testing. We performed a Welch test respecting the potentially inhomogeneous variance (see Appendix F.3 and [28], p. 447) at level  $\alpha = 0.01$  for the hypothesis that two halves of the image have equal brightness, and the test is not able to reject the hypothesis. The detection power of the test is analyzed in simulations of the image structure, which show that the test can detect differences of 0.5 in the brightness with roughly 80% probability. Because the focal brightness is calibrated from the data, a variable brightness between different microscopes, samples or parts of a sample does not compromise the counting ability.

Our simulations showed that the counting error, apart from possible inaccuracies in the MCF model, depends mainly on the number of repeated measurements and the brightness. The achievable number of measurements is limited by the required recording time and the number of possible switching cycles per fluorophore. Future reversibly switchable fluorophores with a lower tendency to bleach may enable a larger number of measurements.

The theoretical considerations as well as the experimental results were limited to 2D images and thin samples corresponding to the commonly used RESOLFT recording mode. The extension to 3D is uncomplicated in the theory and additionally requires only a sufficiently good calibration of the 3D shape of the MCF.

The excess variance of the signal features a different convolution kernel from the mean signal (see Eq. (17)), so that Poisson-likelihood based image deconvolution methods cannot be expected to be accurate. However, for our images the standard Richardson-Lucy method performs sufficiently well. In general, special deconvolution techniques adapted to the particular statistics of RESOLFT images might be needed.

Our counting framework maps the number of switchable fluorophores. To quantify the number of biological molecules of interest, the degree, efficiency and specificity of the labeling has to be controlled as well. With an optimal labeling of the target structure with stable labels, this technique should provide an accurate counting of molecules on the nanoscale.

## Appendix

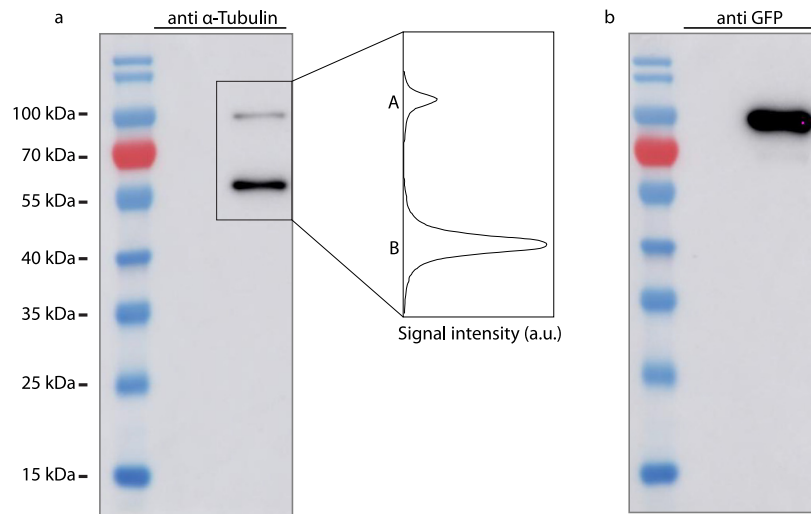
### A. Materials and methods

#### A.1. Sample preparation for switching kinetics measurements

A plasmid, encoding the sequence for rsEGFP2 was cloned into the high-copy expression vector pQE31 (Qiagen) and used in *E. Coli* cells to obtain highly concentrated proteins in a cellular environment. Direct observation of protein properties in the cytosol was accomplished by immobilizing the bacteria in agarose gel (2%) on standard coverslips (20× 20 mm, #1.5) mounted to an object slide.

#### A.2. Western Blot Analysis of *Drosophila melanogaster* larvae

The front half of one wandering third instar larvae expressing rsEGFP2- $\alpha$ -tubulin was dissected in 1x PBS to remove all tissues except body wall muscles. Subsequently the cuticle with attached body wall muscles was transferred to a new Eppendorf tube and mixed with Laemmli Buffer (Sigma-Aldrich, St. Louis, Missouri, USA) and denatured for 10 min at 95° C before loading onto a SDS-polyacrylamide gel. The protein-lysate of several larvae was mixed and the amount of approximately one larva was loaded into each lane of the gel. The SDS gel was blotted using a semi-dry blotter onto a Nitrocellulose membrane. The protein-blot was blocked using 5% fat free milk powder dissolved in 1 x PBS for 1 hr. The membrane was cut into several pieces and incubated with antibodies against GFP and alpha-tubulin. The polyclonal  $\alpha$ -tubulin antiserum



**Fig. 6.** Western blot analysis of body wall muscle protein lysate of rsEGFP2- $\alpha$ -tubulin expressing *Drosophila melanogaster* third instar larvae. **a:** Comparison of the signal from rsEGFP2- $\alpha$ -tubulin (A) to the signal of unlabeled  $\alpha$ -tubulin (B) via labeling using antiserum against  $\alpha$ -tubulin. Ratio was determined on the total signal intensity of the respective peak. **b:** Control staining using antiserum against GFP.

(Abcam, Cambridge, England) and the anti-GFP antibody (Clontech, Mountain View, United States) were used in a concentration of 1:2500 in 2,5% fat free milk powder dissolved in 1x PBS for 1 hr. HRP-conjugated secondary antibody (Jackson ImmunoResearch, Suffolk, UK) was used at a dilution of 1:5.000 in 2,5% fat free milk powder dissolved in 1x PBS for 1 hr and detected with an ECL-kit (Perkin Elmer Life Science, Waltham, USA) using an Amersham Imager 600 (GE Healthcare, Little Chalfont, UK).

### A.3. Imaging parameters for the RESOLFT images

The RESOLFT image shown in Fig. 5 was taken as described in [23], using an illumination with  $2.8\mu\text{W}$  for  $15\mu\text{s}$  for the diffraction-limited activation with a 405 nm laser, an illumination with  $17.8\mu\text{W}$  for  $500\mu\text{s}$  using a doughnut-shaped focus of 488 nm light, and finally a diffraction-limited excitation with  $24.2\mu\text{W}$  using 488 nm light for  $100\mu\text{s}$ . All powers were measured in the back aperture of the objective lens. For the RESOLFT image, the first  $30\mu\text{s}$  of the readout time were integrated.

## B. Two-state Markov switching model with linear rates

We assume a stochastic model for the memory-less switching between a single fluorescent on-state and a single non-fluorescent off-state. A scheme of the model is depicted in Fig. 2(a). The switching rates  $k_{\text{on}}$  and  $k_{\text{off}}$  determine the time evolution of the state populations. The time evolution of the on-state population  $p_{\text{on}}$  is given by an exponential decay with an effective switching rate  $k = k_{\text{on}} + k_{\text{off}}$ , starting from the initial state  $p_0$  and ending in the equilibrium state  $p_{\infty}$ :

$$p_{\text{on}}(t) = p_{\infty} + (p_0 - p_{\infty}) \exp(-kt), \quad (22)$$

where the equilibrium state is given by

$$p_{\infty} = \frac{k_{\text{on}}}{k_{\text{on}} + k_{\text{off}}}. \quad (23)$$



With the population of the off-state being  $1 - p_{\text{on}}(t)$ , the two-state system is fully characterized. For light driven rates that depend linearly on the intensity of the applied light we can define the switching rates as

$$\begin{aligned} k_{\text{on}} &= \sigma_{\text{on}}\lambda/(hc)I, \\ k_{\text{off}} &= \sigma_{\text{off}}\lambda/(hc)I. \end{aligned} \quad (24)$$

where  $\sigma_{\text{on}}$  and  $\sigma_{\text{off}}$  are effective absorption cross-sections for the on- and off-switching, respectively,  $hc/\lambda$  is the energy of a single photon and  $I$  is the intensity of the applied light.

Defining the effective total absorption cross-section  $\sigma = \sigma_{\text{on}} + \sigma_{\text{off}}$ , the on-state population only depends on the initial state and the irradiation dose  $It$ :

$$p_{\text{on}}(It) = p_{\infty} + (p_0 - p_{\infty}) \exp(-\sigma\lambda/(hc)It). \quad (25)$$

### C. Brightness and number of fluorophores on single clusters

For the case of a single, isolated cluster consisting of  $n_0$  molecules all at the same position, the estimation of the focal readout brightness  $b$  as well as the number of fluorophores is especially straightforward. We assume that  $N$  measurements are performed with the cluster location coinciding with the focal center. All fluorophores in the cluster then are subjected to the same effective activation probability  $p_{\text{on}} = p_{\text{act}}(0)$  and the average readout signal is simply  $b$ .

