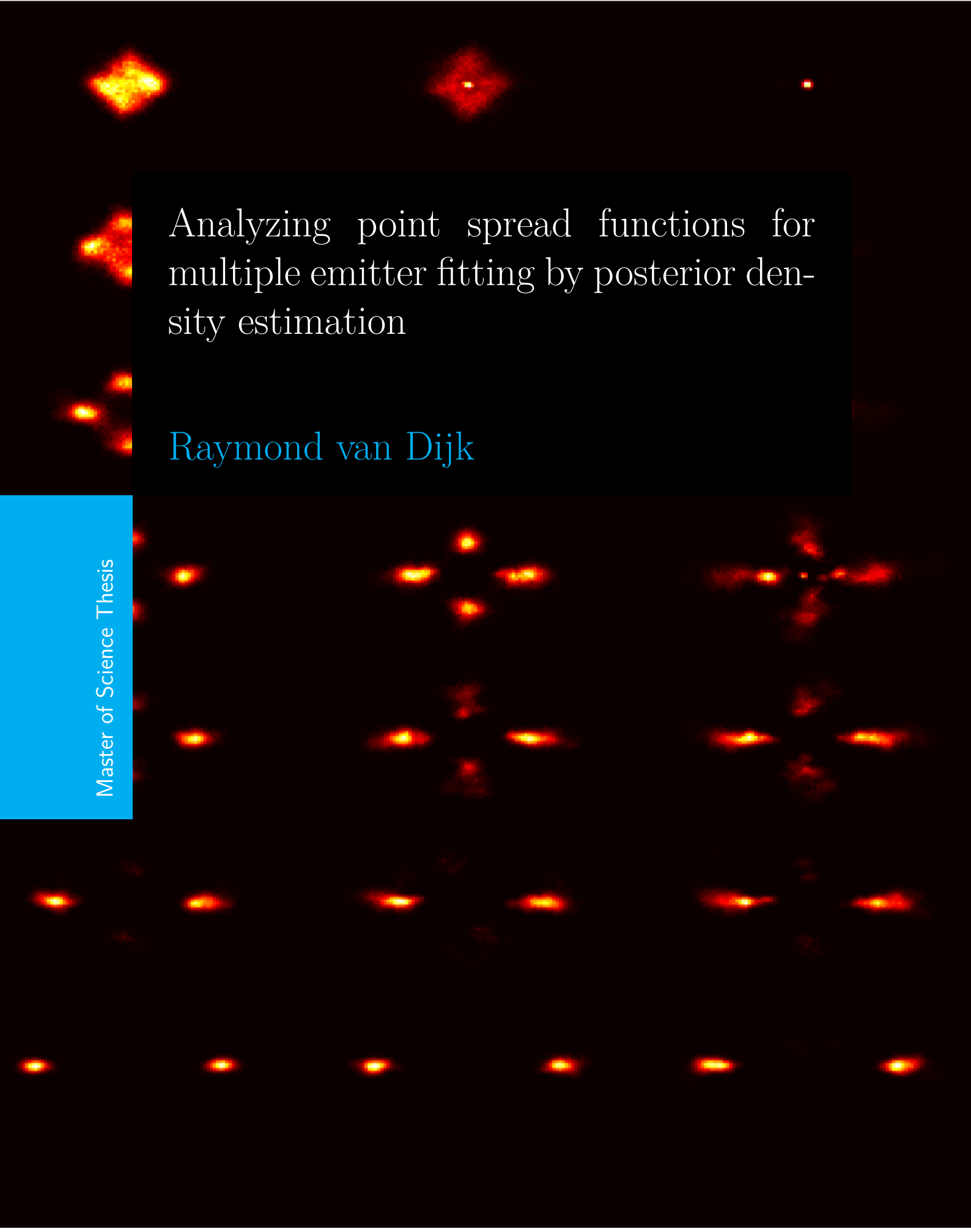


Analyzing point spread functions for multiple emitter fitting by posterior density estimation

Raymond van Dijk

Master of Science Thesis

TU Delft Delft University of Technology

Delft Center for Systems and Control

Analyzing point spread functions for multiple emitter fitting by posterior density estimation

Master of Science Thesis

For the degree of Master of Science in Systems and Control at Delft University of Technology

Raymond van Dijk

July 31, 2022

Faculty of Mechanical, Maritime and Materials Engineering (3mE) · Delft University of Technology

Copyright © Delft Center for Systems and Control (DCSC)
All rights reserved.

Delft University of Technology
Department of
Delft Center for Systems and Control (dcsc)

The undersigned hereby certify that they have read and recommend to the Faculty of Mechanical, Maritime and Materials Engineering (3mE) for acceptance a thesis entitled

Analyzing point spread functions for multiple emitter fitting by posterior density estimation

by

Raymond van Dijk

in partial fulfillment of the requirements for the degree of

Master of Science Systems and Control

Dated: July 31, 2022

Supervisor(s):

Dr. ir. C. S. Smith

Jelmer Cnossen

Reader(s):

Prof. dr. ir. B. de Schutter

Dr. K. Grußmayer

Dr. ir. S. L. Caneva

Abstract

The microscope is an essential tool for biologists. Since the late 16th century, it has given researchers a better understanding of cell processes and greatly advanced healthcare. In this century, Single molecule localization microscopy (SMLM) has revolutionized optical microscopy by breaking the optical diffraction limit. Sparsely activating emitters in a sample labeled with fluorophores, the object can be reconstructed by estimating their positions using the system point spread function (PSF). These localization algorithms are the state of the art in optical imaging, using unbiased estimators to reach the theoretical minimum uncertainty, or Cramér-Rao lower bound (CRLB).

While SMLM works well when emitters are sparsely activated, overlap of the emitter images is inevitable for thick or densely labeled samples. When SMLM is used on such images, the estimates become biased and the algorithm cannot find the correct number of emitters. Most techniques also make a deterministic estimate and are incapable of representing the uncertainty of estimates for dense samples.

A three-dimensional, Bayesian multiple emitter fitting algorithm is constructed using reversible jump Markov chain Monte Carlo (RJMCMC). While following the structure of Bayesian multiple-emitter fitting (BAMF) [1], novel RJMCMC moves are designed to sample the parameters. The algorithm also jumps through models, estimating the number of emitters. It asymptotically samples from the posterior, revealing uncertainties in three-dimensional imaging that other techniques are incapable of imaging.

The algorithm was tested with astigmatic and biplane imaging. It has proven capable of consistently finding the correct model when a prior on emitter intensity is used. When separating two emitters, posterior density reconstruction revealed non-Gaussian emitter position uncertainties. Upon further investigation, the posterior density was found to be multimodal, with both modes representative of the data and indistinguishable in terms of likelihood. This shows the algorithm can quantify three-dimensional PSF degeneracy [2] and can become a vital tool for researchers to analyze their imaging setup. We also expect it to be especially effective when combined with modulation-enhanced localization microscopy (meLM) [3] techniques.

Table of Contents

List of Figures

Preface

When I began this journey, I could not yet envision the challenge that was ahead of me. Writing this thesis tested not only my academic ability, but strength of character and independence. I can only be grateful to have had this opportunity to develop myself.

Working in the optics lab has been a special experience. Despite the woes of the global pandemic, the scientific staff have created a welcoming and encouraging environment for students to conduct their research in. Joined by a passionate group of fellow students, the lab was stimulating both in an academic and a social sense.

I would like to take this opportunity to express my gratitude. Firstly, to my supervisors, Carlas Smith and Jelmer Cnossen, for guiding my work and providing valuable insights. You have opened my eyes on so many occasions, helping me tackle even the most complex problems. Secondly, towards Dylan Kalisvaart, who also provided key insights and helped write the manuscript. Your selflessness and dedication is inspiring. Thirdly, my thanks go out to the students I had the pleasure of getting to know during my thesis. Let it be known that you are a wonderful group of ambitious and helpful people. Interacting with one another and getting to watch all of these projects develop was delightful. Lastly, I wish to thank my family and friends. You have kept me going through these challenging times and encouraged me to give it my absolute best.

Delft, University of Technology
July 31, 2022

Raymond van Dijk

"Wonder is the first of all the passions."
— *Rene Descartes*

Chapter 1

Introduction

Since the invention of the microscope, scientists have used it to better their understanding of biological processes within organisms. As lens makers improved their skills, more powerful lenses with less flaws became available and microscopes improved in resolution. Optical microscopy however has a fundamental limit on resolution, known as Abbe's diffraction limit:

$$d = \frac{\lambda}{2NA} \tag{1-1}$$

with d the resolution, λ the wavelength, and NA the numerical aperture of the lens. Although optical microscopy has been surpassed in resolution by electron and atomic force microscopy, the ability to do live cell imaging still makes it a crucial tool for biologists. Scientists therefore throughout the years developed a series of techniques to break the diffraction limit.

This chapter will give an introduction into fluorescence microscopy. Section 1-1 discusses the optical setup used for fluorescence microscopy. Section 1-2 introduces the image formation model and the optical impulse response, also known as the point spread function (PSF). In section 1-3, resolution is discussed. Section 1-4 explains the principles behind breaking the diffraction limit through localization microscopy. Finally, section 1-5 outlines the research topic for this thesis.

1-1 Fluorescence microscopy setup

To understand the possibilities and limitations of super-resolution techniques using fluorescence microscopy, the setup is briefly discussed. Light in a fluorescence microscope takes the following path:

- an excitation beam is filtered and focused onto the sample
- the excitation light activates the fluorescent label, which in turn starts emitting light
- the emission light is filtered and focused onto the camera

Understanding the sample helps identify which properties are desirable to image, while analyzing the path of the light within the sample will give an insight into the factors that can limit the resolution. The path of both the excitation and the emission light can then be studied to help measure the desired properties and find methods to maximise resolution. Figure 1-1 shows a simple diagram of a fluorescence microscopy setup.

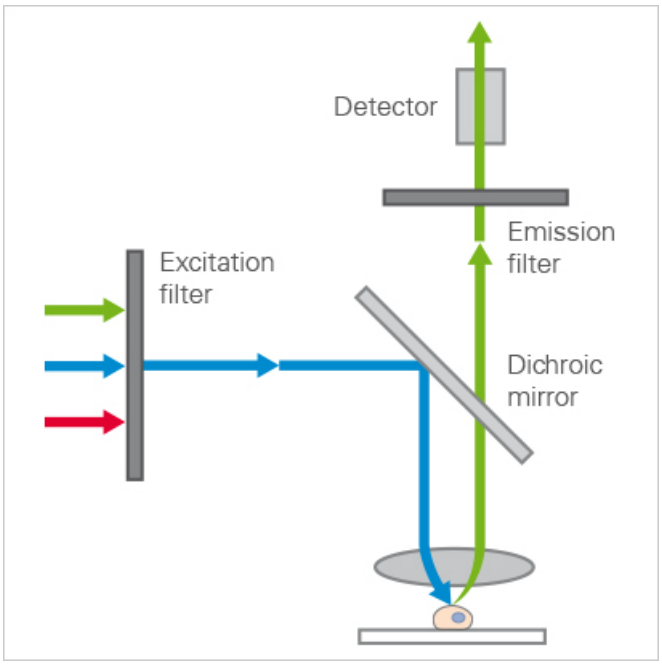


Figure 1-1: An example of a simple fluorescence microscopy setup. A light source is filtered such that it matches the absorption spectrum of the fluorescent labels in the sample. It then meets a dichroic mirror, which joins the excitation path with the emission path, illuminating the sample. The emission light from the sample is shifted in wavelength, allowing it to pass the dichroic mirror. After passing another filter that matches the emission spectrum, it is finally focused onto the detector to record the image. Source: https://Ibidi.com

1-1-1 The sample

To make the samples fluoresce they are labeled with a fluorescent stain or protein. When the sample is illuminated and a photon that matches the absorption spectrum of the label is incident on the fluorescent molecule, it is absorbed, transitioning the molecule from its ground state S_0 to an excited state S_1, as depicted in the Jablonski diagram of Figure 1-2. Through non-radiative decay, some of the absorbed energy is dissipated in a process called vibrational relaxation, lowering the vibrational level of the excited state. Then, through spontaneous

emission, the molecule emits a photon as it is returned to its ground state. The molecule can also be brought to its ground state by an incident photon in a process called stimulated emission. Note that decay to the ground state can also be achieved solely through non-radiative decay, meaning the ratio of emission to absorption photons is < 1. This ratio is named the quantum yield of the fluorophore. Due to the energy dissipated through non-radiative decay, the emitted photon carries less energy than the absorbed photon. This results in a Stokes shift, meaning the emitted photon has a longer wavelength than the absorbed photon. An example of fluorophore absorption and emission spectra is shown in Figure 1-2.

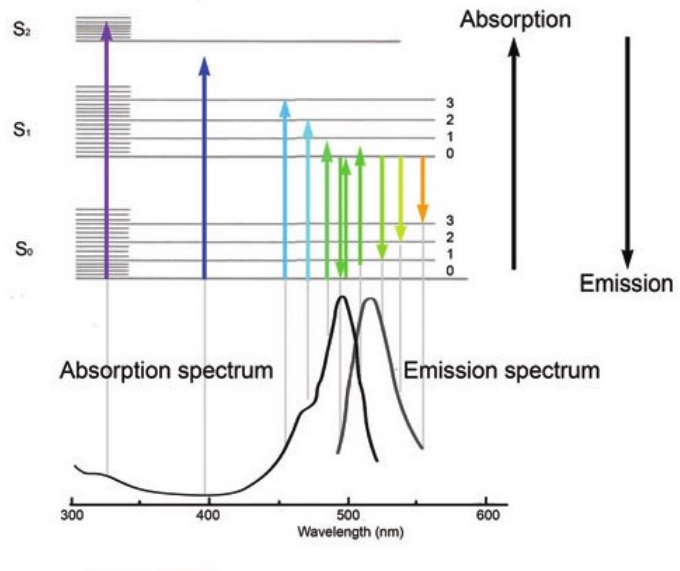


Figure 1-2: Top: Jablonski diagram for photon absorption and emission by a fluorescent molecule. When illuminated with light within the absorption spectrum, the fluorophore absorbs the photon and transitions from its ground state, S_0, to an excited state, S_1. while in state S_1, vibrational losses dissipate some of the energy. Finally, the fluorophore returns from S_1 to S_0 through either emission of a photon or dispersion of heat. Bottom: corresponding absorption and emission spectra of fluorescein isothiocyanate (FITC). Source: image adapted from Lichtman and Conchello, Nature methods 2005

Fluorescent labeling introduces both a series of useful properties as well as constraints to the system. First and foremost, the switching between on and off states allows for recording of frames in which sparse subsets of molecules are emitting. This enables us to achieve resolution beyond the diffraction limit through Single molecule localization microscopy (SMLM), which will be discussed in section 1-4. Through the difference in emission and absorption spectra, the illumination source can be filtered out of the image, as shown in Figure 1-2. This results in a high contrast image with a good Signal-to-background ratio (SBR). Furthermore, the ability to label specific molecules and do in vivo imaging allows researchers to closely study the processes within cells through particle tracking.

A downside however is that the fluorescent label, or emitter, can only activate a finite amount of times, after which photobleaching occurs and the label is turned off permanently. This can happen within minutes, depending on the illumination, limiting the acquisition time and

illumination intensity. Additionally, when working with thick samples, out of focus emitters will activate, resulting in a lot of background and more difficulty in collecting frames with sparsely activated emitters.

1-1-2 Excitation path

For this work, we will focus on epifluorescence, which is simple yet effective. To activate fluorescence in the sample, an illumination source with a wavelength corresponding to the absorption spectrum of the label is required. This leads to setups using either a laser or a white light source in combination with a filter. After filtering, the excitation source can be brought into the sample through various methods. Using a dichroic mirror, which is transparent to light in its passband while reflecting other wavelengths, the excitation path is joined with the emission path, as shown in Figure 1-1. This will send light through the entire sample which may cause some issues when imaging thicker samples, as discussed in the previous subsection.

1-1-3 Emission path

The emission path traces from the sample through the sample coverslip and objective onto the dichroic mirror, which is transparent to the emission and is finally relayed to the camera with a 4f system. As extra precaution an emission filter is added to remove any stray excitation light. As the emitters blink independently, the resulting light is incoherent. This means no interference should occur and the image can be considered a sum of emissions from a collection of point sources. Neither is the light polarized, due to variations in orientation of the emitters.

1-2 Image formation model

An image is formed by a light source illuminating the object and the object interacting (reflecting, absorbing, fluorescing) with this light. For incoherent imaging, each point on the object can be treated as a point source and the image can be constructed by superposition of the system response to each source.

Propagation of a point source through the imaging system can be described by its point spread function. Assuming uniform illumination and the PSF being shift-invariant, the image can now be constructed:

$$I(x_i, y_i) = \iint O(u, v)h(x_o - u, y_o - v)dudv \tag{1-2}$$

with $O(x, y)$ the object, $h(x, y)$ the PSF, x_o and y_o the coordinates in object plane, and x_i and y_i the coordinates in the image plane. The object and image planes are related through magnification M as $(x_i, y_i) = (x_o/M, y_o/M)$. Note that the formula for image formation is a convolution integral, meaning the object is blurred by the PSF to form the image. Using the Fourier transform results in:

$$\hat{I}(f_{x_i}, f_{y_i}) = \hat{O}(f_{x_o}, f_{y_o})\hat{h}(f_{x_o}, f_{y_o}) \tag{1-3}$$

with f the spatial frequencies in object or image plane, $\hat{O} = \mathcal{F}\{O\}$ the Fourier transform of the object, $\hat{I}$ that of the image, and $\hat{h}$ that of the PSF, which is the system optical transfer function (OTF). The image can thus also be interpreted as the result of the object being filtered by the OTF of the system.

1-2-1 Point spread functions

Several PSF models can be applied to fluorescence microscopy. In the paraxial regime, that is for systems with a relatively small NA, the PSF for the incoherent system can be modeled as the square of the Airy Disk:

$$h(x, y) = \left[\frac{2J_1\left(2\pi NA\sqrt{x^2 + y^2}/\lambda\right)}{2\pi NA\sqrt{x^2 + y^2}/\lambda} \right]^2 \tag{1-4}$$

with J_1 the order 1 Bessel function of the first kind and λ the emission wavelength. Note that this PSF model assumes ideal imaging, so no aberrations are present. However, the microscopes used often have high NA objectives as system resolution scales inversely with NA. For these powerful lenses a more accurate model would be the vectorial PSF [4]:

$$H(\vec{r}) = \frac{N}{3} \sum_{l=x,y} \sum_{j=x,y,z} |w_{lj}(\vec{r})|^2 + \frac{b}{a^2} \tag{1-5}$$

where $\vec{r}$ is the coordinate vector, N the number of pixels, $w_{lj}(\vec{r})$ the electric field components, b the background photons, and a the pixel size. The electric field components are given by:

$$w_{lj}(\vec{r}) = \frac{1}{w_n} \int_{|\vec{\rho}| \leq 1} d^2\rho A(\vec{\rho}) \exp\left(\frac{2\pi i W(\vec{\rho})}{\lambda}\right) q_{lj}(\vec{\rho}) \exp(-i\vec{k}(\vec{\rho}) \cdot \vec{r}) \tag{1-6}$$

where $\vec{\rho}$ are the pupil coordinates, w_n is a normalization factor, $A(\vec{\rho})$ the amplitude, q_{lj} polarization vector components, $W(\vec{\rho})$ the aberration function, and $k(\vec{\rho})$ the wavevector. This model is significantly more complex, but accounts for both aberration and polarization effects.

For 2D imaging using fluorescent molecules, the PSF can be approximated with a Gaussian function:

$$h(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(- \frac{x^2 + y^2}{2\sigma^2} \right) \tag{1-7}$$

with $\sigma = \lambda/(4NA)$ the standard deviation. This approximation is accurate when the fluorescent molecules can be modelled as freely rotating dipoles, for instance in fluid solution. The molecules take a vastly different orientation in about 50-100 ps, while emitting for 1-10 ms [5]. The result is that during capture, the vectorial PSF is averaged over a collection of orientations, giving a PSF that closely resembles a Gaussian with an additional shoulder, as in figure 1-5.

3D point spread functions

The PSFs up to this point assume the point source is in focal plane, which is not always the case. When studying thick samples, it can be helpful to modify the PSF to encode the axial position into its shape. Examples of such PSFs are the astigmatic [6] and the tetrapod [7, 8] PSF.

The astigmatic PSF can be formed by using a cylindrical lens. This splits the focal plane into two focal lines, one above and one below the original focus, which focus the image along perpendicular directions, as in Figure 1-3. The focus is now defined to be in the middle of these two planes, where the circle of least confusion occurs. The resulting PSF is well approximated by a 2D Gaussian in focal plane which elongates into an ellipse when moving out of focus:

$$h(x,y) = \frac{1}{2\pi\sigma_x(z)\sigma_y(z)} \exp\left(-\frac{x^2}{2\sigma_x(z)^2} - \frac{y^2}{2\sigma_y(z)^2} \right) \tag{1-9}$$

$$\sigma_x(z) = \sigma_{0x}\sqrt{1 + \frac{(-z)^2}{d^2} + A_x\frac{(-z)^3}{d^2} + B_x\frac{(-z)^4}{d^2}} \tag{1-11}$$

$$\sigma_y(z) = \sigma_{0y}\sqrt{1 + \frac{z^2}{d^2} + A_y\frac{z^3}{d^2} + B_y\frac{z^4}{d^2}} \tag{1-12}$$

with d the depth of focus depending on the objective, σ_{0x}, σ_{0y} the x and y direction width in focus, and A and B some constants depending on the optical system. This assumes $z = 0$ to be at the circle of least confusion and that the x direction is focused above the focal plane.