#### C.1. Analytic results

In the following, we assume that the average number of background counts  $d$  and the activation probability  $p_{\text{on}}$  are known with no or negligible error. Following Eq. (13), the mean  $m$  and the variance  $v$  of the measured signal are then given by

$$\begin{aligned} m &= bn_0p_{\text{on}} + d, \\ v &= m + b^2n_0p_{\text{on}}(1 - p_{\text{on}}). \end{aligned} \quad (26)$$

Equating mean and variance with the sample mean  $\hat{m}$  and the sample variance  $\hat{v}$  from  $N$  measured signals  $\{Y_j\}_{j=1,\dots,N}$  and solving Eq. (26) for number and focal brightness, gives the estimators  $\hat{n}_0$  and  $\hat{b}$  for the number of molecules  $n_0$  and the focal brightness  $b$  using the method of moments

$$\begin{aligned} \hat{n}_0 &= \frac{1 - p_{\text{on}}}{p_{\text{on}}} \frac{(\hat{m} - d)^2}{\hat{v} - \hat{m}}, \\ \hat{b} &= \frac{1}{1 - p_{\text{on}}} \frac{\hat{v} - \hat{m}}{\hat{m} - d}. \end{aligned} \quad (27)$$

We approximate the standard deviations of the estimator for the mean  $\Delta m = \sqrt{\text{Var}[\hat{m}]}$  and of the estimator for the variance  $\Delta v = \sqrt{\text{Var}[\hat{v}]}$  assuming that the signal follows a normal distribution with the known expectation and variance. This yields

$$\begin{aligned} (\Delta m)^2 &= v/N, \\ (\Delta v)^2 &= \frac{2v^2}{N-1}. \end{aligned} \quad (28)$$

By propagating the error for the estimators (27), we get the following result for the relative errors of the estimation of the number and focal brightness:

$$\begin{aligned} \frac{\Delta b}{b} &\simeq \sqrt{\frac{(\Delta v)^2}{(v-m)^2} + \frac{(\Delta m)^2(\sqrt{v}-d)^2}{(v-m)^2(m-d)^2}}, \\ \frac{\Delta n_0}{n_0} &\simeq \sqrt{\frac{(\Delta v)^2}{(v-m)^2} + (\Delta m)^2 \left( \frac{2}{m-d} + \frac{1}{v-m} \right)^2}. \end{aligned} \quad (29)$$

By only keeping the leading term  $(\Delta v)^2/(v - m)^2$  under the square root we find a joint, simple expression for a lower bound of the relative error of both the estimated number and the estimated brightness:

$$\frac{\Delta b}{b}, \frac{\Delta n_0}{n_0} \gtrsim \sqrt{\frac{2}{N-1}} \left( 1 + \frac{1}{b(1-p_{\text{on}})} \left( 1 + \frac{1}{\text{SBR}} \right) \right). \quad (30)$$

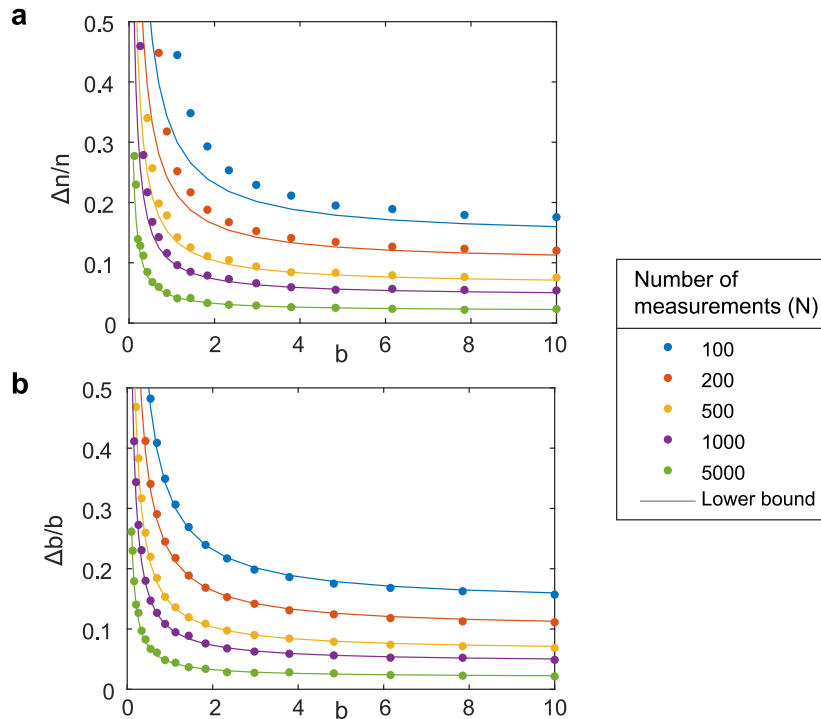
Here SBR denotes the signal to background ratio, given by

$$\text{SBR} = \frac{m-d}{d}. \quad (31)$$

Interestingly, the expression (30) for the lower bound of the relative error does not depend on the number of molecules except in the contribution of the signal to background ratio. This result is very similar to corresponding analyses in fluorescence fluctuations theory ([30]). The number of measurements  $N$ , the brightness  $b$  and the activation probability  $p_{\text{on}}$  mainly determine the relative errors.

### C.2. Simulation

We simulated RESOLFT data for the simple case of measuring a single, isolated cluster of  $n_0$  fluorophores located directly at the focus. The brightness  $b$ , assumed to be equal for all molecules, and the number of repetitions of the measurements were varied. The number and the brightness were estimated according to Eq. (27). The relative errors of the estimation are shown



**Fig. 7.** Simulation of isolated clusters of molecules. Relative standard deviation of the estimated number (a) and estimated brightness (b) for  $n_0 = 20$  molecules in one cluster, with an activation probability of 20%. Above a minimum threshold brightness, the counting error depends mainly on the number of measurements ( $N$ ). Full simulation results are plotted as dots, the calculated expressions of Eq. (30) are shown as lines.

in Fig. 7. The lower bounds derived analytically are shown as comparison. The simulations confirm that the true number and brightness can be inferred accurately. However, the estimates for the number and focal brightness are highly correlated and will, in general, be scattered along a hyperbola in number and brightness space. The relative error of brightness and number is rather independent of the number of molecules and also rather independent of the brightness above a certain brightness threshold of 1-2 photons. The error decreases with the square root of the number of repetitions. Above 500 measurements, the relative error is below 10%.

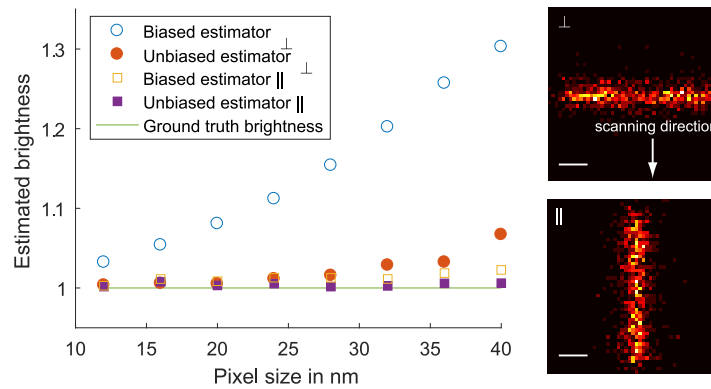
#### D. Estimator of the variance for shifted images

Due to the image acquisition procedure described in Section 3.2, the recorded image is split in one direction to deliver two independent images  $Y_1$  and  $Y_2$  with square pixels of the same structure but shifted by half a pixel with respect to each other. Depending on the structure, this can lead to a bias in the estimation of the focal brightness, as there are two contributions to the variance: first, the variance due to the switching fluorophores and the photon detection, which is included in our model, and second, the variance due to changes of the structure on the scale of the pixel shift. In the estimator for the focal brightness devised in Section 2.3, this second source of variance is not included. However, the bias due to this effect can be suppressed by estimating the pixel-wise variance in the following way: we examine the squared deviation of the signal in each pixel of one of the images  $Y_1$  to the average signal of the pixels measured directly before and after this pixel in  $Y_2$ :

$$\hat{v}_1(\vec{s}) = \left( Y_1(\vec{s}) - \frac{1}{2} (Y_2(\vec{s}_-) + Y_2(\vec{s}_+)) \right)^2. \quad (32)$$

If the variance of  $Y_2(\vec{s}_-)$  equals the variance of  $Y_2(\vec{s}_+)$  and of  $Y_1(\vec{s})$ , the expectation value of this quantity is  $\mathbb{E}[\hat{v}_1(\vec{s})] = 3/2 \text{Var}[Y_1(\vec{s})]$ . Defining  $Y'_2 = \frac{1}{2} (Y_2(\vec{s}_-) + Y_2(\vec{s}_+))$ , the variance is estimated using

$$\hat{v}_{\text{shift}}(\vec{s}) = \frac{2}{3} (Y_1(\vec{s}) - Y'_2(\vec{s}))^2. \quad (33)$$



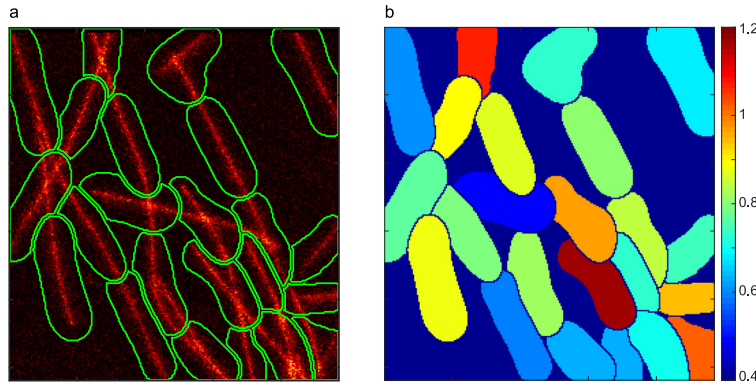
**Fig. 8.** Influence of the pixel shift between the two images  $Y_1$  and  $Y_2$  after splitting the recorded image. Simulations of images with line-like structures parallel ( $\parallel$ ) and perpendicular ( $\perp$ ) to the scanning direction. The biased estimator simply ignores the pixel shift, while the unbiased estimator implements Eq. (33). Scale bars, 200nm.

Such variance estimators were analyzed in [27]. To show that this estimator reduces the bias due to the structure in the image, we conducted simulations of line-like structures, both parallel and perpendicular to the scanning direction. The scanning process was simulated realistically

by evaluating the signal at positions shifted by half a pixel. The simulation results shown in Fig. 8 demonstrate, that the bias is pronounced mainly for structures that are perpendicular to the scanning direction and is reduced strongly using the estimator (33).

### E. Estimation of the brightness in regions within images of rsEGFP2- $\alpha$ -tubulin

The data of Fig. 5(a) was taken and regions were defined (see Fig. 9(a)). Then the brightness was estimated equivalently to Sec. 3.2 but only with the data of each region. The estimated brightnesses are shown in Fig. 9(b). The average estimated brightness of all regions is 0.79 with a standard deviation of 0.16. A systematically inhomogeneous brightness distribution like for example a gradient of the brightness cannot be observed. If the regions would have been made smaller, the fluctuations in the estimation of the brightness would naturally have increased.



**Fig. 9.** Estimation of the brightness in regions of the image of Fig. 5(a). The regions were chosen manually to adapt to the structure and include similar amounts of structure. Their boundaries are shown as green lines in **a**. In **b** the estimated brightness for each region using only data from this region is shown as color in the respective regions.

Additionally, the differences in estimated brightness using only the left and right half of the image, the upper and lower half of the image and the inner and outer half of the image were calculated. The differences in the estimated brightness between these half-image regions was always smaller than 2%.

### F. Variance of the estimated focal readout brightness $b$

The estimator  $\hat{b}$  in (17) with the bias correction from Section D has the form

$$\hat{b} = \frac{H_1}{H_2} \frac{\frac{1}{|\bar{X}|} \sum_{\bar{x} \in X} \left[ \frac{2}{3} \left( Y(\bar{x}) - \frac{1}{2} Y(\bar{x}_-) - \frac{1}{2} Y(\bar{x}_+) \right)^2 - Y(\bar{x}) \right]}{\frac{1}{|\bar{X}|} \sum_{\bar{x} \in X} \left( Y(\bar{x}) - \hat{d}(\bar{x}) \right)}, \quad (34)$$

where  $Y(\bar{x})$  are random variables following the model from Eqs. (9), (10) and (12), and  $\hat{d}(\bar{x})$  is an estimator for the intensity of the Poisson background  $d(\bar{x})$ . The constants  $H_1$  and  $H_2$  are defined in the line following Eq. (15) in the main text (see also Assumption 2 below). For simplicity of notation, we do not write explicitly in Eq. (34) and in the following the sample splitting used in Section D to estimate the variance. Instead, we implement the idea of Section D by subtracting the mean of the neighbor pixels values  $\frac{1}{2} (Y(\bar{x}_-) + Y(\bar{x}_+))$  in the numerator of (34). Here,  $\bar{x}_-$  and  $\bar{x}_+$  denote the positions of the pixels before and after  $\bar{x}$  along the scanning direction (see Section D for details).

In this section we derive an asymptotic expression for the variance of  $\hat{b}$  in the limit of many fluorophores. This limit is realized by a series of experiments where the sampled region  $X'_k$  grows in size with  $k$  and where the number of fluorophores  $n_0^{(k)}$  in  $X'_k$  also grows with  $k$ . Each experiment yields a set of observations  $\{Y_k(\vec{x})\}_{\vec{x} \in X'_k}$ , which correspond to photon counts at the scanning positions  $\vec{x} \in X'_k$ . In the following, all limits correspond to the asymptotics where the size of the sampling region  $|X'_k|$  and the number of fluorophores  $n_0^{(k)}$  in  $X'_k$  tend to infinity.

Note that in the estimator (34) we need an estimator  $\hat{d}_k(\vec{x})$  for the intensity of the Poisson background. The estimator is constructed by selecting a subset  $\tilde{X}_k$  of the sampled region where no signal is present, and using the mean signal in that region as an estimator for the background, i.e.,

$$\hat{d}_k(\vec{x}) = \frac{1}{|\tilde{X}_k|} \sum_{\vec{s} \in \tilde{X}_k} Y_k(\vec{s}). \tag{35}$$

In particular, the estimator  $\hat{d}_k(\vec{x})$  is spatially constant, and hence is expected to perform badly if the true background has strong variations in its intensity. If the background has approximately constant intensity, then  $\hat{d}_k(\vec{x})$  is expected to perform better the larger the set  $\tilde{X}_k$  is. We will hence assume that the set  $\tilde{X}_k$  grows in size as  $k \rightarrow \infty$ , but it remains small in comparison with  $X'_k$ . Finally, in order to simplify the computations, we will split the sampled region  $X'_k$  into the sets  $\tilde{X}_k$  and  $X_k = X'_k \setminus \tilde{X}_k$ , and use the first region to estimate the Poisson background, and the second to estimate the variance and the expectation of the signal. Hence, for each  $k \in \mathbb{N}$ , our estimator for the brightness is given by

$$\hat{b}_k = \frac{H_1^{(k)} \frac{1}{|\tilde{X}_k|} \sum_{\vec{x} \in X_k} \left( \frac{2}{3} \left( Y_k(\vec{x}) - \frac{1}{2} Y_k(\vec{x}_-) - \frac{1}{2} Y_k(\vec{x}_+) \right)^2 - Y_k(\vec{x}) \right)}{H_2^{(k)} \frac{1}{|\tilde{X}_k|} \sum_{\vec{x} \in X_k} \left( Y_k(\vec{x}) - \hat{d}_k(\vec{x}) \right)}, \tag{36}$$

where  $H_1^{(k)} = \sum_{\vec{x} \in \tilde{X}_k} h_{\text{read}}(\vec{x}) p_{\text{act}}(\vec{x})$  and  $H_2^{(k)} = \sum_{\vec{x} \in X_k} h_{\text{read}}^2(\vec{x}) p_{\text{act}}(\vec{x}) (1 - p_{\text{act}}(\vec{x}))$ .

We make the following assumptions:

- 1) The density of fluorophores converges to a constant  $\rho \in (0, \infty)$ , i.e.,

$$\lim_{k \rightarrow \infty} \frac{n_0^{(k)}}{|X'_k|} = \rho.$$

Moreover, the *local density* of fluorophores is also bounded by a constant  $\tilde{\rho} \in (0, \infty)$ , i.e.,

$$\lim_{k \rightarrow \infty} \max_{\vec{x} \in X_k} \{\text{Number of molecules at } \vec{x}\} = \tilde{\rho}.$$

- 2) The PSFs  $h_{\text{read}}$  and  $p_{\text{act}}$  do not change with  $k$ , and have bounded support:  $\text{supp } h_{\text{read}} \subset S$ ,  $\text{supp } p_{\text{act}} \subset S$ , with  $|S| < \infty$ . Moreover, the quantities

$$H_1^{(k)} = \sum_{\vec{x} \in \tilde{X}_k} h_{\text{read}}(\vec{x}) p_{\text{act}}(\vec{x}),$$

$$H_2^{(k)} = \sum_{\vec{x} \in X_k} h_{\text{read}}^2(\vec{x}) p_{\text{act}}(\vec{x}) (1 - p_{\text{act}}(\vec{x})),$$

converge to constants  $H_1$  and  $H_2$  as  $k \rightarrow \infty$ .

3) The background noise is independently Poisson distributed  $D(\vec{x}) \sim \text{Poisson}(d(\vec{x}))$  with intensity satisfying

$$\lim_{k \rightarrow \infty} \frac{1}{|X_k|} \sum_{\vec{x} \in X_k} d(\vec{x}) = \bar{D} \in (0, \infty).$$

Our analysis below uses the fact that  $\bar{D} > 0$ , i.e. that there is background noise. We remark nevertheless that under good experimental conditions, the average noise level  $\bar{D}$  might be quite small.

Finally, the region  $\tilde{X}_k$  is chosen such that the estimator  $\hat{d}_k(\vec{x})$  in (35) satisfies

$$\begin{aligned} \lim_{k \rightarrow \infty} \mathbb{E}[\hat{d}_k(\vec{x})] &= \bar{D}, \\ \text{Var} \left[ \sum_{\vec{x} \in \tilde{X}_k} \hat{d}_k(\vec{x}) \right] &= C_1 |X_k| + o(|X_k|), \\ \mathbb{E}[\hat{d}_k(\vec{x})^4] &\leq C_2, \end{aligned}$$

as  $k \rightarrow \infty$ , for some constants  $C_1, C_2 > 0$ . Given the structure of  $\hat{d}_k(\vec{x})$  in (35), these hold for instance if the signal in the region  $\tilde{X}_k$  is pure background, the background intensity is constant, and  $\lim_{k \rightarrow \infty} \frac{|\tilde{X}_k|}{|X_k|} \in (0, \infty)$ .