The tetrapod PSF is formed by modulating the wavefront, using for example a spatial light modulator (SLM). As the tetrapod PSF is moved out of focus, it spreads out into two main lobes along perpendicular directions depending on the depth, as in Figure 1-4. The tetrapod PSF can be optimized for different ranges of depths, as demonstrated by Shechtman et al. [8].

Fitting experimental point spread functions

To account for aberrations inherent to the optical system or calibrate for engineered PSFs, the PSF model can also be fitted to experimental data. An efficient method to do this is by

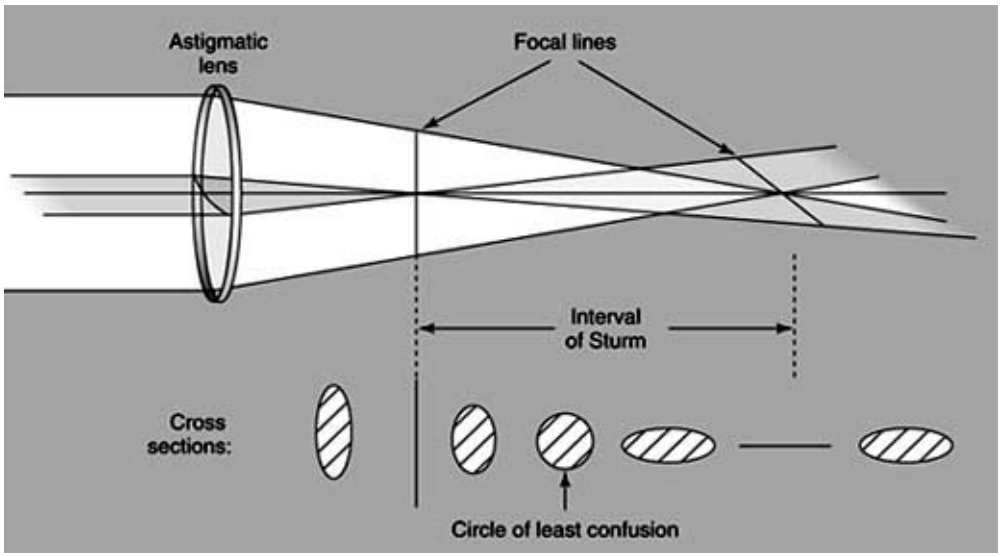


Figure 1-3: Demonstration of astigmatism forming two focal lines. The circle of least confusion, which results in the best image quality, is in the middle of both focal lines. Source: image adapted from http://optometryzone.com

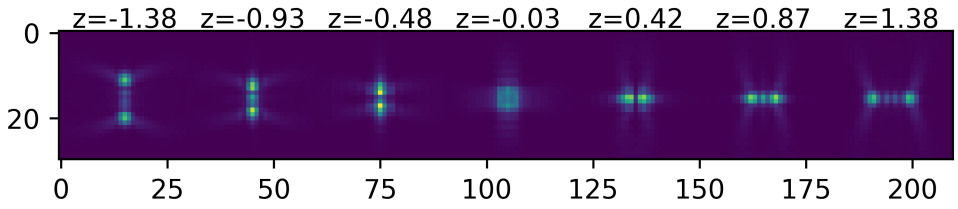


Figure 1-4: Visualisation of the tetrapod PSF at different depths, optimized for a range of $[-1.4 \quad 1.4]$ μm.

using cubic splines [2]. Splines approximate functions by constructing a series of piecewise polynomials that fit the data. The polynomials are constrained at the ends of their intervals such that $f_i(t_{i+1}) = f_{i+1}(t_i)$ and $df_i(t_{i+1})/dx = df_{i+1}(t_i)/dx$. A 1D cubic spline can be described as follows:

$$f_i(x) = a_i \frac{x - t_i}{\delta t}^3 + b_i \frac{x - t_i}{\delta t}^2 + c_i \frac{x - t_i}{\delta t} + d_i \tag{1-13}$$

$$t_i \leq x \leq t_{i+1} \tag{1-14}$$

$$\delta t = t_{i+1} - t_i \tag{1-15}$$

with $f_i(x)$ approximating the function on the interval $[t_i, t_{i+1}]$. Varying the interval size can increase the accuracy of the fit, but also increases the number of coefficients of the model. An interval with a length of half the image pixel size was shown to be appropriate for accurately modeling PSFs while keeping the amount of coefficients moderate. Extending cubic splines to 3D to capture a 3D PSF results in:

$$f_{i,j,k}(x,y,z) = \sum_{m=0}^{3}\sum_{n=0}^{3}\sum_{o=0}^{3} a_{i,j,k,m,n,o}\left(\frac{x-t_i}{\delta t}\right)^m\left(\frac{y-t_j}{\delta t}\right)^n\left(\frac{z-u_k}{\delta u}\right)^o \tag{1-16}$$

$$t_j \leq y \leq t_{j+1} \tag{1-17}$$

$$u_k \leq z \leq u_{k+1} \tag{1-18}$$

$$\delta u = u_{k+1} - u_k \tag{1-19}$$

with $[u_k, u_{k+1}]$ the interval in the axial direction. The coefficients $a_{i,j,k,m,n,o}$ can now be determined by fitting to either simulated or experimental 3D PSF data.

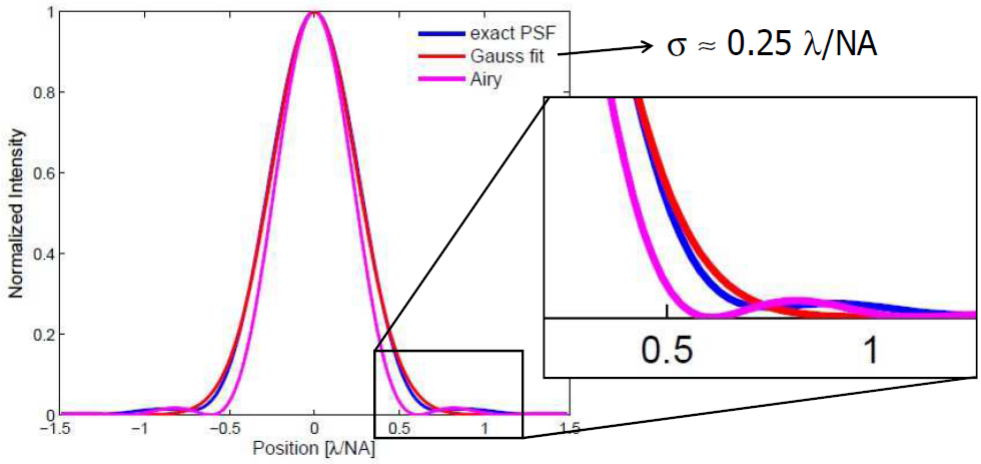


Figure 1-5: Plot of the exact PSF, the Gaussian approximation, and the PSF for the paraxial regime. The exact PSF resembles a Gaussian with a shoulder and is well approximated with a Gaussian distribution using $\sigma = \lambda/(4NA)$. Source: image adapted from lecture slides of TU Delft course AP3121: Imaging systems, S. Stallinga.

1-3 Resolution

There are several methods to quantify resolution. As in the beginning of this chapter, Abbe's diffraction limit defines the smallest resolvable length as $0.5\lambda/NA$. Rayleigh's resolution criterion however states that two point sources are resolved when the maximum of the image of one point source coincides with the minimum of the image of the other source. Applying this to the Airy pattern, the sources are resolved when the maximum lies in the first dark ring of the pattern, resulting in a resolution of $0.61\lambda/NA$. Both give a theoretical resolution limit for the system based on the numerical aperture and emission wavelength, or by evaluating the PSF. The resolution can also be verified experimentally. The full width half maximum (FWHM), defined as the width at which the signal is greater than half of its maximum, can be used to get an idea of the width of the PSF. Test targets have also been devised to verify the resolution of the system. They often include patterns of increasing spatial frequency in various orientations, such as parallel lines of decreasing thickness. One such target is shown in figure 1-7.

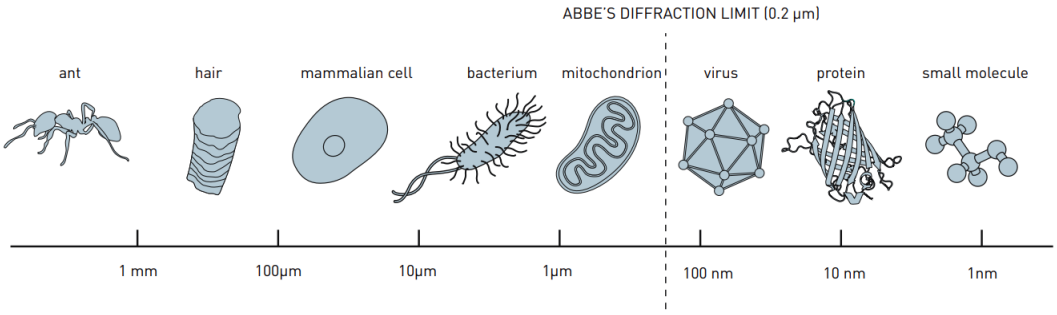


Figure 1-6: Visualisation of the scale of microscopic elements. Conventional microscopy is not capable of imaging viruses or smaller particles, as it can resolve elements no smaller than 200 nm. Source: NobelPrize.org, ©Johan Jarnestad, The Royal Swedish Academy of Sciences.

As an example of the resolution of a fluorescence microscope, suppose we are imaging with a system that has a numerical aperture of 1.3 and our sample is labeled with Cyanine-5 (Cy5) [9], which has a peak in its emission spectrum at 670 nm. Following Abbe's diffraction limit, this system can resolve two emitters if the distance between them is no less than 258 nm. Following Figure 1-6, if we wish to image the structures that make up viruses or proteins, we need a technique that can go well beyond the diffraction limit, which brings us to super-resolution microscopy.

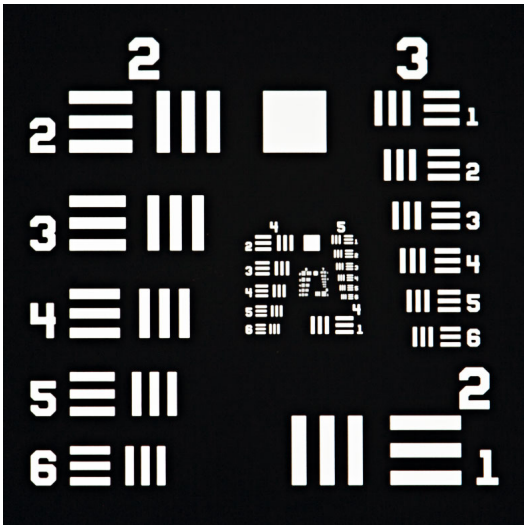


Figure 1-7: 1951 USAF resolution test target, as defined by the U.S. Air Force MIL-STD-150A standard. Source: Thorlabs.com

1-4 Super-resolution microscopy

This section evaluates several methods to go beyond the resolution limit under various conditions. Though the the excitation path allows for interesting means of imaging beyond the diffraction limit [10, 11], in this work the focus will be on localization microscopy.

1-4-1 Single molecule localization microscopy

In 1995, Eric Betzig proposed a method to break the diffraction limit [12]. Normally when imaging an object the PSFs of the various point sources overlap, making it impossible to distinguish their individual positions. However, if one were able to separate these point sources along an orthogonal axis, their parameters could be estimated given the PSF. Fluorescence microscopy allows us to separate emitters in time through sparse activation, letting us acquire frames where the subset of active emitters has little to no overlap. By capturing a series of frames, each with different active emitters, it becomes possible to localize all the labelled molecules within the sample. The image can then be reconstructed at a resolution depending solely on the uncertainty of the estimated emitter positions, going well beyond the diffraction limit. Over a decade later, two fluorescence microscopy techniques emerged that break the diffraction limit through Single molecule localization microscopy (SMLM).

Photo-activated localization microscopy Photo-activated localization microscopy (PALM) [13] is an SMLM method that works by activating sparse subsets of emitters using two lasers. An excitation laser is used to continuously illuminate the sample, bringing the emitters to their excited state. Meanwhile, an activation laser is pulsed at a wavelength below the absorption peak. This activation laser is used to increase the density of active emitters up to a point where they were still resolvable. This increase in density of active emitters ensured super-resolution could be achieved faster. Following the activation pulse, the camera captures frames as the excitation laser slowly bleaches part of the emitters. As all emitters are bleached, the activation laser pulses again to activate another subset of emitters, repeating until all emitters are bleached. The emitters in the collected frames are then localized using the microscope PSF and a super-resolution image is reconstructed from the emitter positions and their uncertainties.

Stochastic optical reconstruction microscopy Stochastic optical reconstruction microscopy (STORM) [9] uses both a red and a green laser to switch the emitters from dark states to active states. It uses the properties of the photoswitchable Cy5 dye, which can both be excited and brought to a dark state by the red laser. First, a pulse from the red laser switches all emitters to their dark state. Then, for an imaging cycle, the green laser sends a pulse to activate a subset of the emitters. Then, while illuminated by the red laser, these emitters continue to fluoresce until switched off. Later variations of STORM use different dyes and work with just a single excitation laser.

STORM, PALM, and other derivative SMLM techniques can achieve resolutions in the order of tens of nanometers, reaching the CRLB for localization microscopy [14]. Yet they are also limited by the photostability of the dyes and the required laser power. They also remain localization methods for single molecules, struggling in applications where sparse activation is not always achieved.

1-4-2 High density localization microscopy

Localization techniques for densely labeled samples can obtain more information in shorter time intervals or image densely labeled samples that SMLM can not. As opposed to SMLM,

where generally a region of interest (ROI) can be found with only one active emitter, techniques for densely labeled samples also need to be able to identify the number of active emitters in the ROI. Compressed sensing as used in compressed sensing STORM (CS-STORM) has shown to be effective for fitting densely labeled samples. The Fast Localization algorithm based on Continuous-spatial formulation (FALCON) [15] extends on this technique, turning a discrete position estimate from compressed sensing into a continuous domain estimate with a Taylor series approximation. Sun et al. [16] extend on FALCON by applying variational inference to localization. The variational expectation-maximization (vEM) is initialized by FALCON to avoid entrapment in local minima, further refining the estimated positions.

Another method to achieve super-resolution, though not strictly localization microscopy, is super-resolution optical fluctuation imaging (SOFI) [17]. It is based on temporal correlation and stochastic fluctuations of the emitters. SOFI images the brightness and correlation of the object, giving a resolution improvement of $\sqrt{2}$. Use of higher order cumulant functions can in theory improve the resolution endlessly, but in practice the contrast of the reconstruction becomes a problem. Super-resolution through radial fluctuations (SRRF) [18] also uses temporal correlation, evaluating the cumulants of the radial symmetry of the image instead of the intensity. SRRF does not suffer from the contrast issues that SOFI has, as radial symmetry is a normalized parameter for all emitters unlike brightness. As a result, SRRF can resolve high density images where SMLM methods may fail and at high densities is capable of achieving super-resolution in 100 frames, enabling live-cell super-resolution.

Deep neural networks were also shown to be highly effective in dense samples. Deep-STORM [19] uses a deep convolutional encoder-decoder network, encoding individual frames into a feature respresentation and then decoding the features directly into a super-resolution reconstruction. It can be trained on simulated or experimental data and reconstructs dense frames where SMLM normally fails. DeepSTORM3D [20] and DECODE [21] extend these principles to 3D, training deep neural networks on simulated or experimental data of thick samples. DeepSTORM3D uses the tetrapod PSF, while also using the network to find an optimal PSF for multiple emitter fitting. DECODE combines emitter detection and localization, while also providing the detection probabilities and localization uncertainties.

Finally, a Bayesian approach may solve both the problem of finding a model and estimating the parameters simultaneously. Bayesian multiple-emitter fitting (BAMF) [1] uses reversible jump Markov chain Monte Carlo (RJMCMC) to approximate the posterior probability distribution of the parameters as well as the model. It runs SMLM code [14] to find prior information for emitter intensity combining with user provided priors on emitter count and uniform priors on position to not introduce bias in localization. The priors and likelihoods are then used to estimate the posterior using the RJMCMC algorithm and returning a maximum a posteriori (MAP) estimate of the emitter count.

BAMF proposes a series of jumps for the RJMCMC chain, consisting of three types. The first proposes a jump within the model, optimizing the parameters without changing the number of emitters. The second allows RJMCMC to move to different parameter spaces by changing the number of emitters. The final type is used to convert signal emitters to background emitters and vice versa. All jumps are shown in figure 1-8, explaining the BAMF algorithm. After RJMCMC settles on an appropriate model and posterior, it is used to initialize Markov chain

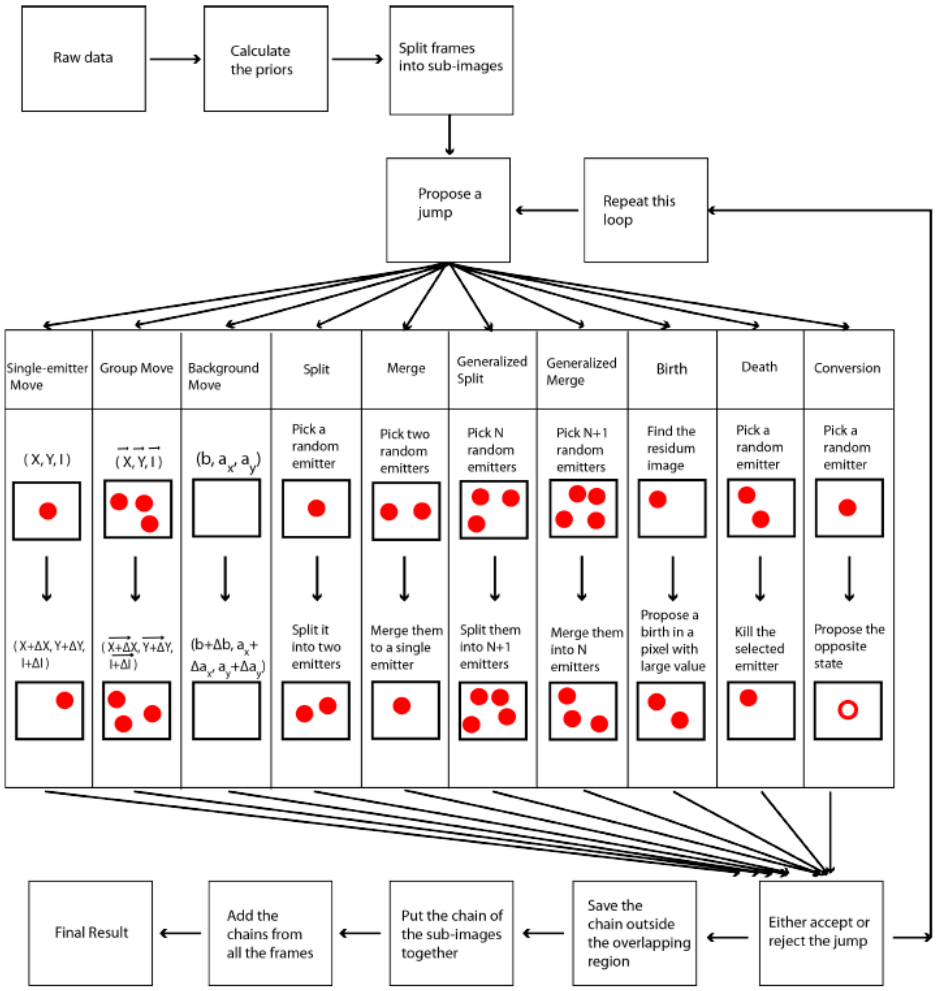


Figure 1-8: Outline of the BAMF algorithm. First, the priors are calculated from the raw data. The frames are then split into sub-images, and RJMCMC is applied to approximate the posterior distribution. The ten jump proposals for the Markov chain are shown: three for moves within a given model, 3 pairs of jumps between parameter spaces, and one for conversion of signal and background emitters. Source: Fazel et al., [1]

Monte Carlo (MCMC) estimation based on the maximum a posteriori model of number of emitters (MAPN). No longer having to jump between parameter spaces, MCMC can do more within model jumps to better approximate the posterior. This MCMC chain then returns the posteriors from each frame, which are then averaged to form the reconstruction.

1-5 Research topic

Despite the plethora of advanced super-resolution techniques available, it has become evident that localization microscopy still struggles when too many emitters are active in a frame. This is of particular importance for 3D imaging, where out of focus emitters are also activated and cause a lot of overlap in the frame. Work on multiple emitter fitting also allows researchers to pack more information into each frame, allowing for shorter acquisition times and more favorable illumination intensities. Though several methods exist to do multiple emitter fitting in 2D, the research on 3D multiple emitter fitting is limited.

To contribute to multiple emitter fitting in 3D, the goal of this work will be to extend the principles of BAMF [1] to a 3D Bayesian multiple emitter fitting algorithm. The Bayesian approach to model identification was shown to be accurate, relying mostly on the prior on intensity and emitter count. By focusing on the astigmatic and tetrapod PSFs, the resulting algorithm should work for even relatively simple setups and requires no learning. The astigmatic PSF has a good chance of converging with 3D RJMCMC localization, as RJMCMC was designed to fit Gaussian mixture model (GMM)s and it is well approximated by a 2D Gaussian. The tetrapod PSF may provide more of a challenge, while successful implementation should result in better axial resolution. The challenges will consist of how to define appropriate priors and getting the algorithm to converge.

The remainder of the thesis is structured as follows. Chapter two will work out some of the math required for localization microscopy, explaining Maximum Likelihood Estimation (MLE), Cramér-Rao lower bound (CRLB), MAP estimation, and posterior probability density estimation. Chapter three explores MCMC methods for asymptotically approximating the posterior density, as well as giving an introduction to RJMCMC for model estimation. Chapter four will present the main body of the thesis in a paper format, evaluating the effectiveness of 3D localization using RJMCMC. Finally, chapter five holds the conclusion to the thesis.

Chapter 2

Parameter estimation

As a precursor to localization microscopy, this chapter will elaborate some of the estimators used to reconstruct the object. To estimate the position of the emitter, a model for the detected photon count of the camera pixels is required. Using a high gain camera such as an electron multiplying charge-coupled device (EMCCD), the readout noise will be negligible. The pixel output can then be described by a Poisson process:

$$P(n_k|\mu_k(\theta))_k = \frac{\mu_k(\theta)^{n_k}}{n_k!}e^{-\mu_k(\theta)} \tag{2-1}$$

With $P(n_k|\mu_k(\theta))_k$ the probability of the pixel value n_k given expected value $\mu_k(\theta)$ in the k^{th} pixel. The parameter vector θ contains the set of parameters up for estimation, such as emitter position, emitter intensity, and background photon count.

2-1 Maximum likelihood Estimation

Using this stochastic model, Maximum Likelihood Estimation (MLE) can be used to find an unbiased, mimimum variance estimate of the emitter parameters [14]. The expected value $\mu_k(\theta)$ follows from the image formation model, assuming only a single emitter contributes to the measurement:

$$\mu_k(\theta) = \theta_I \int_{A_k} h(u-\theta_x, v-\theta_y, \theta_z)dudv + \theta_b \tag{2-2}$$

with θ_I the emitter intensity, A_k the pixel area, $(\theta_x, \theta_y, \theta_z)$ the emitter position, and θ_b the number of background photons. The parameter vector becomes $\theta = \begin{bmatrix}\theta_x & \theta_y & \theta_z & \theta_I & \theta_b\end{bmatrix}^T$.

For a region of interest (ROI), the likelihood of a collection of pixel readouts then becomes the product of the likelihood for each pixel:

$$L(D|\theta) = P\Big(\begin{bmatrix} n_1 & n_2 & ... & n_k \end{bmatrix} \Big| \mu(\theta) \Big) = \prod_{k=1}^{N} \frac{\mu(\theta)^{n_k}}{n_k!} e^{-\mu(\theta)} \tag{2-3}$$

with N the number of pixels in the region of interest and D the data vector collecting all the pixel readouts. The parameter vector can now be estimated by finding the maximum of the likelihood function. As the natural logarithm is monotonically increasing, $\arg\max_x \ln f(x) = \arg\max_x f(x)$ and the log likelihood can be used as:

$$\ln\left(L(D|\theta)\right) = \sum_{k=1}^{N} n_k \ln\left(\mu_k(\theta)\right) - \mu_k(\theta) - \ln(n_k!) \tag{2-4}$$

Finding the MLE for the parameters now becomes a root finding problem for the partial derivatives of the log likelihood:

$$\frac{\partial \ln(L)}{\partial \theta_i} = \frac{\partial \ln(L)}{\partial \mu_k} \frac{\partial \mu_k}{\partial \theta_i} = \sum_{k=1}^{N} \left(\frac{n_k}{\mu_k} - 1 \right) \frac{\partial \mu_k}{\partial \theta_i} \tag{2-5}$$

which can be solved iteratively. Given an initial estimate sufficiently close to the true parameter values, the Newton-Raphson method will converge to the estimate that maximizes the likelihood. Note that up to (2-5), no assumptions about the point spread function (PSF) were made, thus the derivation is applicable for any PSF model. Given a 3D PSF model, MLE is also used to find the axial position [22] despite this direction being perpendicular to the imaging plane.

2-2 Variance of the estimate: Cramér-Rao lower bound

To characterise the localization precision, the variance of the parameter estimates is used. The Cramér-Rao lower bound (CRLB) defines the minimum variance of the estimate as the inverse of the Fisher information matrix, which is given as:

$$I_{ij}(\theta) = E\left[\frac{\partial \ln(L)}{\partial \theta_i} \frac{\partial \ln(L)}{\partial \theta_j} \right] \tag{2-6}$$

$$\tag{2-7}$$

Substituting (2-5) gives:

$$I_{ij}(\theta) = E\left[\sum_{k=1}^{N}\left(\frac{(n_k-\mu_k)}{\mu_k}\right)^2 \frac{\partial\mu_k}{\partial\theta_i}\frac{\partial\mu_k}{\partial\theta_j}\right] \tag{2-8}$$

$$I_{ij}(\theta) = E\left[\sum_{k=1}^{N}(n_k-\mu_k)^2\frac{1}{\mu_k^2}\frac{\partial\mu_k}{\partial\theta_i}\frac{\partial\mu_k}{\partial\theta_j}\right] \tag{2-9}$$

with $\mu_k(\theta)$ as μ_k, for simplicity. Now substituting the variance of the Poisson process $E[(n_k-\mu_k)^2] = \mu_k$ results in:

$$I_{ij}(\theta) = \sum_{k=1}^{N}\frac{1}{\mu_k}\frac{\partial\mu_k}{\partial\theta_i}\frac{\partial\mu_k}{\partial\theta_j} \tag{2-10}$$

The variance of the parameters is then found as the inverse at the optimum, $I_{ij}(\hat{\theta})$. The CRLB can now be calculated for various PSF models, allowing comparison of their theoretical minimum uncertainty. It can also be used as validation for localization performance in experiments.

Example: CRLB for 1D Gaussian PSF To demonstrate the CRLB calculations for the minimum variance, a 1D example is given. Assuming negligible pixel size, direct measurement of photon count, and no background, the model is as follows:

$$\mu_k(\theta_x,\theta_I) = \frac{\theta_I}{\sqrt{2\pi}}\exp\left(-\frac{(x-\theta_x)^2}{2\sigma^2}\right) \tag{2-11}$$

Equation (2-5) can then be simplified under the given assumptions as:

$$\sum_{k=1}^{N}\mu_k = \theta_I \tag{2-12}$$

$$\frac{\partial\ln(L)}{\partial\theta_i} = \sum_{k=1}^{N}\frac{n_k}{\mu_k}\frac{\partial\mu_k}{\partial\theta_i} - \frac{\partial\sum_{k=1}^{N}\mu_k}{\partial\theta_i} = \sum_{k=1}^{N}\frac{n_k}{\mu_k}\frac{\partial\mu_k}{\partial\theta_i} - \frac{\partial\theta_I}{\partial\theta_i} \tag{2-13}$$

The first and second derivative with respect to the emitter x position become:

$$\frac{\partial\ln(L)}{\partial\theta_x} = \sum_{k=1}^{N} -n_k\frac{\theta_x - x}{\sigma^2} \tag{2-14}$$

$$\frac{\partial^2\ln(L)}{\partial\theta_x^2} = \sum_{k=1}^{N} -\frac{n_k}{\sigma^2} \tag{2-15}$$

And thus the CRLB for the standard deviation of the emitter position follows as:

$$\sigma_{\theta_x,CRLB} = \left[-\frac{\partial^2 \ln(L)}{\partial \theta_x^2}\right]^{-\frac{1}{2}} = \frac{\sigma}{\sqrt{\theta_I}} \tag{2-16}$$

which shows that the theoretical minimum for standard deviation of the position estimates scales inversely with the root of the photon count.

2-3 Maximum a posteriori probability estimation

When prior information of the parameters is known, the maximum a posteriori (MAP) estimate can be used to incorporate this information. The MAP maximises the probability of the parameters conditional to the data:

$$P(\theta|D) = \frac{P(\theta)P(D|\theta)}{P(D)} \tag{2-17}$$

which follows from Bayes' rule as the prior $P(\theta)$ times the likelihood $P(D|\theta)$ over the evidence $P(D)$. The MAP estimate then follows from solving:

$$\hat{\theta}_{MAP} : \frac{\partial}{\partial \theta}\left[\frac{P(\theta)P(D|\theta)}{P(D)}\right] = \frac{\partial}{\partial \theta}\left[P(\theta)P(D|\theta)\right] = 0 \tag{2-18}$$

Note that $P(D)$ is not necessary for calculating the MAP estimate as it does not depend on θ. When the prior is uniformly distributed, the MAP and MLE estimates are identical.

2-4 Posterior probability density estimation

Bayesian inference, as opposed to MLE and MAP estimators, aims to approximate the conditional probability density function $P(\theta|D)$, instead of trying to find a deterministic estimate for the parameters. This density function can be found by expanding the evidence from (2-17):

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{\int P(D|\theta)P(\theta)d\theta} \tag{2-19}$$

As Bayesian inference tries to estimate a distribution of θ, often a hyperprior, namely prior information of the distribution of θ, is introduced, resulting in the prior $P(\theta|\alpha)$ with α a set of hyperparameters. Han and Song [23] demonstrate how smart choice of hyperprior distributions can be used to account for both spatial and temporal sparsity patterns to reconstruct

an object using compressed sensing.

Generally, a closed form solution of the integration in (2-19) only exists when the prior distribution is conjugate to the posterior probability. This is often not the case, and thus the posterior distribution cannot be found directly. There are however numerical methods that can approximate the distribution. Markov chain Monte Carlo (MCMC) methods are one approach. It asymptotically samples from the posterior distribution using Markov chains. Storing these samples then allows for reconstruction of the posterior density using histograms. Finding the posterior distribution now becomes a sampling problem, only limited by the size of the data and complexity of the models.

Variational inference (VI) [24] is another approach to solving this problem. It is faster and therefore scales better for larger sets of data. VI replaces the sampling problem of MCMC with an optimization problem. First, a family of approximate densities Q is made to approach $P(\theta|D)$. Then, the Kullback-Leibler divergence from the posterior is used to find the member of Q that best approximates the posterior, finally using it in place of the posterior to do inference. The choice of Q can limit the complexity, thus improving speed but reducing the accuracy of the fit.

Chapter 3

Markov Chain Monte Carlo

In this chapter, Markov chain Monte Carlo (MCMC) [25] methods will be discussed. In Bayesian inference problems, MCMC methods are used to sample from an unknown posterior distribution. The first section of this chapter discusses the theory behind MCMC. Section two describes a simple MCMC method, the Metropolis-Hastings algorithm [26]. Section three presents reversible jump Markov chain Monte Carlo (RJMCMC) for estimating both the model and the parameter distribution simultaneously. Finally, section four discusses convergence for black box problems and post-processing.

section five discusses the practice of MCMC, such as tuning the hyperparameters, chain mixing, and label switching.

3-1 Theory of Markov chain Monte Carlo

MCMC methods do Monte Carlo simulations that sample from Markov chains to asymptotically approximate a distribution. Monte Carlo simulation refers to computational techniques that approximate calculations by using random sampling, while Markov chains are explained in definition 3.1.

Definition 3.1. ***(Markov chain)*** *Given a sequence of random variables* $\begin{bmatrix} X_1 & X_2 & X_3 & ... & X_t \end{bmatrix}$ *from stochastic process* $\{X_t\}$*, if the Markov property holds:*

$$P(X_t|X_{t-1}, X_{t-2}, ..., X_0) = P(X_t|X_{t-1})$$

then the sequence is a Markov chain and the process is a Markov process.

To understand MCMC, it is also important to understand stationarity and reversibility, given in definitions 3.2 and 3.3, respectively.

Definition 3.2. ***(Stationarity)*** *Given a stochastic process* $\{X_t\}$*, if for every* τ *it holds that*

$$P(X_0, X_1, ..., X_t) = P(X_\tau, X_{\tau+1}, ..., X_{\tau+t})$$

then the process is stationary.

Definition 3.3. ***(Reversibility)*** *Given a Markov process $\{X_t\}$, if the process upholds detailed balance:*

$$P(X_t|X_{t-1}) = P(X_{t-1}|X_t)$$

for $\forall t \in \mathbb{R}$ then the Markov chain is reversible.

3-1-1 Ordinary Monte Carlo

Monte Carlo simulation, or Ordinary Monte Carlo (OMC), can be considered a special case of MCMC. It works as follows. Say you want to calculate the expected value of a function:

$$\mu = E\left[g(X)\right] \tag{3-1}$$

with g some real function and X a stochastic process. Now suppose no analytical solution can be found. If it is possible to generate independent samples X_1, X_2, up to X_n that are identically distributed to X, the expected value can be approximated as:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} g(X_i) \tag{3-2}$$

with $\hat{\mu}$ the estimate of the expected value, which is normally distributed as $\mathcal{N}(\mu, \frac{\sigma^2}{n})$ following from the central limit theorem (CLT), with $\sigma^2 = var\{g(X)\}$. OMC can therefore asymptotically approximate the expected value of a function that has no analytical solution.

3-1-2 Markov chain Monte Carlo

In MCMC, the samples from distribution X are now simulated with a Markov chain. Suppose the Markov chain is stationary and has the same initial distribution as X, the expected value can again be approximated following the CLT as in equation (3-1). In this case, the variance changes due to the dependence in the Markov chain as:

$$\sigma^2 = var\{g(X_0)\} + 2\sum_{k=1}^{n} cov\{G(X_0), G(X_k)\} \tag{3-3}$$

as $n \to \infty$. However, in practice the distribution of X is likely to be unknown. If it were known, then OMC could just be applied directly and there would be no need to use Markov chains. The main strength of MCMC algorithms is the ability to asymptotically reach a stationary distribution that matches that of X, thereby making it possible to reconstruct unknown probability distributions. One such MCMC algorithm is the Metropolis-Hastings (MH) algorithm.

3-2 Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm [26] is a generalization of the Metropolis algorithm [27]. It can sample from an unknown target distribution $\pi(\theta)$ given a known, similar distribution $f(\theta)$. The MH algorithm works as follows. First, the Markov chain is initiated with θ_0, either chosen randomly or based on some prior knowledge of the distribution. Then, a candidate θ' is generated from a proposal function:

$$\theta' \sim q(\theta'|\theta_i) \tag{3-4}$$

for i up to n candidates and with $q(\theta'|\theta_i)$ the proposal distribution. For each candidate, the acceptance rate α is then calculated, and the Markov chain for θ is updated:

$$\alpha = \min\left\{1, \frac{f(\theta')}{f(\theta_i)}\frac{q(\theta_i|\theta')}{q(\theta'|\theta_i)}\right\} \tag{3-5}$$

$$\theta_{i+1} = \begin{cases} \theta', & \text{if } \alpha > u \\ \theta_i, & \text{otherwise} \end{cases} \tag{3-6}$$

$$u \sim \mathcal{U}(0,1) \tag{3-7}$$

with $q(\theta_i|\theta')q(\theta'|\theta_i)$ the so-called Hastings term. The Hastings term guarantees each jump in the chain is reversible, which ensures asymptotic convergence of the Markov chain to the target distribution using an arbitrary proposal distribution. We thus have a method to reconstruct an unknown distribution and find its expected value. Applying the MH algorithm to a Bayesian estimation problem and using a random walk proposal distribution, it is now possible to find the posterior distribution of parameter vector θ given model k, outlined in pseudocode in algorithm 1. Note that the Hastings term drops out when using symmetrical proposal distributions.

Algorithm 1 random walk Metropolis-Hastings algorithm

1: initialize θ_0
2: **for** $i = 0, 1, 2, ..., N_{samples}$ **do**
3: Sample from $\theta' \sim \mathcal{N}(\theta_i, \sigma_{RW})$
4: Calculate the acceptance rate: $\alpha = \frac{P(\theta'|k)}{P(\theta_i|k)}\frac{P(D|\theta',k)}{P(D|\theta_i,k)}$
5: Draw a random number from $u \sim \mathcal{U}(0,1)$
6: If $\alpha > u$, accept the jump and set $\theta_{i+1} = \theta'$
7: Otherwise, reject the jump and set $\theta_{i+1} = \theta_i$
8: **end for**

3-3 Reversible jump Markov chain Monte Carlo

Though MCMC algorithms can reliably sample from the target distribution, they require that the model is known. For many practical problems, including localization microscopy

for densely labeled samples, this is not the case. To approximate distributions for both the parameters and the model simultaneously, reversible jump Markov chain Monte Carlo (RJMCMC) [28, 29] was developed. RJMCMC considers a finite model space with a known prior, which allows it to asymptotically sample from the joint posterior of the model and parameter space. This is achieved by setting up a framework for constructing reversible, trans-dimensional move pairs. These model space moves require a generalized version of the MH acceptance rate:

$$\alpha = \min\left\{1, \frac{P(\theta', k', D) r_m(\theta')}{P(\theta, k, D) r_m(\theta) q(u)} \left|\frac{\partial(\theta')}{\partial(\theta, u)}\right|\right\} \tag{3-8}$$

with $r_m(\theta)$ the within pair probability of selecting either move and $q(u)$ the probability of drawing random number u used to generate new parameters for the increased model space, or found deterministically when removing parameters and reducing model space. An RJMCMC algorithm is thus set up by constructing a set of parameter space (such as from the random walk MH algorithm) and model space moves, then choosing which to run at each iteration, either randomly or in order. The construction of one such pair is given in the following example.

Example: Simple RJMCMC algorithm using Birth and Death move pair Suppose a posterior probability distribution, of which the prior and likelihood are known, has to be approximated with a Gaussian mixture model (GMM):

$$P(x) = \sum_{i=1}^{k} p_i \mathcal{N}(\mu_i, \sigma_i) \tag{3-9}$$

with x a realisation, k the model parameter for number of elements, p_i the element weight, and μ_i and σ_i the parameters for the i^{th} mixture element. for N realisations, the data D follows from:

$$P(D) = P(x_1)P(x_2)\dots P(x_N) = \prod_{j=1}^{N} \sum_{i=1}^{k} p_i \mathcal{N}(\mu_i, \sigma_i) \tag{3-10}$$

$$\sum_{i=1}^{k} p_i = 1 \tag{3-11}$$

The parameter vector $\theta \in \mathbb{R}^{3k}$ now consists of the means, standard deviations, and element weights of k Gaussian distributions:

$$\theta = \begin{bmatrix} \mu_0 & \sigma_0 & p_0 & \mu_1 & \sigma_1 & p_1 & \dots & \mu_k & \sigma_k & p_k \end{bmatrix}$$

To move a single element, a random walk MH move is constructed. It randomly selects one element i^* and then updates the parameters:

$$\mu'_{i^*} = \mu_{i^*} + \mathcal{N}(0, \delta\mu)$$
$$\sigma i^{*\prime} = \sigma_{i^*} + \mathcal{N}(0, \delta\sigma)$$

with $\delta\mu$ and $\delta\sigma$ some user defined parameters for setting the random walk step size. For the constrained weights p, an extra parameter w is introduced:

$$\log(w'_i) = \log(w_i) + \mathcal{N}(0, \delta w)$$
$$p_i = w_i / \sum_{l=1}^{k} w_l$$

with δw another tuning parameter. The acceptance rate for the single move is now simply:

$$\alpha_{single} = \min\left\{1, \frac{P(\theta', k, D)}{P(\theta, k, D)}\right\}$$

with θ' the updated parameters. To move from model k to model $k+1$, a birth move is set up, proposing the following jump:

$$\theta' = h(\theta, u_1, u_2, u_3) = \begin{bmatrix}\theta & u_1 & u_2 & u_3\end{bmatrix}$$
$$u_1 = \mu_{k+1} \sim \mathcal{U}(\min(D), \max(D))$$
$$u_2 = \sigma_{k+1} \sim \mathcal{U}(\sigma_{min}, \sigma_{max})$$
$$u_3 = p_{k+1} \sim \beta(1, k)$$
$$p_{i \neq k+1} = p_i(1 - p_{k+1})$$

Here, the new mean is sampled uniformly from the range of all measured data, the new standard deviation is sampled uniformly from some user defined bounds, and the new weight sampled from a beta distribution while rescaling the other weights. The acceptance rate is then constructed as:

$$\alpha_{birth} = \min\left\{1, \frac{P(\theta', k+1, D)}{P(\theta, k, D)} \frac{r_m(\theta')}{r_m(\theta)} \frac{1}{q(u)} \left|\frac{\partial(\theta')}{\partial(\theta, u)}\right|\right\}$$
$$= \min\left\{1, \frac{P(\theta', k+1, D)}{P(\theta, k, D)} \frac{r_m(\theta')}{r_m(\theta)} \frac{u_3^k}{P(u_1)P(u_2)P(u_3)}\right\}$$

To make a reversible pair, the inverse of the birth move is constructed. The death move reduces a model of $k+1$ to k sources by randomly omitting one:

$$\theta = h'(\theta')$$
$$p_{i \neq i*} = p_i/(1-p_{i*})$$

with $i*$ the element to be removed. Inverting the birth acceptance rate, the death acceptance rate becomes:

$$\alpha_{death} = \min\left\{1, \frac{P(\theta,k,D)}{P(\theta',k+1,D)}\frac{r_m(\theta)}{r_m(\theta')}\frac{P(u_1)P(u_2)P(u_3)}{u_3^k}\right\}$$
$$= \min\left\{1, \frac{P(\theta,k,D)}{P(\theta',k+1,D)}\frac{r_m(\theta)}{r_m(\theta')}\frac{1}{p_{k+1}^k}\right\}$$

Finally, setting the within pair probability of selecting either a birth or death to be equal yields $r_m(\theta') = r_m(\theta)$, simplifying the acceptance rates further. The full algorithm then follows the pseudocode in algorithm 2.

Algorithm 2 simple RJMCMC algorithm

1: initialize θ_0, k_0
2: **for** $i = 0, 1, 2, \ldots, N_{samples}$ **do**
3: Select a move with probability P_{single}, P_{birth}, P_{death}
4: execute the move and calculate the acceptance rate
5: Draw a random number from $u \sim \mathcal{U}(0,1)$
6: If $\alpha > u$, accept the jump and set $\theta_{i+1} = \theta'$
7: Otherwise, reject the jump and set $\theta_{i+1} = \theta_i$
8: **end for**

3-4 The practice of Markov chain Monte Carlo

While in theory MH and other MCMC algorithms asymptotically sample from the target distribution, their application requires some tuning to be practical. With the high computational demand of MCMC, convergence may take too long and thus this method suffers from the curse of dimensionality. What follows are some guidelines to ensure proper convergence while keeping the computational load reasonable.