Define the quantities

$$\begin{aligned} w_k(\vec{x}) &= \mathbb{E} \left[ \frac{2}{3} \left( Y_k(\vec{x}) - \frac{1}{2} Y_k(\vec{x}_-) - \frac{1}{2} Y_k(\vec{x}_+) \right)^2 - Y_k(\vec{x}) \right], \\ W_k(\vec{x}) &= \frac{2}{3} \left( Y_k(\vec{x}) - \frac{1}{2} Y_k(\vec{x}_-) - \frac{1}{2} Y_k(\vec{x}_+) \right)^2 - Y_k(\vec{x}) - w_k(\vec{x}), \\ z_k(\vec{x}) &= \mathbb{E} [ Y_k(\vec{x}) - \hat{d}_k(\vec{x}) ], \\ Z_k(\vec{x}) &= Y_k(\vec{x}) - \hat{d}_k(\vec{x}) - z_k(\vec{x}), \end{aligned}$$

Then the estimator  $\hat{b}_k$  in (36) can be written as

$$\hat{b}_k := \frac{H_1 \frac{1}{|X_k|} \sum_{\vec{x} \in X_k} W_k(\vec{x}) + w_k(\vec{x})}{H_2 \frac{1}{|X_k|} \sum_{\vec{x} \in X_k} Z_k(\vec{x}) + z_k(\vec{x})}.$$

In the following we will show that the estimator  $\hat{b}_k$  satisfies a central limit theorem of the form

$$\sqrt{|X_k|} \left( \hat{b}_k - \mathbb{E}[\hat{b}_k] \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2) \tag{37}$$

as  $k \rightarrow \infty$  for a certain asymptotic variance  $\sigma^2$ . This will be proved in the following steps:

- 1) in Proposition 1, we show that the joint distribution of the quantities  $\frac{1}{\sqrt{|X_k|}} \sum_{\vec{x} \in X_k} W_k(\vec{x})$  and  $\frac{1}{\sqrt{|X_k|}} \sum_{\vec{x} \in X_k} Z_k(\vec{x})$  is asymptotically normal;
- 2) in Proposition 2, we apply the delta method to show that (37) holds.

In these two steps, we will see that the asymptotic variance  $\sigma^2$  in (37) can be computed explicitly. Its expression is nevertheless very involved, and hence not very informative. In order to derive a

useful approximation for the asymptotic variance of  $\hat{b}_k$ , we will approximate the random variables  $Y_k(\vec{x})$  by normal random variables with matching expectation and variance,

$$\tilde{Y}_k(\vec{x}) \sim \mathcal{N}(\mathbb{E}[Y_k(\vec{x})], \text{Var}[Y_k(\vec{x})]), \quad \vec{x} \in X_k,$$

independent from each other. This approximation is motivated by the fact that the distribution function of a Poisson random variable with large intensity can be well approximated by the distribution function of normal random variable with matching expectation and variance. In fact, the approximation error can be shown to behave like the inverse of the square root of the intensity [31]. In our case, the intensity will be large if there are many fluorophores in the sample, if their activation probability is not too small, and if the background noise is sufficiently large. We remark that our observations  $Y_k(\vec{x})$  are not Poisson, but Poisson conditionally on the molecules being active, which results in an increased variance. We compensate for this by approximating  $Y_k(\vec{x})$  by normal random variables with the same increased variance.

We will further assume that the expectation  $\mathbb{E}[Y_k(\vec{x})]$  and the variance  $\text{Var}[Y_k(\vec{x})]$  vary slowly with the position, i.e.  $\mathbb{E}[Y_k(\vec{x})] \approx \mathbb{E}[Y_k(\vec{x}_\pm)]$  and  $\text{Var}[Y_k(\vec{x})] \approx \text{Var}[Y_k(\vec{x}_\pm)]$ . This assumption is justified by the fact that the pixel size is smaller than the scale at which the PSFs vary.

Under these approximations, we show in Section F.2 that the asymptotic variance  $\sigma^2$  in (37) can be approximated by

$$\tilde{\sigma}^2 = \frac{1}{(\rho b H_1 H_2)^2} \left( \rho b (H_1 + b H_2)^3 + \bar{D} (H_1 + b H_2)^2 + \frac{H_1^2}{|X_k|} \sum_{\vec{x} \in X_k} \left( \frac{35}{9} v_k(\vec{x})^2 - \frac{5}{9} v_k(\vec{x}) m_k(\vec{x})^2 + v_k(\vec{x}) m_k(\vec{x}) \right) \right), \quad (38)$$

for large  $k$ , where  $m_k(\vec{x}) = \mathbb{E}[Y_k(\vec{x})]$  and  $v_k(\vec{x}) = \text{Var}[Y_k(\vec{x})]$ . Using this approximation and the limit theorem (37) for  $\hat{b}_k$ , we estimate the standard deviation of the estimator  $\hat{b}_k$  by  $\frac{\tilde{\sigma}}{\sqrt{|X_k|}}$ . Notice that  $\tilde{\sigma}^2$  behaves like  $b^{-2}$  for small values of  $b$ . This is consistent with the fact that our method of moments estimator is based on the relation  $\text{Var}[Y_k(\vec{x})] - \mathbb{E}[Y_k(\vec{x})] = O(b^2)$ , and hence is expected to perform poorly when  $b$  is small. On the other hand,  $\tilde{\sigma}^2$  behaves like  $b^2$  for large brightness  $b$ , which reflects the natural fact that, the larger the parameter to be estimated, the larger the absolute error made. Finally, notice that the relative standard deviation of  $\hat{b}_k$ , given by  $\frac{\tilde{\sigma}}{b}$ , tends to a constant for large brightness. We verify this effect in the simulations shown in Fig. 7.

We remark that Eq. (38) is a generalization of the result in Section C. In fact, the equation derived there is essentially (38) keeping only the term  $\frac{H_1^2}{|X_k|} \sum_{\vec{x} \in X_k} v_k(\vec{x})^2$ , which is in that situation the dominant term.

Finally, we stress that Eq. (38) is valuable inasmuch as it characterizes the error made by the estimator  $\hat{b}_k$ . Indeed, if we know the parameters of the problem, i.e., PSFs, brightness, noise level and distribution of the fluorophores, then the standard deviation of the estimator  $\hat{b}_k$  can be estimated by  $\frac{\tilde{\sigma}}{\sqrt{|X_k|}}$ . We compute this quantity in the setting of the simulations of Section 3.3, which gives the standard deviation and relative standard deviation

$$\frac{\tilde{\sigma}}{\sqrt{|X_k|}} = 0.041,$$

$$\frac{\tilde{\sigma}}{b\sqrt{|X_k|}} = 0.049,$$

for the value  $b = 0.83$  used in the simulations. The relative standard deviation computed from the simulations in Fig. 5(f) agrees with these predictions.

In the following Sections F.1 and F.2, we prove the asymptotic normality of  $\hat{b}_k$  and derive the expression (38) as an approximation for its asymptotic variance.

F.1. Asymptotic normality of  $\hat{b}_k$

Here we will prove (37) as sketched in the previous subsection. First, we prove asymptotic normality of the joint distribution of  $\frac{1}{\sqrt{|X_k|}} \sum_{\vec{x} \in X_k} W_k(\vec{x})$  and  $\frac{1}{\sqrt{|X_k|}} \sum_{\vec{x} \in X_k} Z_k(\vec{x})$ .

**Proposition 1** *In the model presented in Section F with assumptions 1), 2) and 3), the limit*

$$\Sigma := \lim_{k \rightarrow \infty} \frac{1}{|X_k|} \begin{pmatrix} \text{Var} [\sum_{\vec{x} \in X_k} W_k(\vec{x})] & \text{Cov} (\sum_{\vec{x} \in X_k} W_k(\vec{x}), \sum_{\vec{x} \in X_k} Z_k(\vec{x})) \\ \text{Cov} (\sum_{\vec{x} \in X_k} W_k(\vec{x}), \sum_{\vec{x} \in X_k} Z_k(\vec{x})) & \text{Var} [\sum_{\vec{x} \in X_k} Z_k(\vec{x})] \end{pmatrix}$$

exists and is a covariance matrix. Moreover, we have

$$\frac{1}{\sqrt{|X_k|}} \begin{pmatrix} \sum_{\vec{x} \in X_k} W_k(\vec{x}) \\ \sum_{\vec{x} \in X_k} Z_k(\vec{x}) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma \right) \tag{39}$$

as  $k \rightarrow \infty$ .

The main difficulty in proving Proposition 1 lays in the correlation structure of the random variables  $W_k(\vec{x})$  and  $Z_k(\vec{x})$ . In order to take care of the correlations, we use a central limit theorem from [32] for  $m$ -dependent triangular arrays. In our case we have  $m = 3$ , since the random variable  $W_k(\vec{x})$  is correlated with  $W_k(\vec{x}_-)$  and  $W_k(\vec{x}_+)$  and is independent of all other  $W_k(\vec{x}')$ . For each  $k \in \mathbb{N}$ , the sequence  $W_k(\vec{x})$  forms a 3-dependent sequence of random variables indexed by  $\vec{x}$ .