3-4-1 Convergence for black box problems

Most practical problems are black box problems, meaning the researcher has no knowledge of the target distribution or expected value up for estimation. In these problems, while running MCMC it may appear as if the chain has converged, while it has not. This problem

is called pseudo-convergence and it occurs when the Markov chain struggles to escape from local optima. If the total chain length is too short to escape this optimum, the researcher may falsely conclude that the stationary distribution was found. Note that this problem is different from finding multimodal stationary distributions. A multimodal target distribution may cause pseudo-convergence when the chain struggles to transition between modes, however when the transition is easy pseudo-convergence does not occur. Also note that pseudo-convergence can occur without the target distribution being multimodal. Knowing this, pseudo-convergence can be limited by:

1. running a sufficiently long chain
2. ensuring proper mixing of the chain

A multistart approach to MCMC can also be taken, starting several chains with varying initial distributions and stopping when they all converge to the same spot. However, it is not possible to properly initialize multistart or parallel MCMC without knowledge of the target distribution, hence it is still possible to see pseudo-convergence even with this method. It is therefore more effective to use that computational power on one long run, avoiding false confidence and giving the chain enough opportunity to escape from local optima. This works as long as the moves are capable of escaping local optima given enough jumps. Smaller problems may be run overnight for hundreds of thousands of iterations to make absolutely sure no pseudo-convergence has occurred.

To reduce the number of iterations needed to escape local optima and arrive at the target distribution, there needs to be enough 'mixing' in the Markov chain. This comes down to tuning the hyperparameters such that the jumps in parameters are large enough to overcome big gaps in between optima, while also being not so large that the majority of the jumps gets rejected. The chain length and other hyperparameters can be tuned based on the time series and autocorrelation of previous runs as well as the average acceptance of jumps.

3-4-2 Chain mixing and hyperparameters

For the MH algorithm, an average acceptance rate between 40 to 60% generally ensures good mixing of the chain. Analysing the time series and autocorrelation plots gives information on which direction to take for tuning the hyperparameter. If the autocorrelation takes long to reach zero, the dependence in the chain is too strong, and bigger jumps need to be made. This is then paired with slow trends in the time series plot, again indicating convergence problems. If the autocorrelation reaches zero fast, the jumps are large enough to converge to the target distribution quickly. However, if this leads to long flat lines in the time series, the jumps are in fact too large to effectively sample from the target distribution, requiring an excessively long chain to reconstruct the distribution. The jumps need to be made smaller to sample more effectively.

Having tuned the jump size, the required chain length can be inferred from the autocorrelation plot. The lag τ autocorrelation is given as:

$$R_{XX}(\tau) = \frac{\sum_{i=1}^{N-\tau}(X_i - \bar{X})(X_{i+\tau} - \bar{X})}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \tag{3-12}$$

with R_{XX} the correlation between data X and itself, i the i^{th} element of data X, $\bar{X}$ the mean of X, and τ the lag. The chain length may be set to some integer multiple of the lag required for the autocorrelation to reach zero, say at least 3 times as long. Note that the initial distribution of the chain is likely not the target distribution, therefore initial samples may have to be discarded. This part of the chain is labeled the 'burn-in' fraction and is often necessary in black box problems. Fast convergence through good mixing can help minimize the burn-in length and hyperparameters can even be adapted when transitioning from burn-in to kept samples, using initially large jumps for fast convergence and then reducing their size for more effective sampling.

In Figure 3-1, tuning of the hyperparameters for a random walk MH algorithm is demonstrated on a toy problem. The goal is to sample from the posterior distribution of an unknown parameter θ, whose expected value is 14. The samples are updated using $\theta' = \mathcal{N}(\theta, \delta\sigma)$ and $\theta_0 = 12$, with $\delta\sigma$ up for tuning. The first row of Figure 3-1 demonstrates what happens when this parameter is chosen too small. Using $\delta\sigma = 0.01$, the autocorrelation takes 1600 iterations to reach zero, showing a slow trend and implying strong correlation between samples. This is supported by the time series plot taking a long time to reach the target distribution, and still showing slow trends. As 86% of the jumps are accepted, clearly the jump size is too small, requiring a large burn-in fraction as well as long total chain length to effectively draw independent samples from the target distribution. In the second row, $\delta\sigma$ was set to 0.5, making much larger jumps and converging almost immediately. However, the proposed jumps are so large that only 3% gets accepted, as evidenced by the flat lines in the time series. Though the dependence within the chain is weak and the convergence is fast, these settings again require a very long chain to obtain sufficient samples from the target distribution. Finally, the third row uses $\delta\sigma = 0.05$. The autocorrelation reaches 0 in 450 iterations and stays there, while the chain resembles white noise after reaching the target distribution. This implies weak correlation when sampling from the target distribution. As 62% of the jumps are accepted, these settings should allow for effective sampling while not having to discard a long burn-in chain. The jumps can be made slightly larger to get the average acceptance rate down further and improve convergence more.

3-4-3 Label switching

After running the algorithm and discarding any burn-in, the information has to be extracted from the chains. Depending on the model, label switching may become a problem. When approximating distributions with a GMM, but also when estimating multiple emitter positions in highly dense frames, the likelihood of the positions is invariant to the label of the source. This leads to samples from the ground truth parameters switching labels within the algorithm, leading to false multimodality. For a GMM, Richardson and Green [29] fix this by relabeling the means after each iteration, ordering them as $\mu_0 < \mu_1 < \cdots < \mu_k$. For label switching in two or more dimensions, the samples can be relabeled using the k-means algorithm if the target distribution fits a Gaussian. After relabeling, the target distribution can then be retrieved by constructing a histogram of the chain.

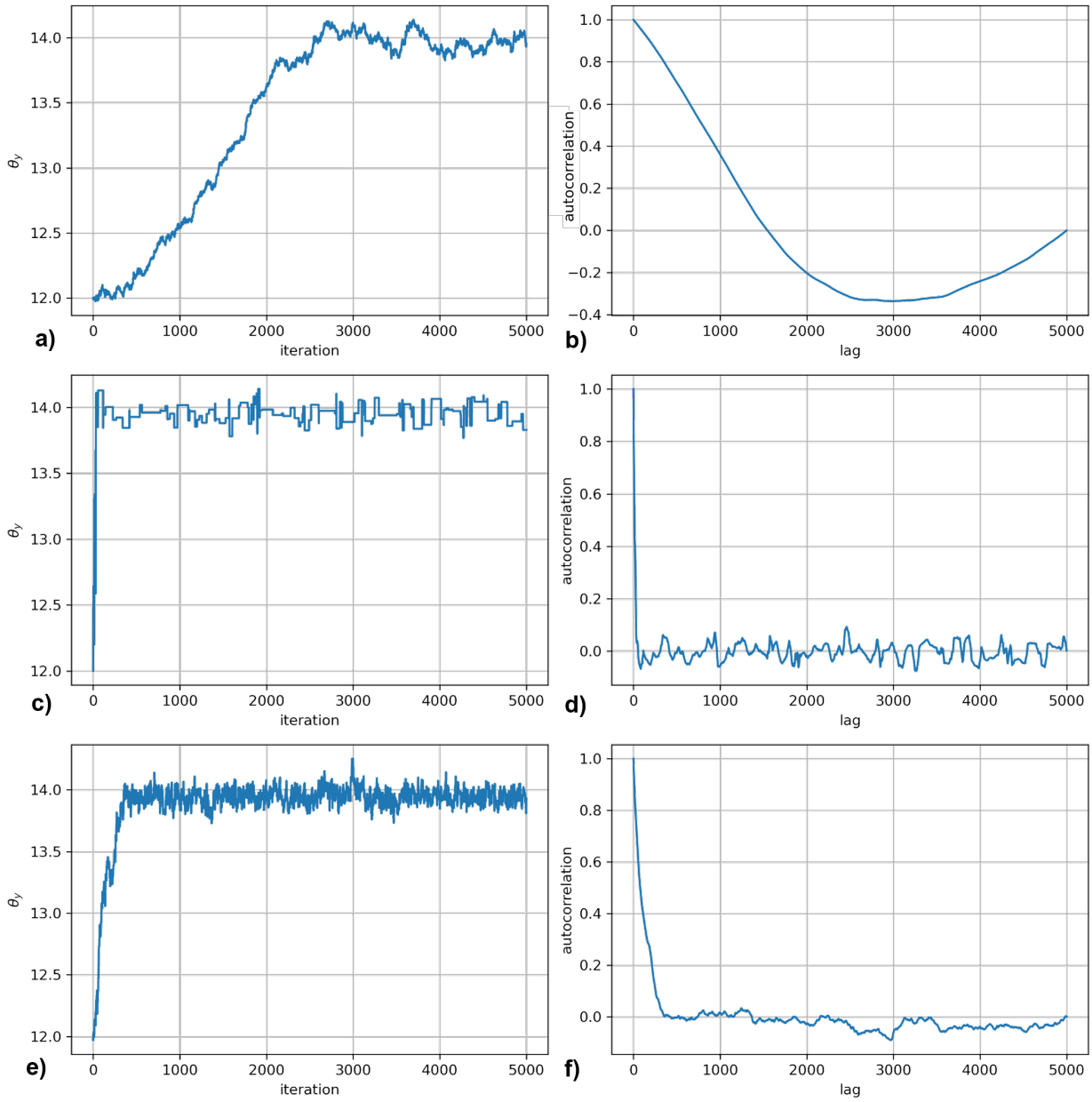


Figure 3-1: Tuning of the random walk MH jump hyperparameter $\delta\sigma$, given parameter jump $\theta' = \mathcal{N}(\theta, \delta\sigma)$ initialized in $\theta_0 = 12$. **a,b)** Time series and autocorrelation plot of samples from θ using $\delta\sigma = 0.01$. The average acceptance rate is 0.86. The autocorrelation crosses 0 after 1500 iterations, indicating strong correlation in the chain and slow convergence, as evidenced by the time series plot. **c,d)** Time series and autocorrelation plot of samples from θ using $\delta\sigma = 0.5$. The average acceptance rate is 0.03. The autocorrelation plot implies fast convergence, however the time series plot shows many flat lines, implying ineffective sampling. **e,f)** Time series and autocorrelation plot of samples from θ using $\delta\sigma = 0.05$. The average acceptance rate is 0.62. The autocorrelation reaches 0 in 450 iterations and the time series plot resembles white noise around the expected value.

Chapter 4

3D Bayesian multiple emitter fitting using reversible jump Markov chain Monte Carlo

In this chapter, the main body of the thesis is presented in the form of a manuscript, including the supplemental document.

Analyzing point spread functions for multiple emitter fitting by posterior density estimation

Raymond van Dijk[1,*], Jelmer Cnossen[1], Dylan Kalisvaart[1], and Carlas S. Smith[1,2,*]

[1]Delft University of Technology, Delft Center for Systems and Control, Delft, 2628 CD, the Netherlands
[2]Department of Imaging Physics, Delft University of Technology, Delft, the Netherlands
[*]vandijkraymond@gmail.com, c.s.smith@tudelft.nl

ABSTRACT

Single-molecule localization microscopy requires sparse activation of emitters to circumvent the diffraction limit. In densely labeled or thick samples, overlap of emitter images is inevitable. Single-molecule localization of these samples results in a biased parameter estimate with a wrong model of the number of emitters. The posterior density of the localization estimate is often incorrectly assumed to be Gaussian. To better estimate the model, parameters, and uncertainties, the posterior density is reconstructed. A three-dimensional Bayesian multiple emitter fitting algorithm was constructed using Reversible jump Markov chain Monte Carlo. It reconstructs the posterior density of both the model and the parameters, namely the three-dimensional position and photon intensity, of overlapping emitters. Imposing a Gaussian prior distribution on emitter intensity helps the algorithm converge to the correct model. The ability of the algorithm to separate two emitters at varying distance was evaluated using an astigmatic point spread function. We found that the posterior distribution of the emitter position is multimodal, when emitters are within two times the point spread function width. This multimodality describes the uncertainty in position that astigmatism introduces in localization microscopy. Biplane imaging was also tested, proving capable of separating emitters up to 0.75 times the point spread function width while free of multimodality. The posteriors seen in astigmatic and biplane imaging demonstrate how the algorithm can identify point spread function degeneracy and evaluate imaging techniques for three-dimensional multiple-emitter fitting performance.

Introduction

Single-molecule localization microscopy (SMLM)[1,2] circumvents the diffraction limit through sparse activation of emitters and reaches theoretical minimum uncertainty[3]. Overlap of emitter images is inevitable in densely labeled samples and thick samples for 3D imaging. As SMLM assumes a single molecule model, denser regions of interest (ROI) result in inaccurate estimates that have to be discarded.

Multiple emitter fitting circumvents this problem by extending the model to account for more than one emitter in the ROI. For two-dimensional localization, various high density localization methods exists, including temporal correlation[4], compressed sensing[5,6], deep learning[7], and posterior density reconstruction[8]. These methods work by simultaneously estimating the model and parameters, or making a model-free reconstruction. Bayesian approaches have the added advantage of including prior information and more accurately representing the uncertainty of model and parameter estimates. Fazel et al.[8] used reversible jump Markov chain Monte Carlo (RJMCMC), a method to sample directly from the posterior distribution, reconstructing the posterior by making a histogram of the samples. The sampler also makes model space jumps, changing the number of parameters while estimating the model.

One of the problems in high density 3D localization microscopy is PSF degeneracy[9]. As the PSF changes over depth, the image of an emitter at a given depth may match that of a sum of emitters at different depths. This increases model and parameter uncertainty, where a k-emitter model can be represented by a different number of emitters at different positions. It thus complicates the use of most 2D multiple emitter fitting methods for 3D, as they misrepresent these uncertainties. Figure 1 b) through d) illustrates these problems with high density imaging and PSF degeneracy.

In this article, we construct a 3D Bayesian localization algorithm using reversible jump Markov chain Monte Carlo (RJMCMC)[10]. The algorithm is described in detail in the Supplementary Note and follows the structure from Fazel et al.[8]. It provides a more accurate reconstruction of the uncertainty through posterior density sampling, constructing probability distributions for the number of emitters and their parameters. In addition, its use of prior information helps to separate emitters. The reconstructed posteriors are used to identify 3D PSF degeneracy in high density imaging.

Figure 1 a) shows a schematic of the algorithm. After gathering the frames and correcting for the camera gain, the user sets priors and hyperparameters that are appropriate for the imaging conditions. Then, an RJMCMC localization algorithm is ran

on each of the frames, finding the posterior of the parameters and number of emitters given the data. Using the posterior, the maximum a posteriori (MAP) number of emitters is selected, and the estimates within this model are used to start a Markov chain Monte Carlo (MCMC) localization run. The MCMC output is used to form the histogram that reconstruct the object.

Results

Two emitter separability for astigmatic imaging

We initially evaluate the high density performance using an astigmatic PSF, as it is the most commonly used PSF for 3D localization. Two emitters were simulated at varying distance to one another, from 3.0 σ_{PSF} to 0.75 σ_{PSF}, to verify model convergence and the ability to separate emitters. This is shown in Figure 2. As MCMC generates samples from the posterior distribution, the reconstructions are made by plotting histograms of the MCMC chains for all of the 100 simulated frames. While the algorithm can separate the two emitters up to a distance of 2.5 σ_{PSF}, around a distance of 2 σ_{PSF} down to 1 σ_{PSF} four peaks can be distinguished, despite the maximum a posteriori (MAP) model finding two emitters. At a distance less than 1 σ_{PSF}, running a k-means clustering algorithm on the chain outputs for two clusters finds both clusters at the same position, in the middle of the frame. This shows the emitters can no longer be separated at distances lower than 1 σ_{PSF}.

To better understand what is happening between 2 σ_{PSF} and 1 σ_{PSF} separation distance, a single frame with four peaks in the reconstruction was analyzed as shown in Supplementary Figure 12. The four peaks formed two pairs of localizations or possible modes, one at the true positions and one perpendicular. A chi-squared test was done to determine if either mode was representative of the frame. Interestingly, the chi-square values of 408 and 407 showed that both modes were representative for the frame, as the chi-square value at the 95% confidence interval is 456. We tested the probability of selecting the correct mode under both of these hypotheses. The probability of error was found to be 49.8%, making the modes indistinguishable in terms of likelihood. This indicates that the algorithm is functioning correctly and the posterior distribution of the emitter position is multimodal.

While the true mode localizations find the ground truth, the alternate mode localizations (at least for this particular astigmatic PSF model) are placed not just perpendicular to the ground truth, but also at a greater depth. At a separation of 0.75 σ_{PSF}, the alternate mode is found at a depth of about -600 nm. This can be explained by the astigmatic PSF characteristics. When moving below the focal plane, the astigmatic PSF stretches along the same axis that separates the emitters. Thus, two emitters in focal plane separated along the x axis can be represented by a pair on the y axis far below focal plane. This same problem occurs when emitters are separated along the y axis, resulting in an alternative pair along the x axis above the focal plane. This implies posterior density reconstruction can be used to analyse 3D PSF quality for high density localization microscopy.

From the XZ plots in Figure 2, it can be seen that as the separation increases, the alternative mode starts to fall outside the provided PSF range of [-1.3, 1.3] μm, disappearing at large emitter separation. As the PSF range gets constrained further, this alternative mode disappears faster. This is consistent with the multimodality imaging the PSF degeneracy, as constraining the PSF range decreases the solution space.

Influence of priors on multimodality

Testing the dependency of the multimodality on the intensity prior yields Figure 3. The image data from Figure 2 was used for localization with four different intensity priors, $P_0(I)$ to $P_3(I)$. $P_0(I)$ and $P_1(I)$ combine a sloped and uniform probability at lower intensities with a Gaussian distribution around the expected count of 2000 photons, while $P_2(I)$ and $P_3(I)$ are strictly Gaussian. The standard deviation of the Gaussian peak in the intensity prior was varied from 150 up to 1000 photons shown in Figure 3 b). The first column of Figure 3 a) matches the results of Figure 2, as it uses the same priors and data, only the random number seed differs. The algorithm separates emitters up to 2.5 σ_{PSF}, multimodality is present from 2 σ_{PSF} down to 1 σ_{PSF}, and emitters are no longer separable within 1 σ_{PSF}. Using $P_1(I)$, column 2 again shows multimodality from 2 down to 1 σ_{PSF}, with failure to separate emitters within 1 σ_{PSF}. Using $P_2(I)$, multimodality is now observed at distances from 2 down to 1.25 σ_{PSF}, failing to separate emitters within 1.25 σ_{PSF}. Finally, using $P_3(I)$, multimodal reconstructions are found from 2 σ_{PSF} down to 1.5 σ_{PSF} distance. Within this distance, the correct number of emitters is not found consistently. Figure 3 c) plots the accuracy of the estimated model for the priors used, calculated by counting the number of correct estimates and dividing by the frame count. For $P_0(I)$ and $P_1(I)$, the model accuracy is greater than 98% over the range of tested distances. For $P_2(I)$, model accuracy only goes below 95% at a distance of 0.75 σ_{PSF}. Prior $P_3(I)$ decreases in model accuracy from a distance of 1.5 σ_{PSF}, going from 94% down to 5% at a distance of 0.75 σ_{PSF}. This shows that Gaussian intensity priors wider than 500 photons cannot consistently separate emitters within 1.5 σ_{PSF} of one another. While three out of four used priors can retrieve the model over 3 σ_{PSF} down to 0.75 σ_{PSF} distance, all of the used priors returned multimodal posterior densities within the range of 2 σ_{PSF} down to 1.5 σ_{PSF}. This again implies that this multimodality is not the result of a flaw in the algorithm or the use of priors and hyperparameters. Also knowing both modes passed a chi-squared test on a 95% confidence interval and having to choose between the two at a near 50% probability of error shows that this multimodality is a result of the imaging

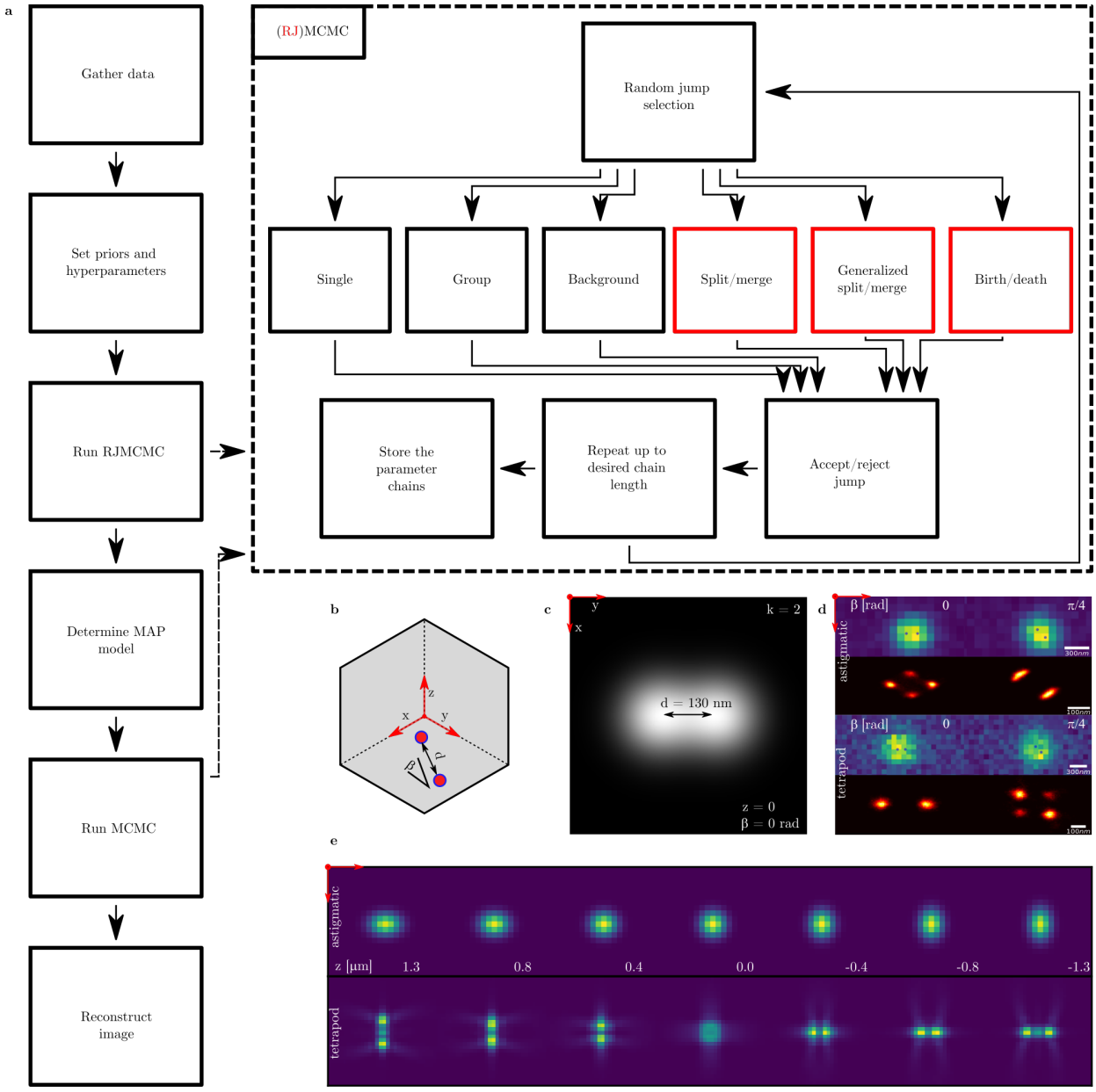


Figure 1. Schematic of the algorithm and problem description. **a)** 3D RJMCMC localization flowchart. Frames are gathered and converted to photon counts. Then, priors and hyperparameters are set. An initial run of RJMCMC samples from the joint posterior of the parameters and model, from which the MAP number of emitters is determined. Using this MAP model, another MCMC run is used to condition the parameter distribution on the estimated model. Finally, the image is reconstructed by plotting histograms of the MCMC chains. The dashed rectangle demonstrates the (RJ)MCMC algorithm. Each loop, a move is randomly selected (RJMCMC moves that act on model space are highlighted in red) and used to propose a new set of parameters. This jump in parameters is accepted or rejected based on the ratio of posteriors. The algorithm repeats this loop, storing the parameters at each iteration to finally output the chain of iterations. **b)** Diagram of two overlapping emitters separated by distance d and under angle β. **c)** Ideal image of two nearby emitters. **d)** Simulated data and reconstruction using astigmatic and tetrapod PSFs. Frames are shown in rows 1 and 3, while the reconstructed posterior distribution is shown in rows 2 and 4. **e)** Z-scan of the PSFs used in d) to generate and reconstruct the data.