**Proof of Proposition 1:** First, we show that  $\Sigma$  is a covariance matrix. Since  $\Sigma$  is symmetric, it is enough to show that

$$t^T \Sigma t > 0 \tag{40}$$

for all  $t \in \mathbb{R}^2$  with  $t \neq (0, 0)^T$ . In the proof of Eq. (45) below we show that, for any  $k \in \mathbb{N}$ ,

$$t^T \Sigma_k t \geq C \tag{41}$$

for any  $t \in \mathbb{R}^2$ ,  $t \neq (0, 0)^T$ , for a constant  $C > 0$  independent of  $k$ , and

$$\Sigma_k := \frac{1}{|X_k|} \begin{pmatrix} \text{Var} [\sum_{\vec{x} \in X_k} W_k(\vec{x})] & \text{Cov} (\sum_{\vec{x} \in X_k} W_k(\vec{x}), \sum_{\vec{x} \in X_k} Z_k(\vec{x})) \\ \text{Cov} (\sum_{\vec{x} \in X_k} W_k(\vec{x}), \sum_{\vec{x} \in X_k} Z_k(\vec{x})) & \text{Var} [\sum_{\vec{x} \in X_k} Z_k(\vec{x})] \end{pmatrix}.$$

Since  $\Sigma = \lim_{k \rightarrow \infty} \Sigma_k$ , Eq. (41) implies (40).

Next, we prove the convergence in (39). For that, we use the Cramér-Wold argument (see Theorem 29.4 in [33]), which in our setting states that (39) holds if and only if the one dimensional limit

$$\frac{1}{\sqrt{|X_k|}} \left( t_1 \sum_{\vec{x} \in X_k} W_k(\vec{x}) + t_2 \sum_{\vec{x} \in X_k} Z_k(\vec{x}) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, t^T \Sigma t) \tag{42}$$

holds for any  $t = (t_1, t_2)$  with  $t_1^2 + t_2^2 = 1$ . Notice that the variance of the right-hand side is

$$t^T \Sigma t = t_1^2 \Sigma_{1,1} + t_2^2 \Sigma_{2,2} + 2t_1 t_2 \Sigma_{1,2} = \lim_{k \rightarrow \infty} \frac{1}{|X_k|} \text{Var} \left[ \sum_{\vec{x} \in X_k} t_1 W_k(\vec{x}) + t_2 Z_k(\vec{x}) \right].$$

In order to prove (42), we apply Theorem 2.1 in [32]. It applies to our situation, since we consider the triangular array  $t_1 W_k(\vec{x}) + t_2 Z_k(\vec{x})$ ,  $k \in \mathbb{N}$ ,  $\vec{x} \in X_k$  of 3-dependent random variables. For



simplicity of the statement and the computations, we will re-index the random variables  $W_k(\vec{x})$  and  $Z_k(\vec{x})$  by an index  $i = 1, \dots, |X_k|$  such that

$$W_{k,i} := W_k(\vec{x}_i), \quad Z_{k,i} := Z_k(\vec{x}_i)$$

are ordered so that  $W_{k,i}$ ,  $W_{k,i-1}$  and  $W_{k,i-2}$  have nonzero correlation. Assume for the moment that, for fixed  $t_1^2 + t_2^2 = 1$ , the conditions

$$\mathbb{E}[|t_1 W_{k,i} + t_2 Z_{k,i}|^{2+\delta}] \leq C_1 < \infty \quad \text{for all } i = 1, \dots, |X_k|, \quad \text{and all } k \in \mathbb{N}, \quad (43)$$

$$\text{Var} \left[ \sum_{i=a}^{a+L-1} t_1 W_{k,i} + t_2 Z_{k,i} \right] \leq C_2 L \quad \text{for all } a = 1, \dots, |X_k| - L + 1, \quad L = 2, \dots, |X_k| \quad \text{and all } k \in \mathbb{N}, \quad (44)$$

$$\text{Var} \left[ \sum_{i=1}^{|X_k|} t_1 W_{k,i} + t_2 Z_{k,i} \right] \geq C_3 |X_k| \quad \text{for all } k \in \mathbb{N}, \quad (45)$$

hold for some constants  $C_1, C_2, C_3 > 0$ . Then Theorem 2.1 in [32] states that

$$\frac{1}{\sqrt{\text{Var} \left[ \sum_{i=1}^{|X_k|} t_1 W_{k,i} + t_2 Z_{k,i} \right]}} \sum_{i=1}^{|X_k|} t_1 W_{k,i} + t_2 Z_{k,i} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

holds as  $k \rightarrow \infty$ . In our original notation, this means that

$$\frac{1}{\sqrt{\text{Var} \left[ \sum_{\vec{x} \in X_k} t_1 W_k(\vec{x}) + t_2 Z_k(\vec{x}) \right]}} \sum_{\vec{x} \in X_k} t_1 W_k(\vec{x}) + t_2 Z_k(\vec{x}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

holds as  $k \rightarrow \infty$ , which is what we wanted to prove. It remains to verify the conditions (43) to (45).

For condition (43) with  $\delta = 2$ , we use that  $|t_1|, |t_2| \leq 1$  and bound

$$\begin{aligned} \mathbb{E}[|t_1 W_{k,i} + t_2 Z_{k,i}|^4] &\leq \mathbb{E}[W_{k,i}^4] + 4|\mathbb{E}[W_{k,i}^3 Z_{k,i}]| + 6\mathbb{E}[W_{k,i}^2 Z_{k,i}^2] + 4|\mathbb{E}[W_{k,i} Z_{k,i}^3]| + \mathbb{E}[Z_{k,i}^4] \\ &\leq C \max_{j=i, i-1, i-2} \left( \mathbb{E}[Y_k(\vec{x}_j)^8] + \mathbb{E}[\hat{d}_k(\vec{x}_i)^4] \right), \end{aligned}$$

where  $C > 0$  is a constant and we used that  $W_{k,i}$  and  $Z_{k,i}$  are polynomials in  $Y_k(\vec{x}_i)$ ,  $Y_k(\vec{x}_{i-1})$ ,  $Y_k(\vec{x}_{i-2})$  and  $\hat{d}_k(\vec{x}_i)$ , the maximal power of the product polynomials is 8, and the lower order moments of random variables can be bounded by the higher order moments. Now, notice that by assumption, the intensity  $d(\vec{x}_i)$  of the background is bounded for all  $\vec{x}_i$ . Hence,  $\mathbb{E}[\hat{d}_k(\vec{x}_i)^4]$  is bounded. On the other hand, the random variable  $Y_k(\vec{x}_j)$  is the sum of independent Poisson random variables with Bernoulli weights (recall the model in Eqs. (9), (10) and (12)). By the assumption of finite support of the PSFs and the finite density of molecules, we can bound  $|Y_k(\vec{x}_j)|$  by the number of fluorophores that contribute to the sum, which is finite, times the maximum of a finite number of Poisson random variables with bounded intensity. The eighth moment of such a random variable is finite, and we conclude that  $\mathbb{E}[Y_k(\vec{x}_j)^8] \leq C < \infty$ . Altogether, we conclude that (43) holds.

For (44), we bound

$$\begin{aligned}
 \text{Var} \left[ \sum_{i=a}^{a+L-1} t_1 W_{k,i} + t_2 Z_{k,i} \right] &= \sum_{i=a}^{a+L-1} \text{Var} [t_1 W_{k,i} + t_2 Z_{k,i}] + \\
 &\quad + 2 \sum_{i=a}^{a+L-1} \text{Cov} [t_1 W_{k,i} + t_2 Z_{k,i}, t_1 W_{k,i-1} + t_2 Z_{k,i-1}] + \\
 &\quad + 2 \sum_{i=a}^{a+L-1} \text{Cov} [t_1 W_{k,i} + t_2 Z_{k,i}, t_1 W_{k,i-2} + t_2 Z_{k,i-2}] \\
 &\leq \sum_{i=a}^{a+L-1} \text{Var} [t_1 W_{k,i} + t_2 Z_{k,i}] + \\
 &\quad + 2 \sum_{i=a}^{a+L-1} \sqrt{\text{Var} [t_1 W_{k,i} + t_2 Z_{k,i}] \text{Var} [t_1 W_{k,i-1} + t_2 Z_{k,i-1}]} + \\
 &\quad + 2 \sum_{i=a}^{a+L-1} \sqrt{\text{Var} [t_1 W_{k,i} + t_2 Z_{k,i}] \text{Var} [t_1 W_{k,i-2} + t_2 Z_{k,i-2}]} \\
 &\leq 5 \sum_{i=a-2}^{a+L-1} \text{Var} [t_1 W_{k,i} + t_2 Z_{k,i}]
 \end{aligned}$$

using that  $\sqrt{xy} \leq \frac{1}{2}(x + y)$  for  $x, y \geq 0$ . Hence to prove (44) it suffices to show that

$$\text{Var} [t_1 W_{k,i} + t_2 Z_{k,i}] \leq C_2$$

for all  $i = 1, \dots, |X_k|$  and a constant  $C_2$  independent of  $k$ . Notice that

$$\begin{aligned}
 \text{Var} [t_1 W_{k,i} + t_2 Z_{k,i}] &\leq t_1^2 \text{Var} [W_{k,i}] + t_2^2 \text{Var} [Z_{k,i}] + 2t_1 t_2 \text{Cov} [W_{k,i}, Z_{k,i}] \\
 &\leq 2 (\text{Var} [W_{k,i}] + \text{Var} [Z_{k,i}])
 \end{aligned}$$

with an argument as above to bound the covariance. Now, the terms  $\text{Var} [W_{k,i}]$  and  $\text{Var} [Z_{k,i}]$  can be shown to be bounded with an argument as in the proof of (43). Indeed, there we showed that the fourth moment of  $W_{k,i}$  is bounded, which implies that  $\text{Var} [W_{k,i}]$  is bounded. The term  $\text{Var} [Z_{k,i}]$  can be bounded in the same way, and we conclude that (44) holds.

For (45), we have to bound the total variance  $\text{Var} \left[ \sum_{i=1}^{|X_k|} t_1 W_{k,i} + t_2 Z_{k,i} \right]$  from below. Obviously, the total variance will be bounded from below if the observations  $Y_k(\vec{x}_i)$  have a variance bounded from below, but to verify the quantitative statement of condition (45) we would require lengthy computations.

Alternatively, we can use the fact that the background noise and the estimator  $\hat{d}_k(\vec{x})$  give a nonzero contribution to the variance. Indeed, writing

$$Z_{k,i} = Y_k(\vec{x}_i) - \hat{d}_k(\vec{x}_i) - \mathbb{E}[Y_k(\vec{x}_i) - \hat{d}_k(\vec{x}_i)],$$

we have

$$\begin{aligned}
 \text{Var} \left[ \sum_{i=1}^{|X_k|} t_1 W_{k,i} + t_2 Z_{k,i} \right] &= \text{Var} \left[ \sum_{i=1}^{|X_k|} t_1 W_{k,i} + t_2 Y_k(\vec{x}_i) \right] + t_2^2 \text{Var} \left[ \sum_{i=1}^{|X_k|} \hat{d}_k(\vec{x}_i) \right] + \\
 &\quad - 2t_2 \text{Cov} \left( \underbrace{\sum_{i=1}^{|X_k|} t_1 W_{k,i} + t_2 Y_k(\vec{x}_i)}_{=0}, \sum_{i=1}^{|X_k|} \hat{d}_k(\vec{x}_i) \right), \tag{46}
 \end{aligned}$$

where the covariance term vanishes since  $\hat{d}_k(\vec{x}_i)$  is independent of  $W_{k,i}$  and  $Y_k(\vec{x}_i)$  by the sample splitting used to compute  $\hat{d}_k(\vec{x}_i)$ . Hence, if we have  $|t_2| \geq \epsilon$  for some fixed  $\epsilon > 0$  to be specified below, assumption 3 at the beginning of Section F gives

$$\text{Var} \left[ \sum_{i=1}^{|X_k|} t_1 W_{k,i} + t_2 Z_{k,i} \right] = \underbrace{\text{Var} \left[ \sum_{i=1}^{|X_k|} t_1 W_{k,i} + t_2 Y_k(\vec{x}_i) \right]}_{\geq 0} + t_2^2 \text{Var} \left[ \sum_{i=1}^{|X_k|} \hat{d}_k(\vec{x}_i) \right] \geq \epsilon^2 C |X_k|$$

holds for  $k$  large enough and a constant  $C > 0$ , which is what we wanted to show. If  $|t_2| < \epsilon$ , then we can express the first term in (46) as the sum of variances and covariances and get

$$\begin{aligned} \text{Var} \left[ \sum_{i=1}^{|X_k|} t_1 W_{k,i} + t_2 Z_{k,i} \right] &= t_1^2 \text{Var} \left[ \sum_{i=1}^{|X_k|} W_{k,i} \right] + t_2^2 \left( \text{Var} \left[ \sum_{i=1}^{|X_k|} Y_k(\vec{x}_i) \right] + \text{Var} \left[ \sum_{i=1}^{|X_k|} \hat{d}_k(\vec{x}_i) \right] \right) + \\ &\quad + 2t_1 t_2 \text{Cov} \left( \sum_{i=1}^{|X_k|} W_{k,i}, \sum_{i=1}^{|X_k|} Y_k(\vec{x}_i) \right). \end{aligned}$$

Note that

$$\text{Var} \left[ \sum_{i=1}^{|X_k|} Y_k(\vec{x}_i) \right] \geq |X_k| \rho (bH_1 + b^2 H_2) + |X_k| \bar{D}$$

for  $k$  large enough, and that

$$\begin{aligned} \left| \text{Cov} \left( \sum_{i=1}^{|X_k|} W_{k,i}, \sum_{i=1}^{|X_k|} Y_k(\vec{x}_i) \right) \right| &\leq \sqrt{\text{Var} \left[ \sum_{i=1}^{|X_k|} W_{k,i} \right] \text{Var} \left[ \sum_{i=1}^{|X_k|} Y_k(\vec{x}_i) \right]} \\ &\leq \max \left\{ \text{Var} \left[ \sum_{i=1}^{|X_k|} W_{k,i} \right], \text{Var} \left[ \sum_{i=1}^{|X_k|} Y_k(\vec{x}_i) \right] \right\} \\ &= \text{Var} \left[ \sum_{i=1}^{|X_k|} W_{k,i} \right] \end{aligned}$$

where the last equality follows by the definition of  $W_{k,i}$ , which gives

$$\text{Var} \left[ \sum_{i=1}^{|X_k|} W_{k,i} \right] \geq \text{Var} \left[ \sum_{i=1}^{|X_k|} Y_k(\vec{x}_i) \right]$$

Altogether we have

$$\begin{aligned} \text{Var} \left[ \sum_{i=1}^{|X_k|} t_1 W_{k,i} + t_2 Z_{k,i} \right] &\geq t_1^2 \text{Var} \left[ \sum_{i=1}^{|X_k|} W_{k,i} \right] + t_2^2 C |X_k| - 2|t_1 t_2| \left| \text{Cov} \left( \sum_{i=1}^{|X_k|} W_{k,i}, \sum_{i=1}^{|X_k|} Y_k(\vec{x}_i) \right) \right| \\ &\geq t_1^2 \text{Var} \left[ \sum_{i=1}^{|X_k|} W_{k,i} \right] + t_2^2 C |X_k| - 2|t_1 t_2| \text{Var} \left[ \sum_{i=1}^{|X_k|} W_{k,i} \right] \\ &= |t_1| (|t_1| - 2|t_2|) \text{Var} \left[ \sum_{i=1}^{|X_k|} W_{k,i} \right] + t_2^2 C |X_k| \end{aligned}$$

Assume without loss of generality that  $t_1, t_2 \geq 0$ . Then, if  $t_2 < \epsilon$  with  $\epsilon = \frac{1}{\sqrt{6}}$ , we have

$$t_1 = \sqrt{1 - t_2^2} > \sqrt{1 - \epsilon^2} = \sqrt{\frac{5}{6}},$$

and hence we have

$$t_1(t_1 - 2t_2) > t_1 \left( \frac{\sqrt{5}}{\sqrt{6}} - \frac{2}{\sqrt{6}} \right) \geq t_1 \frac{0.2}{\sqrt{6}} > 0.05.$$

So if  $t_2 < \epsilon$ , we have shown that

$$\text{Var} \left[ \sum_{i=1}^{|X_k|} t_1 W_{k,i} + t_2 Z_{k,i} \right] \geq 0.05 \text{Var} \left[ \sum_{i=1}^{|X_k|} W_{k,i} \right] + t_2^2 C |X_k|.$$

It is not difficult to show that the variance of  $\sum_{i=1}^{|X_k|} W_{k,i}$  is bounded from below by the variance of the background noise  $\text{Var} \left[ \sum_{i=1}^{|X_k|} D_i \right]$ , which by Assumption 3 behaves like  $\bar{D} |X_k|$  asymptotically. Hence the lower bound follows, and we conclude that (45) holds. ■

We will now use the delta method (see Sections 3.1 and 3.3 in [34]) to prove the asymptotic normality of  $\hat{b}_k$ .

**Proposition 2** *With the notation and assumptions of Section F, we have*

$$\sqrt{|X_k|} (\hat{b}_k - b) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

as  $k \rightarrow \infty$ , where

$$\sigma^2 = \frac{1}{(\rho b H_1 H_2)^2} (H_1^2 \Sigma_{1,1} + b^2 H_2^2 \Sigma_{2,2} - 2b H_1 H_2 \Sigma_{1,2}), \tag{47}$$

with  $\Sigma_{i,j}$  being the entries of the covariance matrix in Proposition 1,  $H_1$  and  $H_2$  the constants in Assumption 2, and  $\rho$  the density in Assumption 1.