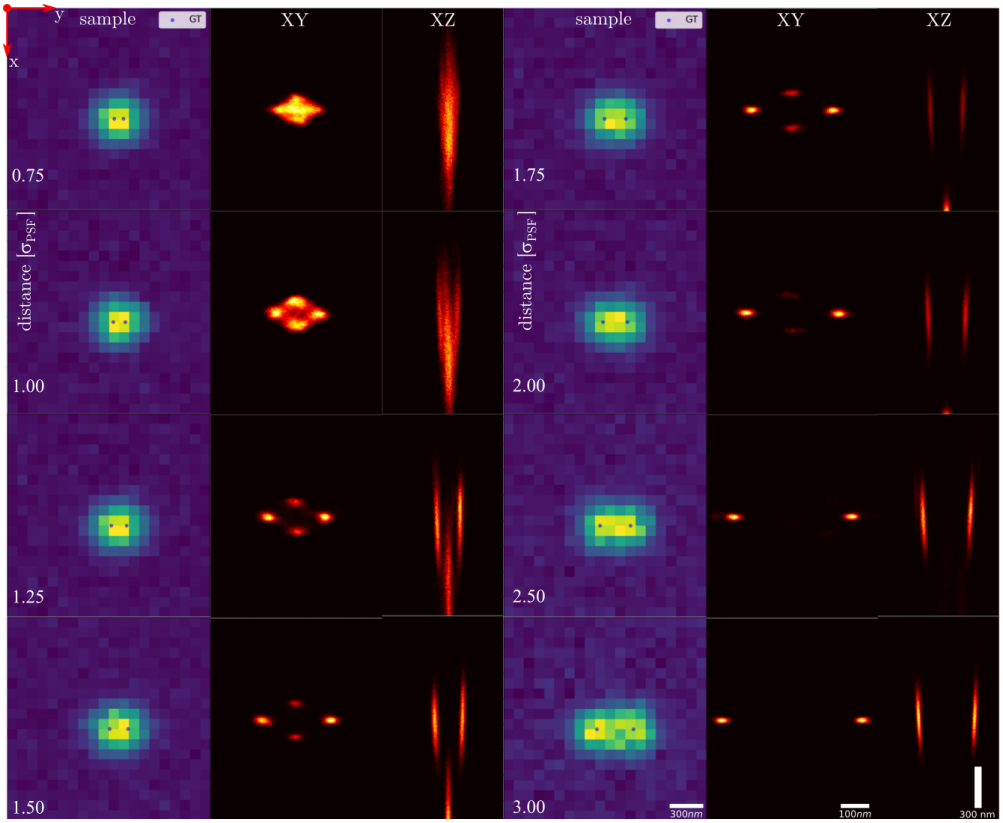


Figure 2. Two emitter separability using an astigmatic PSF, collecting 100 simulated frames into one reconstruction while varying emitter distance. **columns 1 and 4)** Example frames. **columns 2 and 5)** Zoomed in XY plane reconstruction. **columns 3 and 6)** Zoomed in XZ plane reconstruction. Emitters were placed in focal plane, with an intensity of 2000 photons each and a background of 20 photons. The ROI is 20 by 20 pixels, scalebars assume an effective pixel size of 100 nm. The reconstructed image consists of histograms from MCMC chains which used the MAP number of emitters as model.

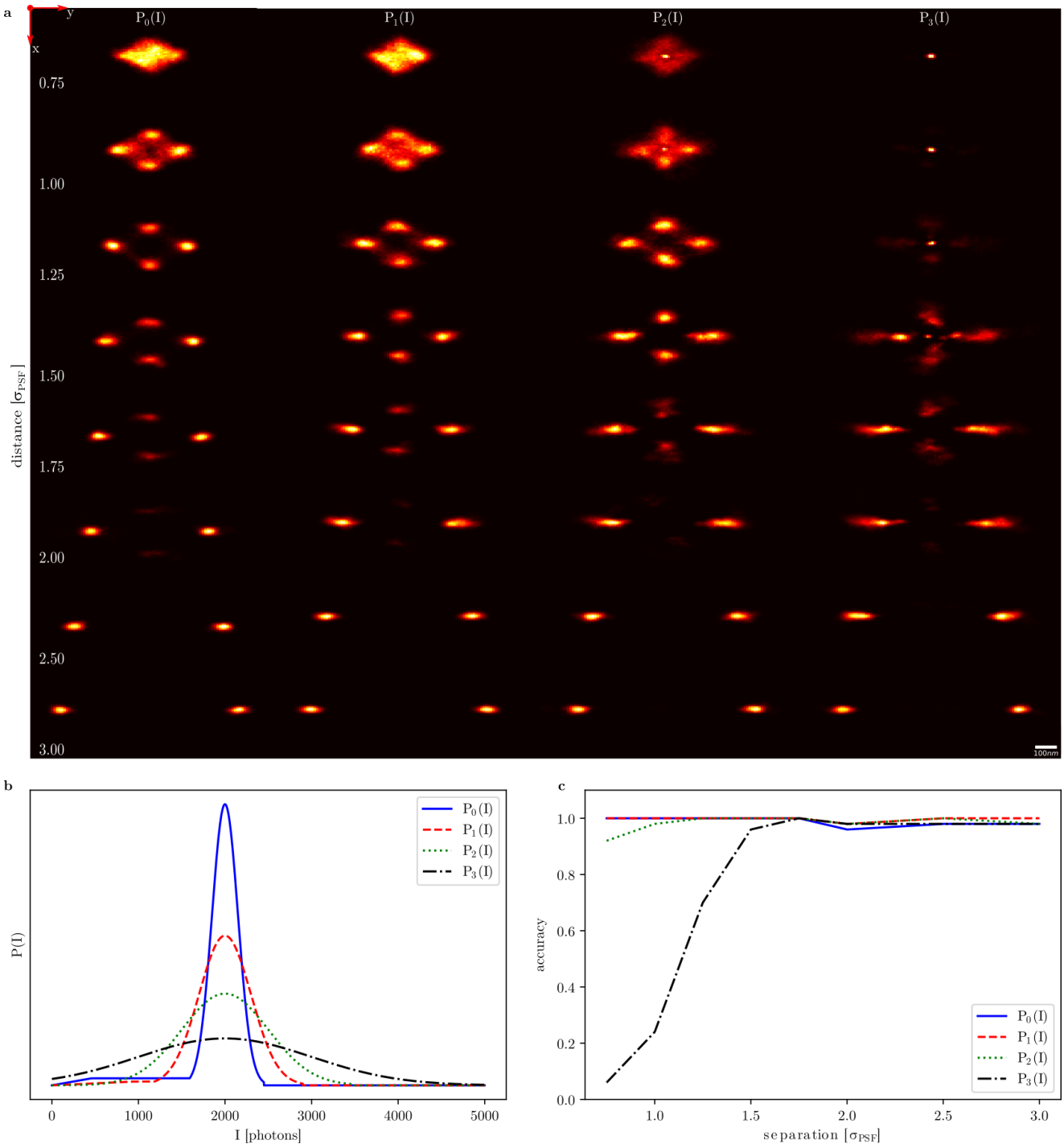


Figure 3. Two emitter separability using an astigmatic PSF, varying the width of the intensity prior, using the same frames as shown in Figure 2. **a)** XY plane reconstructions, each column using a different intensity prior. Images were formed by constructing histograms of the MCMC chains. textbfb) Plots of the respective intensity priors used in each column in a). **c)** Accuracy of the found model as varying over the emitter distance, plotted for each intensity prior used. Accuracy is found using $N_{frames,\hat{k}=k}/N_{frames}$ with $N_{frames,\hat{k}=k}$ the frames where the estimated model matches the ground truth and N_{frames} the total frame count. The widths of the Gaussian peaks in the priors are 150, 300, 500, and 1000 photons, for $P_0(I)$ up to $P_3(I)$, respectively. $P_0(I)$ and $P_1(I)$ additionally use a uniform and sloped probability at lower intensities to facilitate model space moves. Priors were set to 0 at intensities beyond $2000 + 3\sigma_{prior}$.

conditions and in particular the degeneracy of the 3D PSF that was used. The algorithm can thus be used to analyze different PSFs for degeneracy by reconstructing the posterior distribution of parameters.

Two emitter separability using biplane imaging

Changing the PSF used for the two emitter separability test, biplane imaging was tested for its ability to separate emitters and the multimodalities that may occur when doing so. For biplane imaging, the PSF can be approximated by a Gaussian. Unlike the astigmatic PSF, the 3D Gaussian PSF stays radially symmetrical over its range. It can therefore be expected that the same multimodality shown in astigmatic imaging will not be present here. Figure 4 shows that is indeed the case. As the emitter distance varies over the same range of 0.75 to 3 σ_{PSF}, the algorithm consistently finds a model of two emitters while the reconstruction also consists of just two peaks. Under these conditions, biplane imaging can separate two emitters up to a distance of 0.75 σ_{PSF}, entirely free of multimodality. This not only validates the idea that multimodality is caused by 3D PSF degeneracy, it also demonstrates how the algorithm can be used to determine which PSFs suffer the least from this problem and which are best used in dense 3D imaging.

Conclusion

3D localization microscopy struggles with overlapping emitter images, often not being able to determine the number of active emitters in the ROI and leading to inaccurate position estimates. Multiple emitter fitting algorithms can find the number of emitters, but the added complexity of 3D PSFs and PSF degeneracy means that these algorithms often misrepresent the uncertainty of their estimates. We therefore constructed a 3D RJMCMC multiple emitter fitting algorithm to sample from the posterior density.

3D RJMCMC localization can retrieve the correct model with at least 93% accuracy using Gaussian intensity priors with a width of up to 500 photons. In providing the posterior density, it also estimates the parameter uncertainty. Using the reconstructed posterior density of emitter positions not only reconstructs the object, but can aid in understanding PSF degeneracy in multimodal results. For astigmatic and biplane imaging, the algorithm is capable of separating emitters up to a distance of 1 σ_{PSF} and 0.75 σ_{PSF}, respectively, localizing emitters in 3D where SMLM methods would fail. However, astigmatic imaging at these densities will result in multimodal reconstructions. This is an accurate representation of the posterior and a consequence of the 3D PSF structure. Therefore, posterior density reconstruction is the tool of choice to identify potential PSF degeneracy problems in dense 3D localization.

Discussion

Multimodality can be reduced by limiting the PSF depth range. The proposed RJMCMC algorithm has demonstrated cases of multimodality in astigmatic PSF models. Constraining the range reduces the variety of shapes the PSF can take on, limiting PSF degeneracy. Multimodality may also be reduced by encoding the axial position in intensity. Modifying the algorithm to work with total internal reflection fluorescence (TIRF)[11,12] is therefore a promising method to image without multimodality. Modulation-enhanced localization microscopy (meLM)[13] techniques such as ModLoc[14], SIMFLUX[15], and ROSE[16] could all be decoded with 3D RJMCMC localization and a position-dependent intensity prior. A combination of these techniques with RJMCMC may result in a posterior distribution free of multimodality.

In conclusion, the 3D RJMCMC localization algorithm is best used to analyze imaging techniques for their effectiveness in 3D multiple emitter fitting. Though no multimodality was revealed when testing biplane imaging with 3D RJMCMC localization, orientations of emitter pairs were not exhaustively tested. This combined with the slow and memory intensive nature of RJMCMC means we do not recommend using the algorithm for localization. However, it can be used to analyze PSFs for possible multimodality in dense samples. Testing imaging techniques that would require a position dependent intensity prior, such as TIRF or any meLM technique, is also a topic of great interest.

Methods

Image formation model and key probabilities

For multiple emitter fitting, the expected photon count per camera pixel can be described as:

$$\mu_i = \sum_{j=1}^{k} \theta_{I,j} \int_{A_k} H(\theta_{x,j}, \theta_{y,j}, \theta_{z,j}) \mathrm{d}x\mathrm{d}y + \theta_b \tag{1}$$

with μ_i the photon count in pixel i, $\theta_{I,j}$ the intensity of the j^{th} emitter in the frame, k the amount of emitters in the frame, A_k the pixel area, $H(x,y,z)$ the PSF, $\theta_{x,j}$, $\theta_{y,j}$, $\theta_{z,j}$ the 3D position of the j^{th} emitter, and θ_b the background photon count. For a high

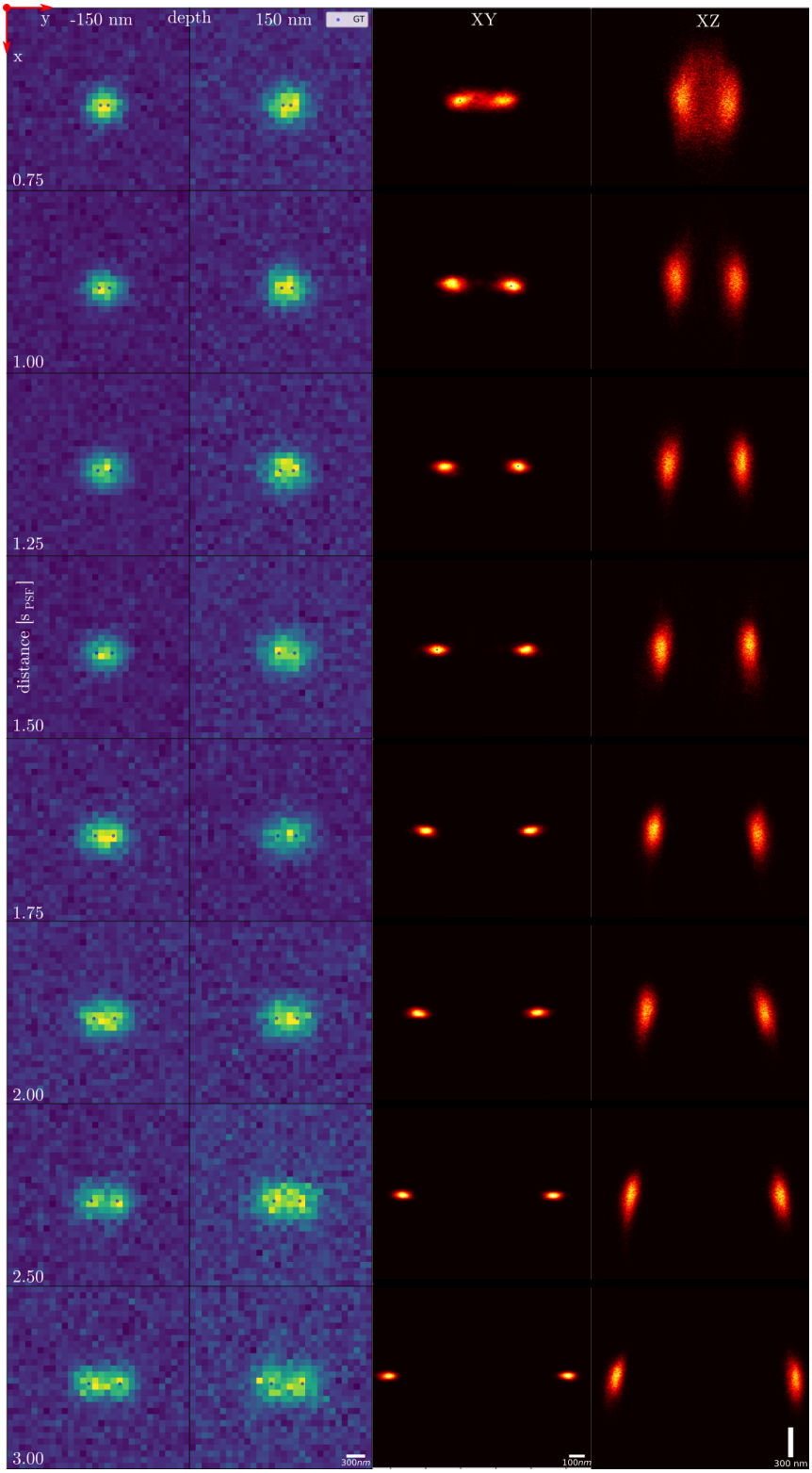


Figure 4. Two emitter separability using biplane imaging with planes separated by 300 nm. 100 pairs of frames were simulated and used to make one reconstruction. As photons are split evenly among the planes, the intensity prior was changed to be a Gaussian with mean 1000 photons and width 150 photons. **columns 1 and 2)** Example frames at both the positive and negative depth. **column 3)** XY plane reconstruction. **column 4)** XZ plane reconstruction. Emitters were placed in focal plane, with an intensity of 2000 photons each and a background of 20 photons. The ROI is 30 by 30 pixels, scalebars assume an effective pixel size of 100 nm. The reconstructed image consists of histograms from MCMC chains which used the MAP number of emitters.

gain camera, such as an electron multiplying charge coupled device (EMCCD), the readout noise is negligible and thus the likelihood function has a Poisson distribution:

$$P(D_i|\theta,k) = \frac{\mu_i^{D_i}\exp(-\mu_i)}{D_i!} \tag{2}$$

with $P(D_i|\theta,k)$ the likelihood of observing measured data D on the i^{th} pixel as function of parameter vector $\theta = \begin{bmatrix}\theta_{x,0} & \theta_{y,0} & \theta_{z,0} & \theta_{I,0} & \dots & \theta_{x,k} & \theta_{y,k} & \theta_{z,k} & \theta_{I,k} & \theta_b\end{bmatrix}$ and number of emitters k. Given the pixels are independent, the likelihood of one frame becomes:

$$P(D|\theta,k) = \prod_{i=1}^{N_p}\frac{\mu_i^{D_i}\exp(-\mu_i)}{D_i!} \tag{3}$$

with N_p the pixel count. The joint posterior distribution of the parameters and model can then be found using Bayes' rule:

$$P(\theta,k|D) = \frac{P(D|\theta,k)P(\theta|k)P(k)}{P(D)} \tag{4}$$

with $P(k)$ the model prior, $P(\theta|k)$ the parameter prior given the model, and $P(D)$ the evidence. The priors can be formulated from earlier attained knowledge of labeling density or emitter intensity, however the evidence term, $P(D) = \int P(D|\theta,k)P(\theta|k)P(k)\mathrm{d}\theta\mathrm{d}k$, only has a closed form solution when the prior distribution is conjugate to the posterior. This is often not the case[17], therefore Reversible jump Markov chain Monte Carlo (RJMCMC)[10] is employed to asymptotically sample from the posterior.

After attaining samples from $P(\theta,k|D)$, the maximum a posteriori (MAP) number of emitters, $\hat{k}$, is used as the true model to condition the parameter estimate on, running a Markov chain Monte Carlo (MCMC) algorithm to find $P(\theta|\hat{k},D)$. This is done to avoid introducing bias in the parameter estimates coming from models of different dimension.