**Proof of Proposition 2:** We prove the proposition by applying the delta method to Eq. (39). Defining

$$\alpha_k := \frac{\sum_{\vec{x} \in X_k} W_k(\vec{x}) + w_k(\vec{x})}{|X_k|},$$

$$\beta_k := \frac{\sum_{\vec{x} \in X_k} Z_k(\vec{x}) + z_k(\vec{x})}{|X_k|},$$

and the limits

$$\alpha := \lim_{k \rightarrow \infty} \frac{\sum_{\vec{x} \in X_k} w_k(\vec{x})}{|X_k|}, \quad \beta := \lim_{k \rightarrow \infty} \frac{\sum_{\vec{x} \in X_k} z_k(\vec{x})}{|X_k|},$$

Eq. (39) can be written as

$$\sqrt{|X_k|} \left( \begin{pmatrix} \alpha_k \\ \beta_k \end{pmatrix} - \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma \right).$$

Using now the delta method with the function  $f(\alpha, \beta) = \frac{\alpha}{\beta}$ , we conclude that

$$\sqrt{|X_k|} \left( \frac{\alpha_k}{\beta_k} - \frac{\alpha}{\beta} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \bar{\sigma}^2)$$

with

$$\bar{\sigma}^2 = \left( \frac{1}{\beta}, -\frac{\alpha}{\beta^2} \right) \Sigma \left( \frac{1}{\beta}, -\frac{\alpha}{\beta^2} \right)^T,$$

where

$$\alpha = \rho b (H_1 + b H_2) + \bar{D}, \quad \beta = \rho b H_1.$$

Multiplying by  $\frac{H_1^{(k)}}{H_2^{(k)}}$ , and using Assumption 2 and the fact that  $b = \frac{H_1}{H_2} \frac{\alpha}{\beta}$ , the claim follows. ■

F.2. Approximation of the variance of  $\hat{b}$

Proposition 2 gives us an explicit expression for the asymptotic variance of  $\hat{b}_k$ . This expression is nevertheless very involved and not very informative. In this section we will approximate this expression for the variance by approximating the Poisson random variables  $Y_k(\vec{x})$  by normal random variables with matching expectation and variance,

$$\tilde{Y}_k(\vec{x}) \sim \mathcal{N}(\mathbb{E}[Y_k(\vec{x})], \text{Var}[Y_k(\vec{x})]), \quad \vec{x} \in X_k,$$

independent from each other. In the following we will use the notation introduced in (13), i.e.,

$$m_k(\vec{x}) = \mathbb{E}[Y_k(\vec{x})], \quad v_k(\vec{x}) = \text{Var}[Y_k(\vec{x})].$$

As stated above, we assume that the expectation and the variance of the signals vary slowly, so that we can approximate  $m_k(\vec{x}) \approx m_k(\vec{x}_\pm)$  and  $v_k(\vec{x}) \approx v_k(\vec{x}_\pm)$ .

Define the quantity

$$\tilde{W}_k(\vec{x}) = \frac{2}{3} \left( \tilde{Y}_k(\vec{x}) - \frac{1}{2} \tilde{Y}_k(\vec{x}_-) - \frac{1}{2} \tilde{Y}_k(\vec{x}_+) \right)^2 - \tilde{Y}_k(\vec{x}).$$

In this setting, we can compute

$$\Sigma_{2,2} = \lim_{k \rightarrow \infty} \frac{1}{|X_k|} \text{Var} \left[ \sum_{\vec{x} \in X_k} \tilde{Y}_k(\vec{x}) - \hat{d}_k(\vec{x}) \right] = \rho b(H_1 + bH_2) + \bar{D}.$$

Moreover, the covariance term between  $\sum_{\vec{x} \in X_k} \tilde{W}_k(\vec{x})$  and  $\sum_{\vec{x} \in X_k} \tilde{Y}_k(\vec{x}) - \hat{d}_k(\vec{x})$  is given by

$$\begin{aligned} \frac{1}{|X_k|} \text{Cov} \left( \sum_{\vec{x} \in X_k} \tilde{W}_k(\vec{x}), \sum_{\vec{x} \in X_k} \tilde{Y}_k(\vec{x}) - \hat{d}_k(\vec{x}) \right) &= \\ &= -\frac{1}{|X_k|} \sum_{\vec{x} \in X_k} v_k(\vec{x}) + \frac{1}{|X_k|} \sum_{\vec{x}_i \in X_k} v_k(\vec{x}_i) \left( m_k(\vec{x}_i) + \frac{1}{3} m_k(\vec{x}_{i-2}) - \frac{4}{3} m_k(\vec{x}_{i-1}) \right) \\ &\quad + \frac{1}{|X_k|} \sum_{\vec{x}_i \in X_k} m_k(\vec{x}) \left( v_k(\vec{x}_i) + \frac{1}{3} v_k(\vec{x}_{i-2}) - \frac{4}{3} v_k(\vec{x}_{i-1}) \right) \\ &\approx -\frac{1}{|X_k|} \sum_{\vec{x} \in X_k} v_k(\vec{x}) \end{aligned}$$

using the assumption that the expectation and the variance vary slowly. In the limit  $k \rightarrow \infty$  this gives

$$\Sigma_{1,2} = \lim_{k \rightarrow \infty} \frac{1}{|X_k|} \text{Cov} \left( \sum_{\vec{x} \in X_k} \tilde{W}_k(\vec{x}), \sum_{\vec{x} \in X_k} \tilde{Y}_k(\vec{x}) - \hat{d}_k(\vec{x}) \right) \approx -\rho b(H_1 + bH_2) - \bar{D}$$

Finally, defining

$$T_k(\vec{x}_i) := m_k(\vec{x}_i) + \frac{1}{3} m_k(\vec{x}_{i-2}) - \frac{4}{3} m_k(\vec{x}_{i-1}),$$

the variance of  $\sum_{\vec{x} \in X_k} \tilde{W}_k(\vec{x})$  is given by

$$\begin{aligned} \text{Var} \left[ \sum_{\vec{x} \in X_k} \tilde{W}_k(\vec{x}) \right] &= \sum_{\vec{x} \in X_k} v_k(\vec{x}) + 2v_k(\vec{x})^2 \\ &+ \sum_{\vec{x}_i \in X_k} v_k(\vec{x}_i) \left( T_k(\vec{x}_i)^2 - T_k(\vec{x}_i) - m_k(\vec{x}_i) + \frac{4}{9} m_k(\vec{x}_{i-1}) m_k(\vec{x}_{i-2}) \right) \\ &+ \sum_{\vec{x}_i \in X_k} v_k(\vec{x}_{i-1}) \left( \frac{16}{9} m_k(\vec{x}_i)^2 + \frac{16}{9} v_k(\vec{x}_i) - \frac{8}{3} m_k(\vec{x}_i) (T_k(\vec{x}_{i-1}) + m_k(\vec{x}_{i-1}) - 1) \right) \\ &+ \sum_{\vec{x}_i \in X_k} v_k(\vec{x}_{i-2}) \left( \frac{1}{9} m_k(\vec{x}_i)^2 + \frac{1}{9} v_k(\vec{x}_i) - \frac{8}{9} m_k(\vec{x}_i) m_k(\vec{x}_{i-1}) + \right. \\ &\quad \left. + \frac{2}{3} m_k(\vec{x}_i) (T_k(\vec{x}_{i-2}) + m_k(\vec{x}_{i-2}) - 1) \right). \end{aligned}$$

Using again the assumption that  $m_k(\vec{x})$  and  $v_k(\vec{x})$  vary slowly, we can approximate

$$\frac{1}{|X_k|} \text{Var} \left[ \sum_{\vec{x} \in X_k} \tilde{W}_k(\vec{x}) \right] \approx \frac{1}{|X_k|} \sum_{\vec{x} \in X_k} v_k(\vec{x}) + \frac{1}{|X_k|} \sum_{\vec{x} \in X_k} \left( \frac{35}{9} v_k(\vec{x})^2 - \frac{5}{9} v_k(\vec{x}) m_k(\vec{x})^2 + v_k(\vec{x}) m_k(\vec{x}) \right),$$

and hence

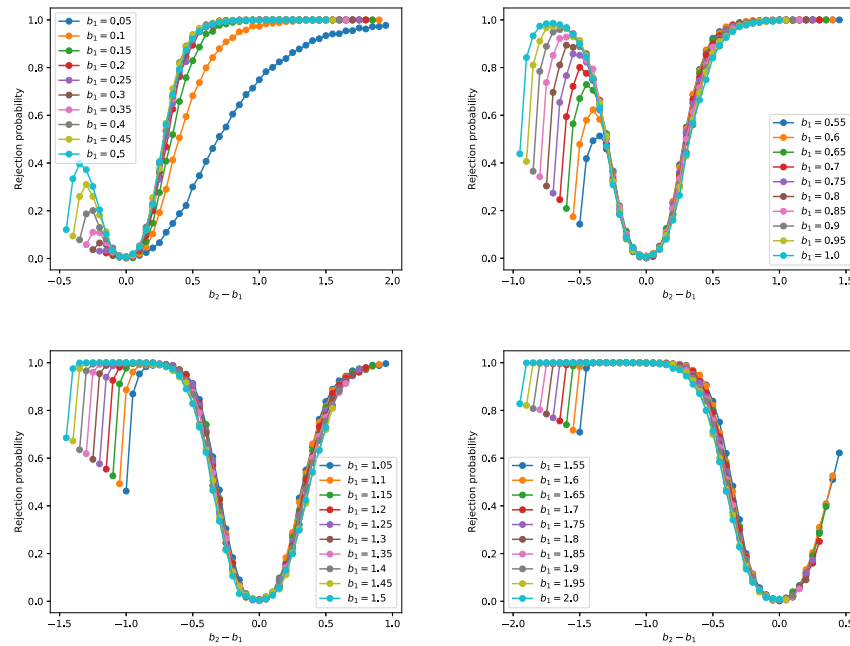
$$\Sigma_{1,1} \approx \rho b(H_1 + bH_2) + \bar{D} + \lim_{k \rightarrow \infty} \frac{1}{|X_k|} \sum_{\vec{x} \in X_k} \left( \frac{35}{9} v_k(\vec{x})^2 - \frac{5}{9} v_k(\vec{x}) m_k(\vec{x})^2 + v_k(\vec{x}) m_k(\vec{x}) \right),$$

since the first term can be computed explicitly and equals  $\rho b(H_1 + bH_2) + \bar{D}$  in the limit  $k \rightarrow \infty$ . The second term can be evaluated numerically. Inserting these expressions for  $\Sigma_{1,1}$ ,  $\Sigma_{1,2}$  and  $\Sigma_{2,2}$  in (47) yields Eq. (38).

### F.3. Verification of the homogeneity of the brightness

To test whether local differences in brightness can be detected by our method, we applied the following testing procedure to a set of simulated measurements for the model structure (see Section 3.3). 1000 images were simulated for every brightness from the set  $B = \{0.05, 0.1, 0.15, \dots, 1.95, 2.0\}$ . Each simulated image was evenly split in 8 parts (2 rows with 4 columns). We then took the 4 parts of the first row from a simulation with brightness  $b_1 \in B$  and the 4 parts of the second row from a simulation with brightness  $b_2 \in B$ . For each part, we estimated the brightness as described in Section 2.3. We then tested for homogeneity by applying the 2-sample Welch test ([28, p. 447]) to the results from the first row and the second row. The Welch test tests whether two samples have equal means without assuming that their variances are equal. Equality of variances can not be assumed here since the two samples stem from different parts of the image which are in general different. We used the test level  $\alpha = 0.01$ .

The relative frequency of rejection of the null hypothesis that the means are equal is shown in Fig. 10 for different values of  $b_1$  and  $b_2$ . It is striking that the power curves are not symmetric; this is a consequence of the fact that the two samples come from different parts of the image. For low values of  $b_2$  the variance of the second sample becomes very high such that the null hypothesis can not be rejected. A related phenomenon occurs for small  $b_1$ . Ignoring this effect, differences of 0.5 in the brightness can be detected with roughly 80% probability at this confidence level.



**Fig. 10.** Empirical power of the Welch test applied to two samples of the simulated images in Fig. 5(e) for different brightnesses. The first sample was taken from a set of four possible area from the upper half and the second sample from four possible areas from the lower half of the simulated images. The brightness of the first sample is  $b_1$  and that of the second sample  $b_2$ . The nominal test level was  $\alpha = 0.01$ .

## Funding

Deutsche Forschungsgemeinschaft (DFG) (CRC 755 (projects A6, A7 and B10)) and RTG 2088 (project B2). S.H., S.J., A.M. acknowledge support of the Cluster of Excellence "Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells".

## Acknowledgments

We are grateful to Mark Bates for initial discussions about the project, Marcel Leutenegger for assistance with the optical setup and Sylvia Löbermann for help with biological sample preparations. We would also like to thank the reviewers from OE for their thoughtful remarks.

## References

1. V. C. Coffman, P. Wu, M. P. Parthun, and J.-Q. Wu, "Cenp-a exceeds microtubule attachment sites in centromere clusters of both budding and fission yeast," *J. Cell Biol.* **195**(4), 563–572 (2011).
2. B. D. Engel, W. B. Ludington, and W. F. Marshall, "Intraflagellar transport particle size scales inversely with flagellar length: revisiting the balance-point length control model," *J. Cell Biol.* **187**(1), 81–89 (2009).
3. C. A. Wurm, D. Neumann, M. A. Latuerbach, B. Harke, A. Egner, S. W. Hell, and S. Jakobs, "Nanoscale distribution of mitochondrial import receptor tom20 is adjusted to cellular conditions and exhibits an inner-cellular gradient," *Proc. Natl. Acad. Sci.* **108**(33), 13546–13551 (2011).
4. S. W. Hell and J. Wichmann, "Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy," *Opt. Lett.* **19**(11), 780–782 (1994).
5. T. A. Klar, S. Jakobs, M. Dyba, A. Egner, and S. W. Hell, "Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission," *Proc. Natl. Acad. Sci.* **97**(15), 8206–8210 (2000).
6. M. G. Gustafsson, "Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution," *Proc. Natl. Acad. Sci.* **102**(37), 13081–13086 (2005).
7. S. Hell, "Toward fluorescence nanoscopy," *Nat. Biotechnol.* **21**(11), 1347–1355 (2003).

8. S. W. Hell, M. Dyba, and S. Jakobs, "Concepts for nanoscale resolution in fluorescence microscopy," *Curr. Opin. Neurobiol.* **14**(5), 599–609 (2004).
9. E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science* **313**(5793), 1642–1645 (2006).
10. M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nat. Methods* **3**(10), 793–796 (2006).
11. S.-H. Lee, J. Y. Shin, A. Lee, and C. Bustamante, "Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM)," *Proc. Natl. Acad. Sci.* **109**(43), 17436–17441 (2012).
12. G. Hummer, F. Fricke, and M. Heilemann, "Model-independent counting of molecules in single-molecule localization microscopy," *Mol. Biol. Cell* **27**(22), 3637–3644 (2016).
13. R. Jungmann, M. S. Avenaño, M. Dai, J. B. Woehrstein, S. S. Agasti, Z. Feiger, A. Rodal, and P. Yin, "Quantitative super-resolution imaging with qPAINT," *Nat. Methods* **13**(5), 439–442 (2016).
14. M. H. Ulbrich and E. Y. Isacoff, "Subunit counting in membrane-bound proteins," *Nat. Methods* **4**(4), 319–321 (2007).
15. M. A. Digman, R. Dalal, A. F. Horwitz, and E. Gratton, "Mapping the number of molecules and brightness in the laser scanning microscope," *Biophys. J.* **94**(6), 2320–2332 (2008).
16. L. Kastrop, H. Blom, C. Eggeling, and S. W. Hell, "Fluorescence fluctuation spectroscopy in subdiffraction focal volumes," *Phys. Rev. Lett.* **94**(17), 178104 (2005).
17. H. Ta, J. Keller, M. Haltmeier, S. K. Saka, J. Schmied, F. Opazo, P. Tinnefeld, A. Munk, and S. W. Hell, "Mapping molecules in scanning far-field fluorescence nanoscopy," *Nat. Commun.* **6**(1), 7977 (2015).
18. T. Grotjohann, I. Testa, M. Leutenegger, H. Bock, N. T. Urban, F. Lavoie-Cardinal, K. I. Willig, C. Eggeling, S. Jakobs, and S. W. Hell, "Diffraction-unlimited all-optical imaging and writing with a photochromic GFP," *Nature* **478**(7368), 204–208 (2011).
19. T. Grotjohann, I. Testa, M. Reuss, T. Brakemann, C. Eggeling, S. W. Hell, and S. Jakobs, "rsEGFP2 enables fast RESOLFT nanoscopy of living cells," *eLife* **1**, e00248 (2012).
20. I. Testa, N. T. Urban, S. Jakobs, C. Eggeling, K. I. Willig, and S. W. Hell, "Nanoscopy of Living Brain Slices with Low Light Levels," *Neuron* **75**(6), 992–1000 (2012).
21. B. Harke, J. Keller, C. K. Ullal, V. Westphal, A. Schönle, and S. W. Hell, "Resolution scaling in STED microscopy," *Opt. Express* **16**(6), 4154–4162 (2008).
22. B. Roubinet, M. L. Bossi, P. Alt, M. Leutenegger, H. Shojaei, S. Schnorrenberg, S. Nizamov, M. Irie, V. N. Belov, and S. W. Hell, "Carboxylated Photoswitchable Diarylethenes for Biolabeling and Super-Resolution RESOLFT Microscopy," *Angew. Chem., Int. Ed.* **55**(49), 15429–15433 (2016).
23. S. Schnorrenberg, T. Grotjohann, G. Vorbrüggen, A. Herzig, S. W. Hell, and S. Jakobs, "In vivo super-resolution RESOLFT microscopy of *Drosophila melanogaster*," *eLife* **5**, e15567 (2016).
24. L. A. Amos and A. Klug, "Arrangement of subunits in flagellar microtubules," *Journal of cell science* **14**, 523–549 (1974).
25. N. Joshi, R. Szeliski, and D. J. Kriegman, "PSF estimation using sharp edge prediction," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, (IEEE, 2008), pp. 1–8.
26. C. Steger, "An unbiased detector of curvilinear structures," *IEEE Trans. Pattern Anal. Machine Intell.* **20**(2), 113–125 (1998).
27. A. Munk, N. Bissantz, T. Wagner, and G. Freitag, "On difference-based variance estimation in nonparametric regression when the covariate is high dimensional," *J. Royal Stat. Soc. Ser. B (Statistical Methodol.)* **67**(1), 19–41 (2005).
28. E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses* (Springer Science & Business Media, 2006).
29. D. Bourgeois and V. Adam, "Reversible photoswitching in fluorescent proteins: a mechanistic view," *IUBMB Life* **64**(6), 482–491 (2012).
30. D. E. Koppel, "Statistical accuracy in fluorescence correlation spectroscopy," *Phys. Rev. A* **10**(6), 1938–1945 (1974).
31. T. T. Cheng, "The normal approximation to the Poisson distribution and a proof of a conjecture of Ramanujan," *Bull. Amer. Math. Soc.* **55**(4), 396–402 (1949).
32. J. P. Romano and M. Wolf, "A more general central limit theorem for m-dependent random variables with unbounded m," *Stat. & probability letters* **47**(2), 115–124 (2000).
33. P. Billingsley, *Probability and Measure* (Wiley, New York, 1979).
34. A. W. van der Vaart, *Asymptotic Statistics* (Cambridge University, 1998).