Priors and hyperparameters

The algorithm takes in priors for 3D position, emitter intensity, background intensity, and emitter count. Throughout the tests, the prior on the lateral emitter position is kept uniform over the ROI plus four extra pixels, to account for the influence of emitters outside the ROI. The axial position prior is uniform over the presumed depth range of the PSF. Emitter intensity was set as a Gaussian distribution, enabling the algorithm to separate the emitters. In practice, it is recommended[8] to estimate the intensity prior using kernel density estimation on intensity data of a previous SMLM run. Finally, the emitter count prior is also kept uniform. Although it is possible estimate the emitter count within a sample given the label density, it can still vary widely on a local scale, thus we keep the prior uniform for smaller ROIs.

Each iteration, the move was randomly selected using user-determined selection probabilities [P_{single}, P_{group}, $P_{background}$, P_{split}, P_{merge}, $P_{g-split}$, $P_{g-merge}$, P_{birth}, P_{death}]. Throughout the tests, the RJMCMC burn-in portion uses [1/5, 1/5, 1/5, 1/15, 1/15, 1/15, 1/15, 1/15, 1/15], while post burn-in [1/4, 1/4, 1/4, 0, 0, 3/32, 3/32, 1/32, 1/32] are used. The burn-in portion uses higher model space move probabilities to ensure more model space mixing, while post burn-in focuses more on parameter space moves. The MCMC portion uses [2/5, 2/5, 1/5, 0, 0, 0, 0, 0, 0], focusing mainly on emitter parameters. Each test runs for 30,000 RJMCMC iterations, using 10,000 of those as burn-in and following them up with 5,000 MCMC iterations. The parameter space moves use random walk samplers, leaving jump sizes [σ_x, σ_y, σ_z, σ_I, σ_b] as parameters for tuning. For good mixing, lateral jump size may vary from 0.05 to 0.1 pixels, axial jump size from 0.07 up to 0.12 μm, emitter intensity between 10 to 40 photons, and background intensity of 1 to 3 photons.

Convergence and precision

To verify convergence of the algorithm, 100 frames with a single active emitter were localized. Emitters in the center of the ROI sampled their intensity randomly from the matching prior and were given a random sub-pixel displacement. Supplementary Figures 3 and 5 show the model and parameter autocorrelation as well as the time series and histogram of the model, for a high and low signal to background, respectively. The algorithm manages to converge to the correct model 100% of the time, while also converging in parameter space. By using only uniform priors, the algorithm yields an unbiased estimate that can be compared to the Cramer-Rao lower bound (CRLB)[3] to verify the precision. For non-uniform priors, the Van Trees inequality (VTI)[18,19] can be used as a Bayesian Cramér-Rao bound to find the theoretically minimum localization error. Supplementary Figure 8 shows a violinplot of the precisions found with 3D RJMCMC localization compared to the CRLB over varying emitter

intensity. It can be seen that the localization precision matches the CRLB over the plotted intensity range. Supplementary Figure 9 shows violinplots for the root mean squared error of the same data. The results show that the algorithm reaches the minimum theoretical uncertainty for low emitter density.

Synthetic data and results

For Figure 2, two emitters were placed in focus and simulated using an astigmatic PSF, their center of mass in the middle of the ROI. The PSF was evaluated using a 3D Gaussian approximation[3], with parameters $[s_{0,x}, \gamma_x, d_x, A_x]$ and $[s_{0,y}, \gamma_y, d_y, A_y]$ of $[\sigma_{PSF}, 2, 3, 0]$ and $[\sigma_{PSF}, -2, 3, 0]$, respectively. Here σ_{PSF} is the width in focal plane, set at 1.2 pixels. The PSF range was set at [-1.3, 1.3] μm, with a ROI size of 20 by 20 pixels. Emitter intensity was fixed at 2,000 photons, with a background intensity of 20 photons. A total of 100 frames were simulated and their 5,000 iteration MCMC chains were merged to finally form the histogram reconstructions of the XY and XZ planes. The histograms were magnified in x and y direction by a factor 2.5 with respect to the sample frames. Move selection probabilities were as in Subsection Priors and hyperparameters, while the jump size hyperparameters were set to [0.1, 0.1, 0.08, 15, 1]. All priors were kept uniform except the emitter intensity, using a Gaussian around 2000 photons with a width of 150 and a small uniform probability between 0 to 1500 that slopes down to 0, as shown in Figure 3 b). The prior is set to 0 for intensities greater than 2450 photons. The number of emitters ranges from 0 to 6 and lateral position estimates may exceed the ROI by four pixels. Background intensity was limited to a range of 1 to 40 photons.

Figure 3 uses the same data and settings as Figure 2, only changing the random number generator seed and the emitter intensity priors used. Priors $P_0(I)$ to $P_3(I)$ all use Gaussian distributions centered around 2000 photons, with a width of 150, 300, 500, and 1000 photons, respectively. Again, $P_0(I)$ and $P_1(I)$ keep a uniform probability sloping to 0 at lower intensities to help facilitate splitting of emitters.

Figure 4 uses 100 frames simulated with biplane imaging, splitting the response of a Gaussian PSF between planes at +150 and -150 nm depth relative to the focal plane. A depth range of [-1, 1] μm was used. Again, the PSF was evaluated with a 3D Gaussian approximation[3] of an experimentally measured astigmatic PSF on the setup described in[20], with [1.70, −4.64, 8.34, 0.00] for the x and y parameters. Emitter intensities and background photons were split evenly across the planes. The ROI is now expanded to 30 by 30 pixels. All hyperparameters and priors used were identical to the previous experiments, except the emitter intensity prior, which is a Gaussian with mean 1000 and width 150 photons. The intensity prior is set to 0 for intensities greater than 1450 photons and again is uniform between 200 and 500 photons, sloping upwards from 0 to 200 photons.

References

1. Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *science* **313**, 1642–1645 (2006).
2. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm). *Nat. methods* **3**, 793–796 (2006).
3. Smith, C. S., Joseph, N., Rieger, B. & Lidke, K. A. Fast, single-molecule localization that achieves theoretically minimum uncertainty. *Nat. methods* **7**, 373–375 (2010).
4. Dertinger, T., Colyer, R., Iyer, G., Weiss, S. & Enderlein, J. Fast, background-free, 3d super-resolution optical fluctuation imaging (sofi). *Proc. Natl. Acad. Sci.* **106**, 22287–22292 (2009).
5. Zhu, L., Zhang, W., Elnatan, D. & Huang, B. Faster storm using compressed sensing. *Nat. methods* **9**, 721 (2012).
6. Min, J. *et al.* Falcon: fast and unbiased reconstruction of high-density super-resolution microscopy data. *Sci. reports* **4**, 1–9 (2014).
7. Nehme, E., Weiss, L. E., Michaeli, T. & Shechtman, Y. Deep-storm: super-resolution single-molecule microscopy by deep learning. *Optica* **5**, 458–464 (2018).
8. Fazel, M. *et al.* Bayesian multiple emitter fitting using reversible jump markov chain monte carlo. *Sci. reports* **9**, 1–10 (2019).
9. Babcock, H. P. & Zhuang, X. Analyzing single molecule localization microscopy data using cubic splines. *Sci. reports* **7**, 1–9 (2017).
10. Green, P. J. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika* **82**, 711–732 (1995).
11. Axelrod, D. Total internal reflection fluorescence microscopy. *Methods cell biology* **30**, 245–270 (1989).
12. Bourg, N. *et al.* Direct optical nanoscopy with axially localized detection. *Nat. Photonics* **9**, 587–593 (2015).

13. Reymond, L., Huser, T., Ruprecht, V. & Wieser, S. Modulation-enhanced localization microscopy. *J. Physics: Photonics* **2**, 041001 (2020).

14. Jouchet, P. *et al.* Nanometric axial localization of single fluorescent molecules with modulated excitation. *Nat. Photonics* **15**, 297–304 (2021).

15. Cnossen, J. *et al.* Localization microscopy at doubled precision with patterned illumination. *Nat. methods* **17**, 59–63 (2020).

16. Gu, L. *et al.* Molecular-scale axial localization by repetitive optical selective exposure. *Nat. Methods* **18**, 369–373 (2021).

17. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *The journal chemical physics* **21**, 1087–1092 (1953).

18. Gill, R. D. & Levit, B. Y. Applications of the van trees inequality: a bayesian cramér-rao bound. *Bernoulli* 59–79 (1995).

19. Kalisvaart, D. *et al.* Precision in iterative modulation enhanced single-molecule localization microscopy. *Biophys. J.* (2022).

20. Hinsdale, T. A., Stallinga, S. & Rieger, B. High-speed multicolor structured illumination microscopy using a hexagonal single mode fiber array. *Biomed. Opt. Express* **12**, 1181–1194 (2021).

Author contributions statement

R.D. constructed the algorithm and wrote the code. J.C. helped with construction of the algorithm and writing the code. C.S.S., J.C., and D.K. provided key insights and helped interpret the results. C.S.S. and J.C. supervised the work. C.S.S. designed the research. All authors contributed to writing the manuscript.

Additional information

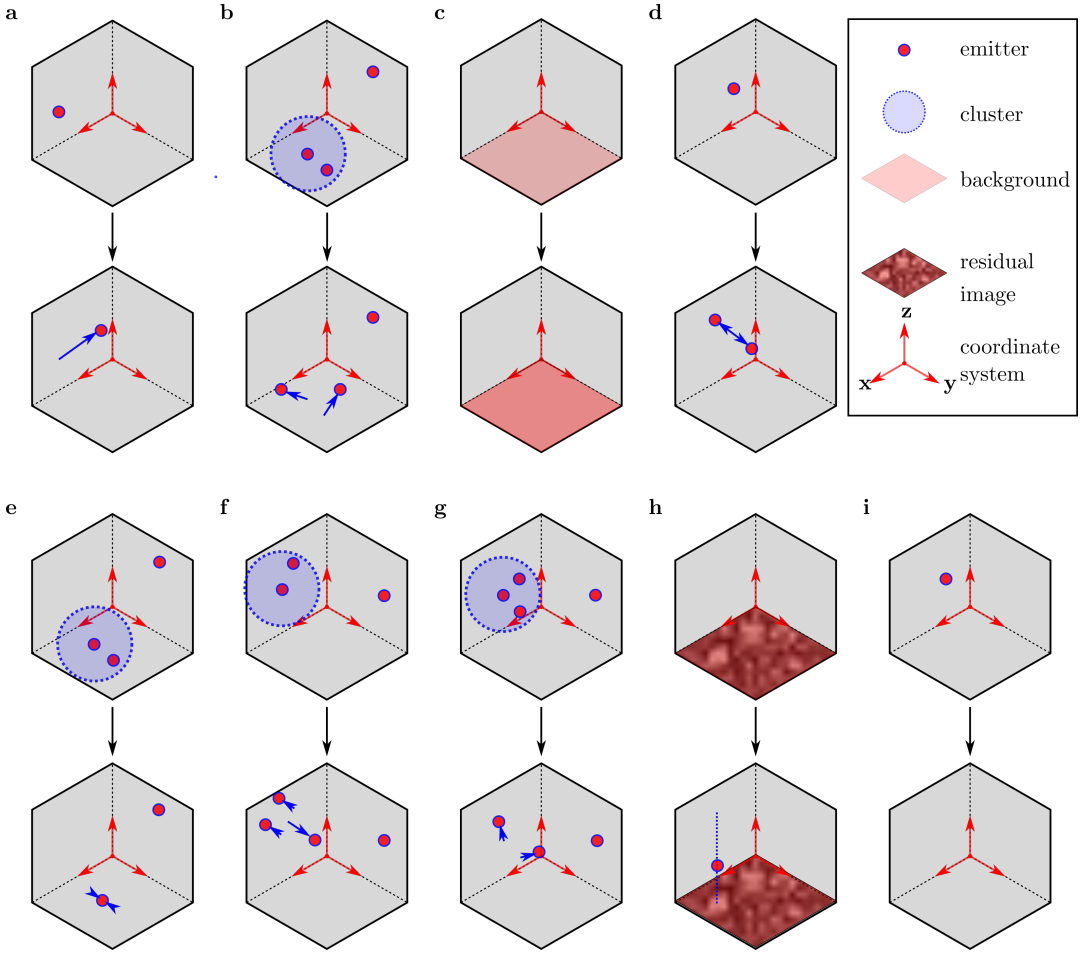
Data availability The data and code are available online from the Github repository at https://github.com/qnano/rjmcmc3D. Two Jupyter notebook examples are also included and can be run with Google colaboratory at https://colab.research.google.com/github/qnano/rjmcmc3D/blob/master/colab_example.ipynb and https://colab.research.google.com/github/qnano/rjmcmc3D/blob/master/colab_example_cspline.ipynb.

Competing interests The authors declare no competing interests.

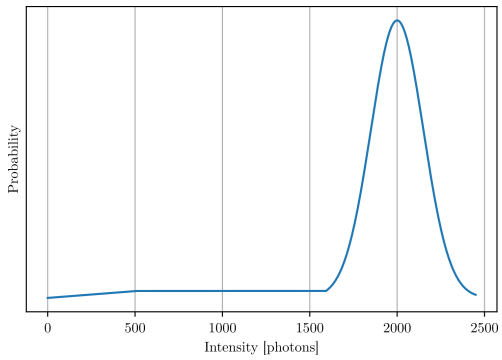
Supporting information for: Analyzing point spread functions for multiple emitter fitting by posterior density estimation

Raymond van Dijk[1,*], **Jelmer Cnossen**[1], **Dylan Kalisvaart**[1], **and Carlas S. Smith**[1,2,*]

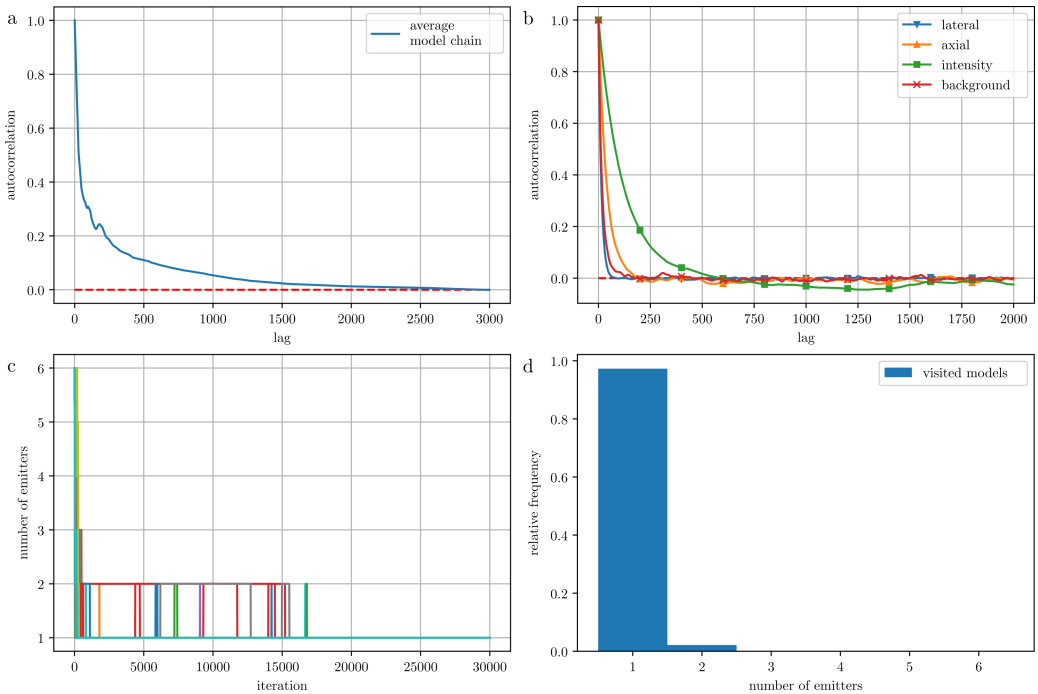
[1]Delft Center for Systems and Control, Delft University of Technology, Delft, the Netherlands
[2]Department of Imaging Physics, Delft University of Technology, Delft, the Netherlands
[*]vandijkraymond@gmail.com, c.s.smith@tudelft.nl



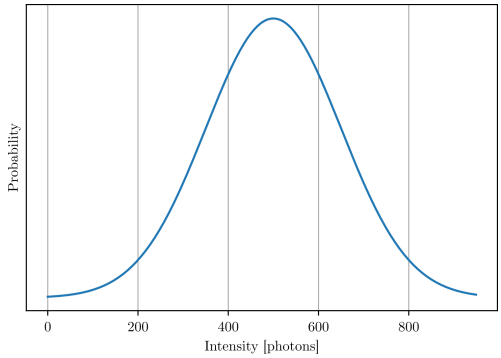
Supplementary Figure 1. A collection of diagrams visualising the proposals that each move makes. **a)** The single emitter move changes the 3D position of an emitter, as well as its intensity. **b)** The group move finds a cluster of emitters and randomly and independently changes their positions and intensities. **c)** The background move changes the estimated background photon count. **d)** Split distributes the intensity of a random emitter over two new emitters. **e)** Merge randomly selects an emitter and searches within a given radius for an emitter to merge it with, combining intensities. **f)** Generalized split finds a cluster of N emitters and splits them off into $N+1$ emitters. **g)** Generalized merge finds a cluster of $N+1$ emitters to merge into N emitters. **h)** Birth constructs a residual image to use as a probability distribution for possible undetected emitter positions, generating a new emitter with a random position. **i)** Death removes a randomly selected emitter from the model.



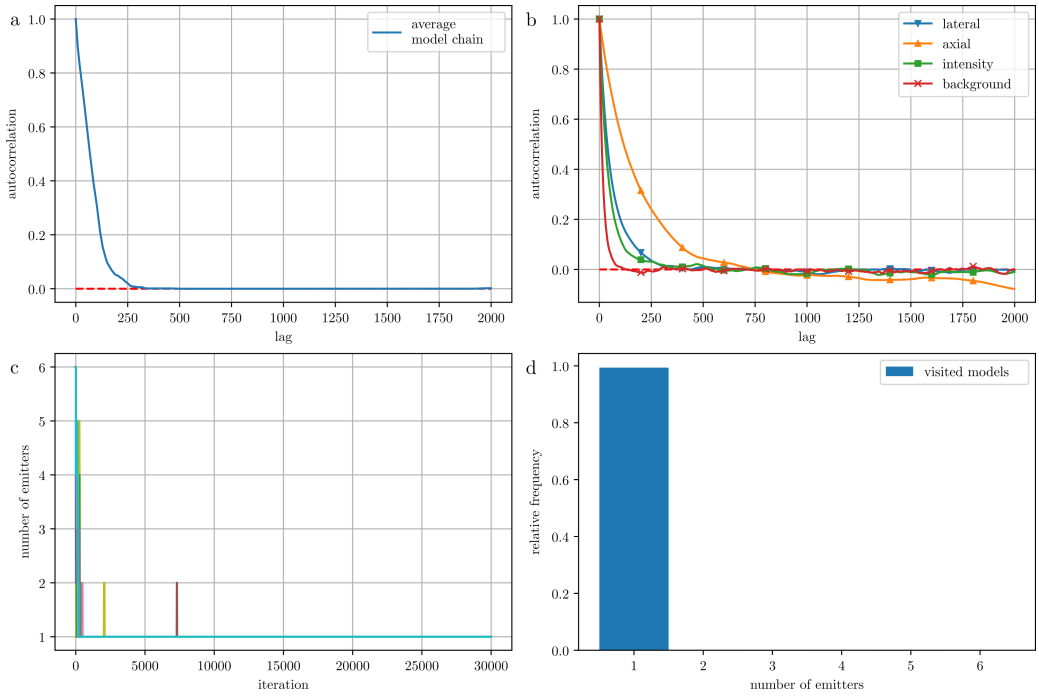
Supplementary Figure 2. plot of the prior on intensity. Used for synthetic data with intensities drawn from $\mathcal{N}(2000, 150)$.



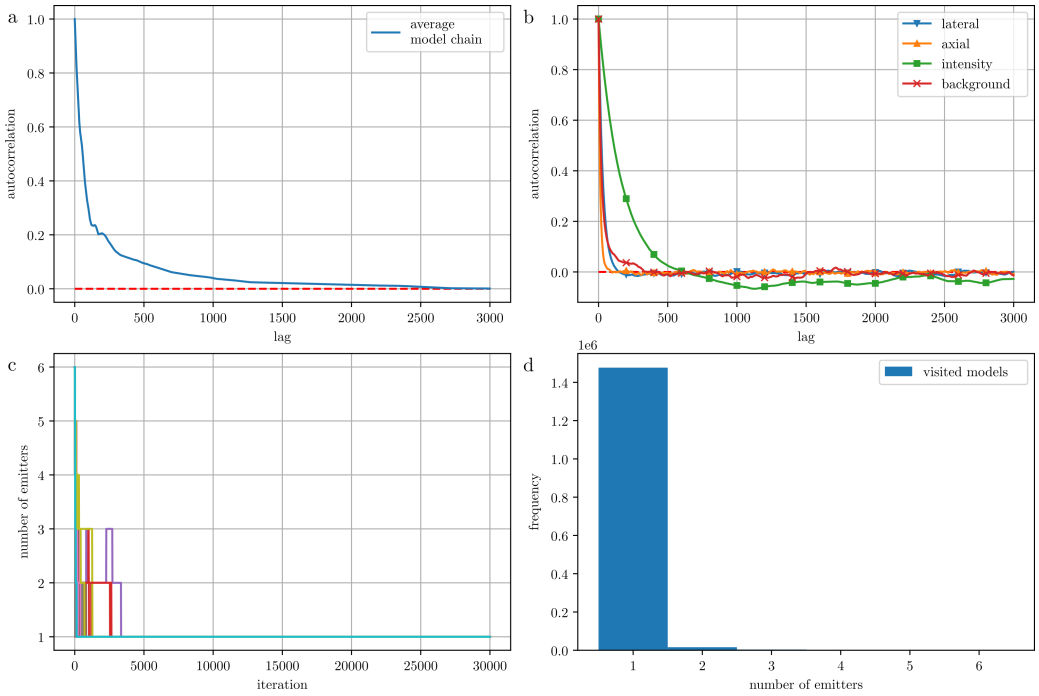
Supplementary Figure 3. Model and parameter space convergence of the algorithm for high SBR frames. **a)** Autocorrelation of the model chain. **b)** Autocorrelation of the parameters. **c)** Time series plot of the model for all frames. **d)** Histogram of all visited models across all frames. Emitter intensities were sampled from $\mathcal{N}(2000, 150)$, background was set to 20 photons. Emitters were placed in focal plane at the center of the region of interest (ROI) and given a sub-pixel shift drawn from $\mathcal{U}(-0.5, 0.5)$. 100 frames were used.



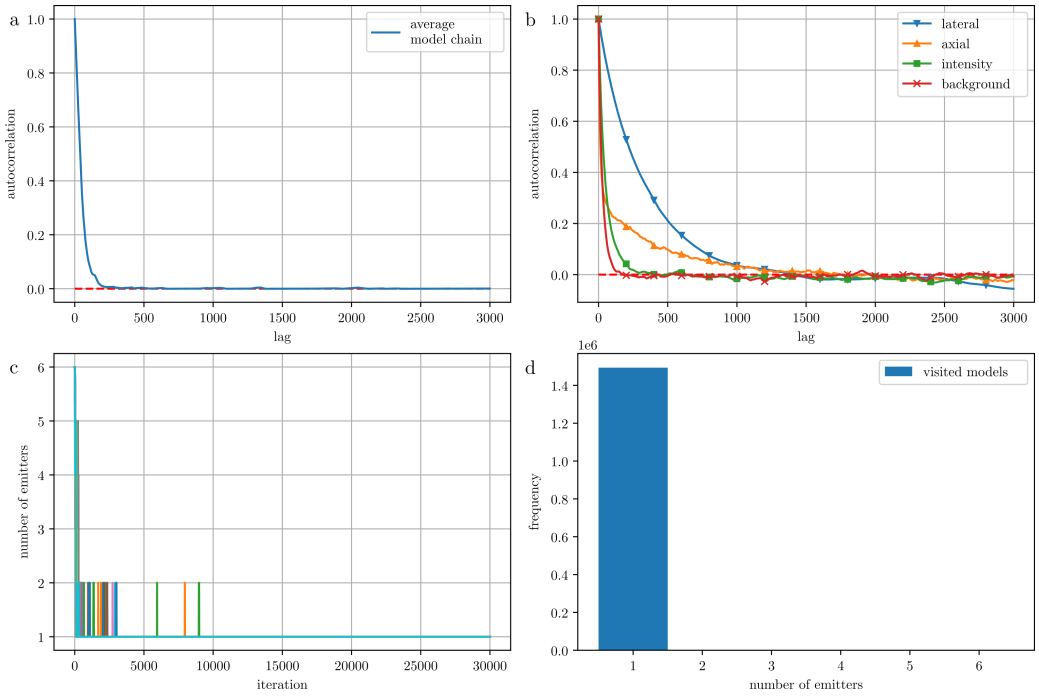
Supplementary Figure 4. plot of the prior on intensity for low SBR. Used for synthetic data with intensities drawn from $\mathcal{N}(500, 150)$.



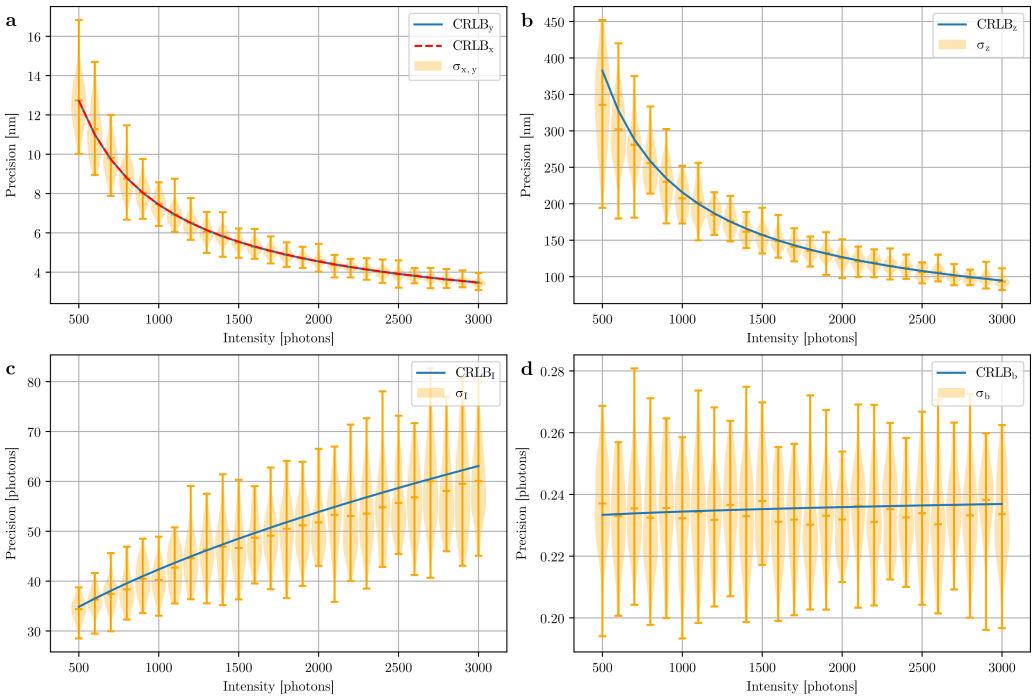
Supplementary Figure 5. Model and parameter space convergence of the algorithm for low SBR frames. **a)** Autocorrelation of the model chain. **b)** Autocorrelation of the parameters. **c)** Time series plot of the model for all frames. **d)** Histogram of all visited models across all frames. Emitter intensities were sampled from $\mathcal{N}(500, 150)$, background was set to 20 photons. Emitters were placed in focal plane at the center of the ROI and given a sub-pixel shift drawn from $\mathcal{U}(-0.5, 0.5)$. 100 frames were used.



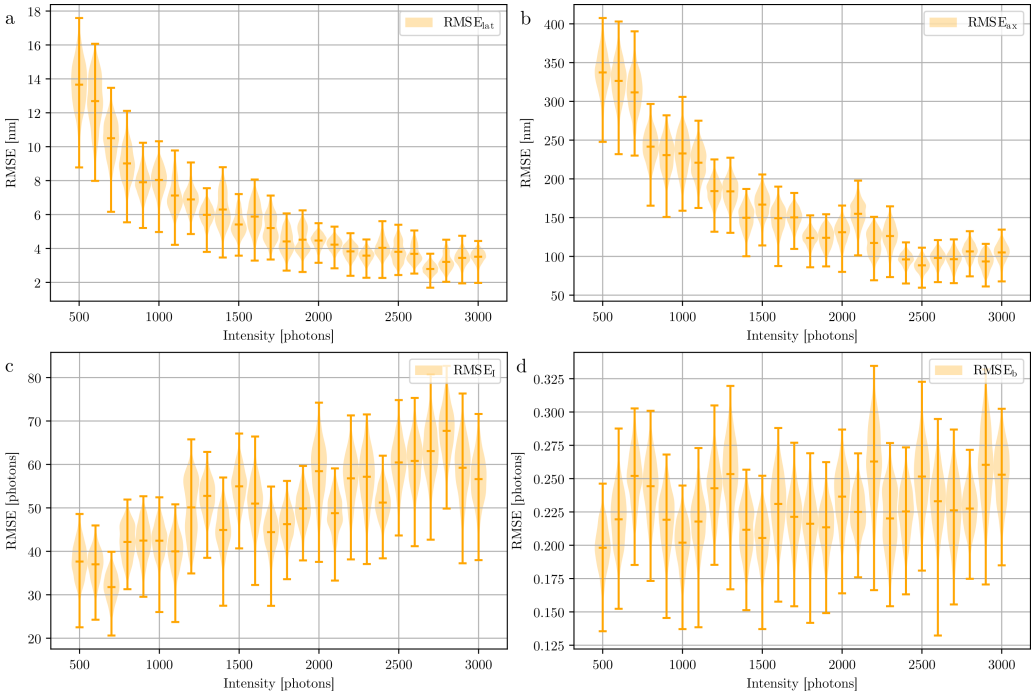
Supplementary Figure 6. Model and parameter space convergence of the algorithm for high SBR frames. **a)** Autocorrelation of the model chain. **b)** Autocorrelation of the parameters. **c)** Time series plot of the model for all frames. **d)** Histogram of all visited models across all frames. Emitter intensities were sampled from $\mathcal{N}(2000, 150)$, background was set to 20 photons. Emitters were placed in focal plane at the center of the ROI and given a sub-pixel shift drawn from $\mathcal{U}(-0.5, 0.5)$. 100 frames were used.



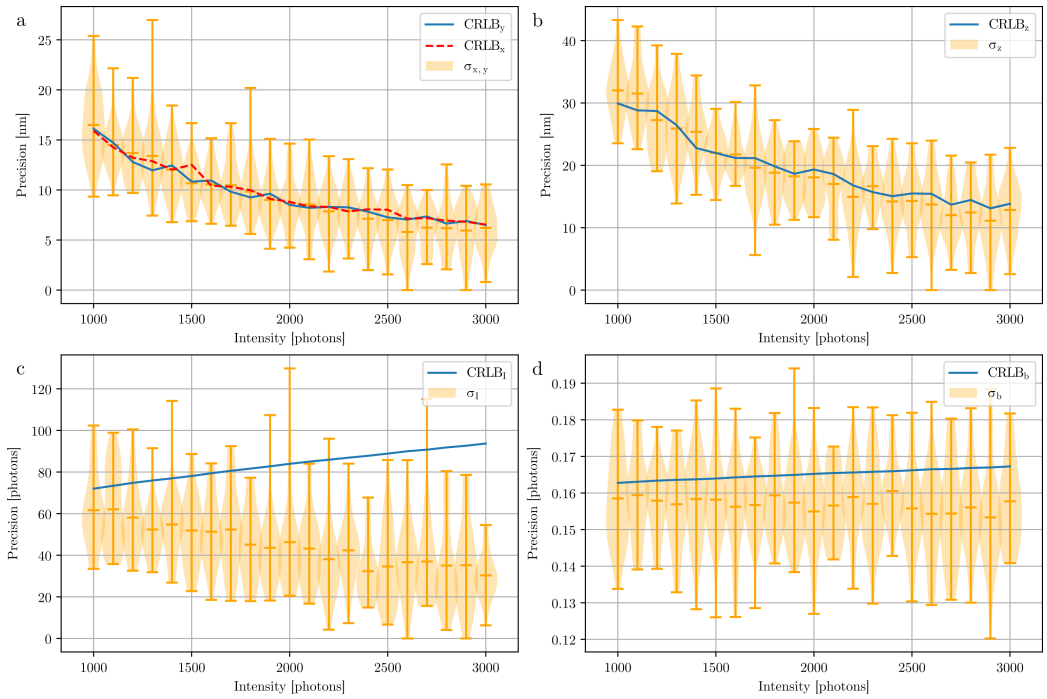
Supplementary Figure 7. Model and parameter space convergence of the algorithm for low SBR frames. **a)** Autocorrelation of the model chain. **b)** Autocorrelation of the parameters. **c)** Time series plot of the model for all frames. **d)** Histogram of all visited models across all frames. Emitter intensities were sampled from $\mathcal{N}(500, 150)$, background was set to 20 photons. Emitters were placed in focal plane at the center of the ROI and given a sub-pixel shift drawn from $\mathcal{U}(-0.5, 0.5)$. 100 frames were used.



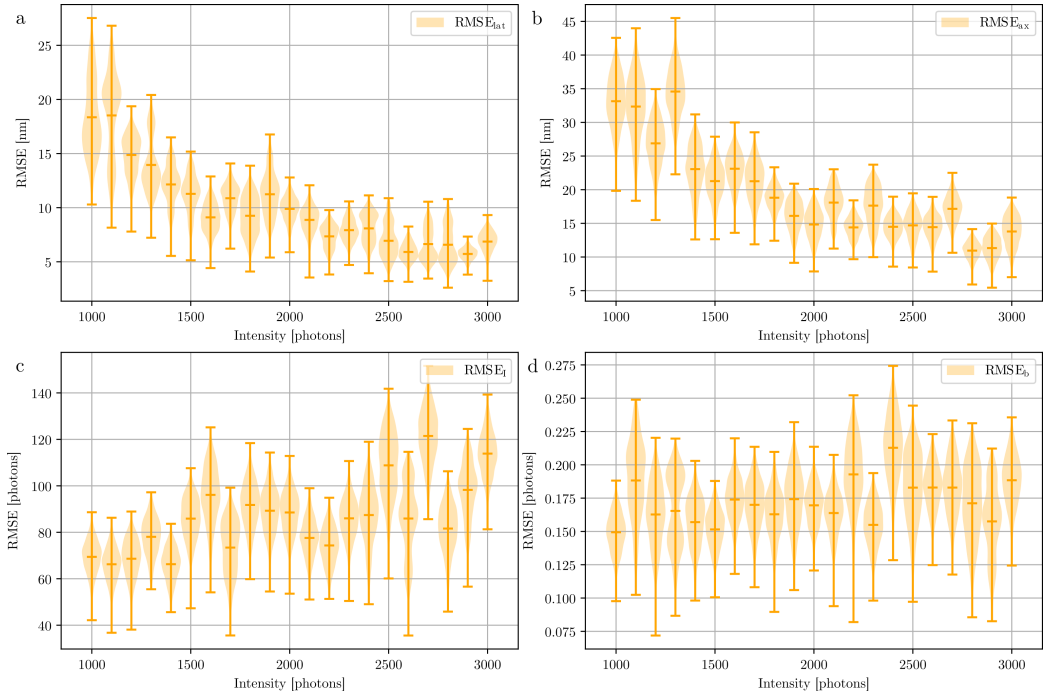
Supplementary Figure 8. Violinplots of parameter estimates over intensity using astigmatic imaging, compared to the CRLB. **a)** Lateral precision violinplot, using an effective pixel size of 100 nm. **b)** Axial precision violinplot. **c)** Violinplot of the emitter intensity precision. **d)** Violinplot of the background intensity precision. Precisions were calculated from the MCMC chain of individual frames, assuming a Gaussian distribution. With a background of 20 photons, 50 frames were used per datapoint from 500 to 3000 emitter intensity photons. Emitters were placed in focal plane at the center of the ROI and given a sub-pixel shift drawn from $\mathscr{U}(-0.5, 0.5)$.



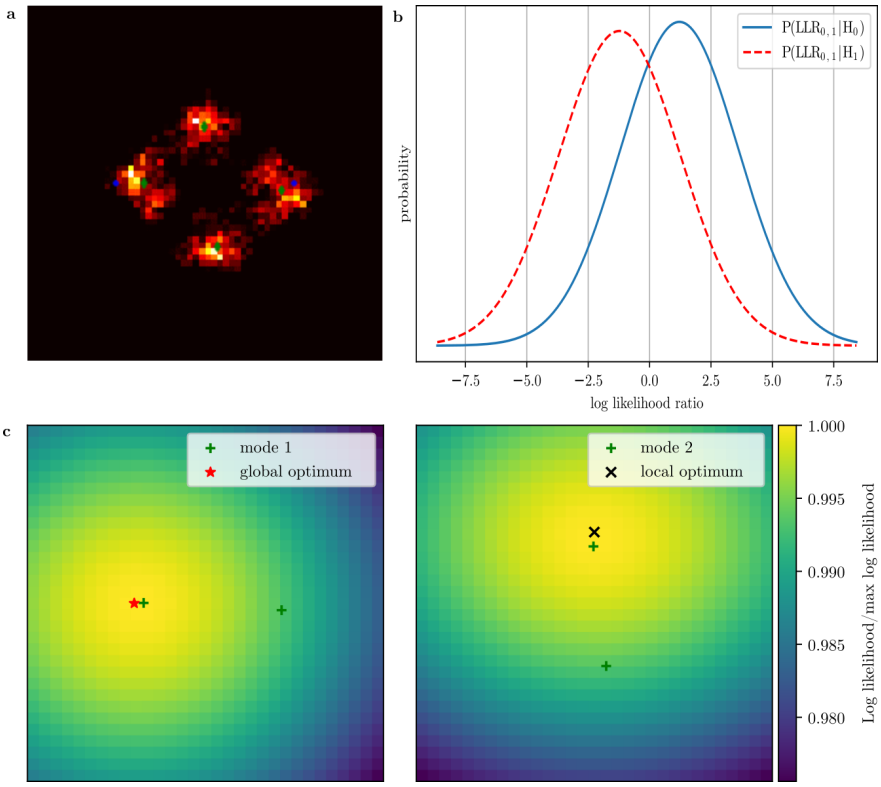
Supplementary Figure 9. Bootstrap resampled root mean square error (RMSE) results for the parameter estimates plotted over intensity using astigmatic imaging. **a)** Violinplot of the lateral RMSE. **b)** Violinplot of the axial RMSE. **c)** Violinplot of the emitter intensity RMSE. **d)** Violinplot of the background intensity RMSE. Results were resampled using 50% of the available data and 1000 runs. With a background of 20 photons, 50 frames were used per datapoint from 500 to 3000 emitter intensity photons. Emitters were placed in focal plane at the center of the ROI and given a sub-pixel shift drawn from $\mathscr{U}(-0.5, 0.5)$.



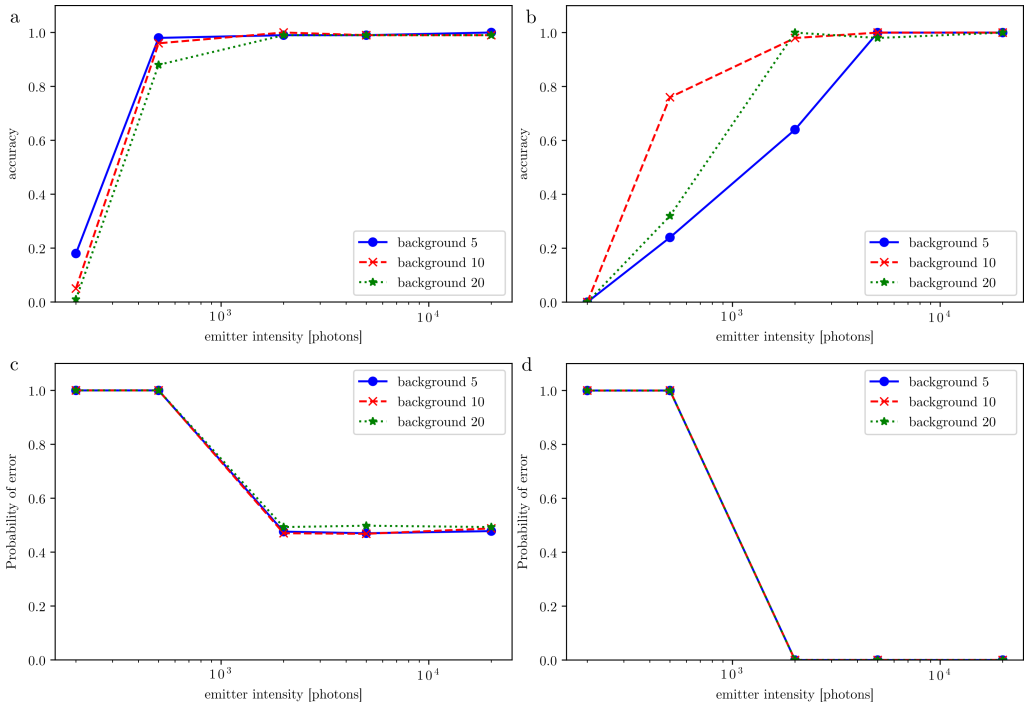
Supplementary Figure 10. Violinplots of the parameter estimates over intensity using a tetrapod PSF, compared to the CRLB. **a)** Lateral precision violinplot, using an effective pixel size of 100 nm. **b)** Axial precision violinplot. **c)** Violinplot of the emitter intensity precision. **d)** Violinplot of the background intensity precision. Precisions were calculated from the MCMC chain of individual frames, assuming a Gaussian distribution. With a background of 20 photons, 50 frames were used per datapoint from 500 to 3000 emitter intensity photons. Emitters were placed in focal plane at the center of the ROI and given a sub-pixel shift drawn from $\mathscr{U}(-0.5, 0.5)$.



Supplementary Figure 11. Bootstrap resampled RMSE results for the parameter estimates using a tetrapod PSF, plotted over intensity. **a)** Violinplot of the lateral RMSE. **b)** Violinplot of the axial RMSE. **c)** Violinplot of the emitter intensity RMSE. **d)** Violinplot of the background intensity RMSE. Results were resampled using 50% of the available data and 1000 runs. With a background of 20 photons, 50 frames were used per datapoint from 500 to 3000 emitter intensity photons. Emitters were placed in focal plane at the center of the ROI and given a sub-pixel shift drawn from $\mathscr{U}(-0.5, 0.5)$.



Supplementary Figure 12. Analysis of multimodality in astigmatic PSF 3D reconstructions for the two emitter problem. Ground truth positions are marked with blue dots, while localizations are marked with green crosses. **a)** Single frame reconstruction for the two emitter problem. **b)** Plots of the log likelihood ratios between the hypothesised modes of the reconstruction in a), using a Gaussian approximation. **c)** Left: fixing the rightmost localization, an emitter is shifted over the ROI, imaging the log likelihood and marking the maximum (red star). Right: the bottom emitter is fixed, and a local optimum is found in the black cross. This finds the pair of modes used in b), one on the ground truth and one perpendicular to it. Calculating the probability of error yields a 49.8% chance of selecting the wrong hypothesis. The expected value of either mode passes a chi-squared test on a 95% confidence interval.



Supplementary Figure 13. Model accuracy and probability of error between modes for the two emitter problem using astigmatic and biplane imaging. Plotted for varying emitter and background intensities. **a)** Model accuracy for astigmatic imaging of the two emitters. **b)** Model accuracy for biplane imaging of the two emitters. **c)** Probability of error when selecting modes with astigmatic imaging. **d)** Probability of error when selecting modes using biplane imaging. Emitters were placed in focal plane, separated by 5/4 σ_{PSF} along the x-axis. All uniform priors except for emitter intensity, which used $\mathcal{N}(I_{sample}, 3.5\sqrt{I_{sample}})$. Model accuracy was calculated by dividing the number of frames where the correct model was found with the total number of frames. Probability of error was calculated using the same methods as for Supplementary Figure 12 described in the Supplementary Note on multimodality in two emitter astigmatic imaging. Reconstructions without alternate mode used a probability of error of 0, while reconstructions that could not separate emitters used a probability of error of 1.

Supplementary Table 1. Move proposal probabilities

move type	probability		
	RJMCMC_{burn}	RJMCMC	MCMC
Single Emitter	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{2}{5}$
Group Move	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{2}{5}$
Background	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{1}{5}$
Split	$\frac{1}{15}$	0	0
Merge	$\frac{1}{15}$	0	0
Generalized Split	$\frac{1}{15}$	$\frac{3}{32}$	0
Generalized Merge	$\frac{1}{15}$	$\frac{3}{32}$	0
Birth	$\frac{1}{15}$	$\frac{1}{32}$	0
Death	$\frac{1}{15}$	$\frac{1}{32}$	0

Supplementary Table 2. RJMCMC hyperparameters for astigmatic PSF, single emitter convergence test

	RJMCMC	MCMC	units
σ_x	0.05	0.05	[pixels]
σ_y	0.05	0.05	[pixels]
σ_z	0.08	0.07	$[\mu m]$
σ_I	18	15	[photons]
σ_b	1	1	[photons]
σ_{PSF}	1.3	1.3	[pixels]

Supplementary Table 3. RJMCMC hyperparameters for tetrapod PSF, single emitter convergence test

	RJMCMC	MCMC	units
σ_x	0.1	0.1	[pixels]
σ_y	0.1	0.1	[pixels]
σ_z	0.08	0.08	$[\mu m]$
σ_I	30	30	[photons]
σ_b	1	1	[photons]
σ_{PSF}	2.1	2.1	[pixels]

Supplementary Table 4. RJMCMC hyperparameters for astigmatic PSF, single emitter precision test

	RJMCMC	MCMC	units
σ_x	0.08	0.08	[pixels]
σ_y	0.08	0.08	[pixels]
σ_z	0.08	0.08	[μm]
σ_I	30	30	[photons]
σ_b	1	1	[photons]
σ_{PSF}	1.3	1.3	[pixels]

Supplementary Note

3D Reversible Jump Markov Chain Monte Carlo localization algorithm

To both estimate the model and find the parameters, Reversible Jump Markov Chain Monte Carlo (RJMCMC) is employed. The RJMCMC algorithm makes jumps in model space, making it possible to recover a distribution of possible model and parameter pairings. For 3D localization, the model only varies in number of emitters active per frame, in which we try to estimate the parameters of the emitters, x, y, and z position as well as intensity I, and the background photons present in the frame, b. With the use of smart priors, 3D RJMCMC localization can retrieve the correct model and localize the emitters.

Localization in 3D using RJMCMC is done as follows:

1. an emitter is initialized randomly
2. RJMCMC is run with the burn-in move probabilities
3. the move probabilities are adjusted for the fraction of the chain that follows the burn-in
4. from the post burn-in RJMCMC chain, the MAP model is selected from the distribution to be used with MCMC
5. MCMC is initialized using the last element from the RJMCMC chain in which the model corresponds to the MAP model
6. MCMC is run with the parameter space move list
7. k-means clustering is used to perform label switching and retrieve the localizations from the chains, setting the cluster count to the MAP model

The RJMCMC and MCMC runs undergo the following steps internally:

1. a move is randomly selected from the move list
2. the move is executed on each frame, generating new parameters and models (if the move is not valid for the frame, the parameters stay fixed)
3. the acceptance rate is calculated and the jump is either accepted or rejected, updating the chain
4. steps 1 to 3 are repeated until the desired chain length is achieved

Priors

The algorithm takes in priors for emitter position, emitter intensity, background intensity, and number of emitters present. Though the prior for number of emitters can be set as a Poisson distribution if the user has knowledge of the labelling density, as the local density when evaluating small ROIs can differ greatly, $P(k)$ is kept uniform:

$$P(k) \sim \mathscr{U}(0, k_{max}) \tag{1}$$

with k_{max} the user defined maximum number of emitters expected in the ROI. For emitter position, the priors are set to be uniform to avoid bias:

$$P(\theta_{x,i}|k) \sim \mathscr{U}(-w_{border}, x_{ROI} + w_{border}) \tag{2}$$
$$P(\theta_{y,i}|k) \sim \mathscr{U}(-w_{border}, y_{ROI} + w_{border}) \tag{3}$$
$$P(\theta_{z,i}|k) \sim \mathscr{U}(z_{min}, z_{max}) \tag{4}$$

with $\theta_{x,i}$, $\theta_{y,i}$, $\theta_{z,i}$ the 3D position of the i^{th} found emitter, x_{ROI}, y_{ROI} the x and y dimensions of the ROI, w_{border} some additional pixels outside the ROI to account for influence of external emitters, and z_{min}, z_{max} the admissible range of z positions.

The prior on background intensity is kept as a simple uniform distribution:

$$P(\theta_b) \sim \mathscr{U}(b_{min}, b_{max}) \tag{5}$$

with b_{min}, b_{max} user defined minimum and maximum expected background values for the experiment. Finally, the prior for intensity should be either determined experimentally by kernel density fitting SMLM results, or directly defined by the user given sufficient knowledge of the emission response. For the testing on synthetic data, often a Gaussian prior around the expected emitter intensity is used, with an additional flat shoulder at lower intensities to facilitate model jumps if necessary, such as:

$$P(\theta_{I,i}|k) = \begin{cases} \frac{1}{150\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\theta_I - 2000}{150}\right)^2\right], & \text{if } 1600 < \theta_I < 2500 \\ 6 \cdot 10^{-5}, & \text{if } 500 < \theta_I < 1600 \\ 6 \cdot 10^{-5} \frac{\theta_I}{500}, & \text{if } 1 < \theta_I < 500 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

The prior on intensity is assumed to be unchanging over depth.

Moves

In RJMCMC, the design of the moves is crucial for exploring the model space. The moves used in 3D RJMCMC localization are:

- a single emitter move, varying the 3D position and intensity of a single emitter
- a group move, varying the 3D position and intensity of a cluster of emitters
- a background move, changing the offset background estimate
- a split move, splitting one emitter into two
- a merge move, merging two emitters into one
- a generalized split move, splitting a cluster of N emitters into $N+1$
- a generalized merge move, merging a cluster of N emitters into $N-1$
- a birth move, spawning an emitter where the difference between measured and expected intensity is high
- a death move, removing an emitter at random

The single emitter, group, and background move work in parameter space while leaving the model space identical. The rest of the moves come in pairs to retain detailed balance and change both the model and the parameters of the estimate. Supplementary Figure 1 schematically shows how each move changes the parameters or model.

move proposal At every iteration of the algorithm, a move is selected according to fixed probabilities defined by the user. Move probabilities can be set separately for the MCMC portion, as well as both the burn-in phase and the post burn-in phase of the RJMCMC portion. For testing on synthetic data, the move probabilities in Supplementary Table 1 were found to work well.

Single emitter move

These first three moves will act on the parameter space only and their acceptance rates can be determined from the Metropolis-Hastings algorithm for the MCMC case.

Metropolis-hastings acceptance rate for parameter space jumps The probability of accepting parameter space jumps is given as:

$$\alpha = \min\left[1, \frac{P(\theta', k, D)}{P(\theta, k, D)} \left| \frac{\partial(\theta')}{\partial(\theta, u)} \right| \right] \tag{7}$$

with α the acceptance rate, using the joint probability $P(\theta, k, D) = P(D|\theta, k)P(\theta|k)P(k)$ to calculate the posterior ratio (ignoring evidence term $P(D)$ as it gets cancelled in the division) and the final fraction as the determinant of the Jacobian of function $\theta'(\theta, u)$ that generates the new parameters. Here $\theta = \left[\theta_{x,0} \quad \theta_{y,0} \quad \theta_{z,0} \quad \theta_{I,0} \quad \dots \quad \theta_{x,i} \quad \theta_{y,i} \quad \theta_{z,i} \quad \theta_{I,i} \quad \theta_b\right]$ is the parameter vector, containing the positions and intensities of i emitters used in the current model along with background

parameter θ_b. Using a random walk sampler for the parameter space moves and cancelling the model prior, the acceptance rate reduces to:

$$\alpha = \min\left[1, \frac{P(\theta', k, D)}{P(\theta, k, D)}\right] = \min\left[1, \frac{P(D|\theta', k)P(\theta'|k)}{P(D|\theta, k)P(\theta|k)}\right] \tag{8}$$

The single emitter move randomly selects one of the i current emitters and generates a new estimate out of its previous parameters as follows:

$$\theta'_{x,j} = \theta_{x,j} + u_x \tag{9}$$
$$\theta'_{y,j} = \theta_{y,j} + u_y \tag{10}$$
$$\theta'_{z,j} = \theta_{z,j} + u_z \tag{11}$$
$$\theta'_{I,j} = \theta_{I,j} + u_I \tag{12}$$

with u drawn from the following normal distributions:

$$P(u_x) \sim \mathcal{N}(0, \sigma_x) \tag{13}$$
$$P(u_y) \sim \mathcal{N}(0, \sigma_y) \tag{14}$$
$$P(u_z) \sim \mathcal{N}(0, \sigma_z) \tag{15}$$
$$P(u_I) \sim \mathcal{N}(0, \sigma_I) \tag{16}$$

and σ_x, σ_y, σ_z, σ_I some user determined hyperparameters to vary the jump size. The posterior ratio in 8 can now be simplified to:

$$\frac{P(D|\theta', k)P(\theta'|k)}{P(D|\theta, k)P(\theta|k)} = \left[\prod_{k_p} \exp\left(\mu_{k_p} - \mu'_{k_p}\right)\left(\frac{\mu'_{k_p}}{\mu_{k_p}}\right)^{n_{k_p}}\left(\prod_{k}^{k_{max}} \frac{P(\theta'_{I,k})}{P(\theta_{I,k})}\right)\right] \tag{17}$$

Assuming the estimated parameters stay within the bounds of the priors, the uniform position priors cancel out and only the likelihood ratio and intensity prior ratio are left.

Group move

The group move first finds a cluster of nearby emitters, and then executes a single emitter move on each of them. It randomly picks one emitter, then searches for emitters within a 2 σ_{PSF} sphere around it to form a cluster. This radius can be treated as a hyper-parameter and tuned, though for Gaussian and astigmatic PSFs it works well when set to the PSF width in focal plane. More complex 3D PSFs may require a fairly large radius for clustering. After finding an eligible cluster and executing the moves, the acceptance rate can be calculated using the same equations as in the single move, namely 8 and 17. If a cluster cannot be found, the move fails.

Background move

The background move is not coupled to any emitter. It updates a flat offset background intensity with a random walk sampler as follows:

$$\theta'_b = \theta_b + u_b \qquad\qquad P(u_b) = \mathcal{N}(0, \sigma_b) \tag{18}$$

with σ_b another hyper-parameter for tuning the background jump size. The acceptance rate from equation 17 can now be reduced even further, assuming the background jump stays within the bounds of its uniform prior:

$$\frac{P(D|\theta', k)P(\theta'|k)}{P(D|\theta, k)P(\theta|k)} = \left[\prod_{k_p} \exp\left(\mu_{k_p} - \mu'_{k_p}\right)\left(\frac{\mu'_{k_p}}{\mu_{k_p}}\right)^{n_{k_p}}\right] \tag{19}$$

which is just the likelihood ratio of the parameters before and after the jump.

Split and Merge pair

The next 6 moves come in pairs and may act on both the parameter and the model space. Their acceptance rates can be determined from the general acceptance rate used in RJMCMC.

Metropolis-Hastings acceptance rate for model space jumps The probability of accepting model space jumps in the general case is given as:

$$\alpha = \min\left\{1, \frac{P(\theta',k',D)r_m(\theta')}{P(\theta,k,D)r_m(\theta)q(u)}\left|\frac{\partial(\theta')}{\partial(\theta,u)}\right|\right\} \tag{20}$$

with $r_m(\theta)$ the probability of choosing the move type (selecting either split or merge) and $q(u)$ the probability density function of the draws from u used to generate the split weights, means, and standard deviations. The split and merge move turn one emitter into two and vice versa. By using this move pair, the algorithm has a move that allows it to distinguish emitters that are in close proximity to one another. Split and merge will uphold the following constraints:

$$\theta_{I,j^*} = \theta_{I,j_1} + \theta_{I,j_2} \tag{21}$$
$$\theta_{I,j^*}\theta_{x,j^*} = \theta_{I,j_1}\theta_{x,j_1} + \theta_{I,j_2}\theta_{x,j_2} \tag{22}$$
$$\theta_{I,j^*}\theta_{y,j^*} = \theta_{I,j_1}\theta_{y,j_1} + \theta_{I,j_2}\theta_{y,j_2} \tag{23}$$
$$\theta_{I,j^*}\theta_{z,j^*} = \theta_{I,j_1}\theta_{z,j_1} + \theta_{I,j_2}\theta_{z,j_2} \tag{24}$$
$$\tag{25}$$

Split For a split, θ' is generated from a randomly selected emitter with index j^* as follows:

$$u_1 \sim \mathscr{U}(0,1) \tag{26}$$
$$u_2 \sim \mathscr{N}(0,\sigma_{PSF}) \tag{27}$$
$$u_3 \sim \mathscr{N}(0,\sigma_{PSF}) \tag{28}$$
$$u_4 \sim \mathscr{N}(0,\sigma_{PSF}) \tag{29}$$
$$\theta_{I,j_1} = \theta_{I,j^*}u_1 \tag{30}$$
$$\theta_{I,j_2} = \theta_{I,j^*}(1-u_1) \tag{31}$$
$$\theta_{x,j_1} = \theta_{x,j^*} + u_2 \tag{32}$$
$$\theta_{y,j_1} = \theta_{y,j^*} + u_3 \tag{33}$$
$$\theta_{z,j_1} = \theta_{z,j^*} + u_4 \tag{34}$$
$$\theta_{x,j^*}\theta_{I,j^*} = \theta_{x,j_1}\theta_{I,j_1} + \theta_{x,j_2}\theta_{I,j_2} \tag{35}$$
$$\theta_{x,j_2} = \theta_{x,j^*} - \frac{u_1u_2}{1-u_1} \tag{36}$$
$$\theta_{y,j^*}\theta_{I,j^*} = \theta_{y,j_1}\theta_{I,j_1} + \theta_{y,j_2}\theta_{I,j_2} \tag{37}$$
$$\theta_{y,j_2} = \theta_{y,j^*} - \frac{u_1u_3}{1-u_1} \tag{38}$$
$$\theta_{z,j^*}\theta_{I,j^*} = \theta_{z,j_1}\theta_{I,j_1} + \theta_{z,j_2}\theta_{I,j_2} \tag{39}$$
$$\theta_{z,j_2} = \theta_{z,j^*} - \frac{u_1u_4}{1-u_1} \tag{40}$$

with the two new emitters retaining the 3D center of mass and total intensity of the emitter they split from. Indices j_1 and j_2 are used to indicate the emitters resulting from a split, or those used in a merge, whereas index j^* is used for the result of a merge or the target of a split.

The ratio of selecting the move types then becomes:

$$\frac{r_m(\theta')}{r_m(\theta)} = \frac{P_{split}}{P_{merge}} \tag{41}$$

and $q(u)$ is:

$$q(u) = P(u_1)P(u_2)P(u_3)P(u_4) \tag{42}$$

Knowing the split move only influences these parameters and leaves the other emitters and background untouched, the emitter to be split and the resulting emitters after the split can be moved to the back for both θ' and (θ, u) vectors. The resulting vectors then become:

$$\theta' = \begin{bmatrix} \dots & \theta_{I,j_1} & \theta_{x,j_1} & \theta_{y,j_1} & \theta_{z,j_1} & \theta_{I,j_2} & \theta_{x,j_2} & \theta_{y,j_2} & \theta_{z,j_2} \end{bmatrix} \tag{43}$$

$$(\theta, u) = \begin{bmatrix} \dots & \theta_{I,j^*} & u_1 & \theta_{x,j^*} & u_2 & \theta_{y,j^*} & u_3 & \theta_{z,j^*} & u_4 \end{bmatrix} \tag{44}$$

the Jacobian can then be constructed as:

$$\frac{\partial(\theta')}{\partial(\theta,u)} = \begin{bmatrix} I_{k-1} & \mathbf{0}_{k-1,1} & & & & \dots & & & \mathbf{0}_{k-1,1} \\ \mathbf{0}_{1,k-1} & u_1 & 0 & 0 & 0 & -\theta_{I,j^*} & 0 & 0 & 0 \\ & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ \vdots & 1-u_1 & 0 & 0 & 0 & -\theta_{I,j^*} & 0 & 0 & 0 \\ & 0 & \frac{-u_2}{(1-u_1)^2} & 1 & \frac{-u_1}{1-u_1} & 0 & 0 & 0 & 0 \\ & 0 & \frac{-u_3}{(1-u_1)^2} & 0 & 0 & 1 & \frac{-u_1}{1-u_1} & 0 & 0 \\ \mathbf{0}_{1,k-1} & 0 & \frac{-u_4}{(1-u_1)^2} & 0 & 0 & 0 & 0 & 1 & \frac{-u_1}{1-u_1} \end{bmatrix}$$

with identity matrix I_{k-1} of dimension $k-1$ for all the unchanged parameters and $\mathbf{0}_{k-1,1}$, $\mathbf{0}_{1,k-1}$ column and row vectors containing zeros. This yields the determinant:

$$\left| \frac{\partial(\theta')}{\partial(\theta,u)} \right| = \frac{\theta_{I,j^*}}{(1-u_1)^3} \tag{45}$$

Setting $P_{split} = P_{merge}$, the final result is then:

$$\alpha = \min\left\{1, A_{split}\right\} \tag{46}$$

$$A_{split} = \frac{P(\theta'|D)r_m(\theta')}{P(\theta|D)r_m(\theta)q(u)} \left| \frac{\partial(\theta')}{\partial(\theta,u)} \right| = \frac{P(\theta',k',D)}{P(\theta,k,D)} \frac{\theta_{I,j^*}}{P(u_1,u_2,u_3,u_4)(1-u_1)^3} \tag{47}$$

Merge For a merge move, an emitter is selected at random, after which it is randomly paired with another emitter within 2 σ_{PSF} of itself. If no emitter can be found within this distance, the move fails. The emitters combine their intensity and center of mass. The acceptance rate can be set to the inverse of that for the split move, finding u_1, u_2, u_3 and u_4 deterministically from equations 30, 32, 33, and 34:

$$u_1 = \frac{\theta_{I,j_1}}{\theta_{I,j^*}} \tag{48}$$

$$u_2 = \theta_{x,j_1} - \theta_{x,j^*} \tag{49}$$

$$u_3 = \theta_{y,j_1} - \theta_{y,j^*} \tag{50}$$

$$u_4 = \theta_{z,j_1} - \theta_{z,j^*} \tag{51}$$

This gives the the acceptance rate:

$$\alpha = \min\left\{1, A_{merge}\right\} \tag{52}$$

$$A_{merge} = A_{split}^{-1} = \frac{P(\theta,k,D)}{P(\theta',k',D)} \frac{P(u_1,u_2,u_3,u_4)(1-u_1)^3}{\theta_{I,j^*}} \tag{53}$$

with θ previously being the parameters before the split and now thus the parameters after the merge and vice versa for θ'.

Generalized Split and Merge

Whereas split and merge work on a single emitter, their generalized versions are helpful for very densely packed frames, allowing the splitting and merging of emitters within clusters of N emitters. Generalized split jumps from N to $N+1$ emitters, while generalized merge does the opposite. These moves use the same clustering as was used for the group move, picking one random emitter and forming an elligible cluster within 2 σ_{PSF} of itself.

Generalized Split Now, instead of distributing the intensity of one emitter to two, the combined intensity of N emitters is distributed over $N+1$, again retaining the center of mass:

$$\sum_{j=1}^{N} \theta_{I,j} = \sum_{j=1}^{N+1} \theta'_{I,j} \tag{54}$$

$$\sum_{j=1}^{N} \theta_{I,j}\theta_{x,j} = \sum_{j=1}^{N+1} \theta'_{I,j}\theta'_{x,j} \tag{55}$$

$$\sum_{j=1}^{N} \theta_{I,j}\theta_{y,j} = \sum_{j=1}^{N+1} \theta'_{I,j}\theta'_{y,j} \tag{56}$$

$$\sum_{j=1}^{N} \theta_{I,j}\theta_{z,j} = \sum_{j=1}^{N+1} \theta'_{I,j}\theta'_{z,j} \tag{57}$$

Putting up some additional constraints to remove the same fraction of intensity for each emitter in the cluster and let the new emitter be formed close to the center of mass of the cluster:

$$\theta'_{I,j} = (1-u_1)\theta_{I,j} \tag{58}$$

$$\theta'_{x,j} = \frac{\theta_{x,j} - u_1\left(\frac{1}{N}\sum_{j=1}^{N}\theta_{x,j} + u_2\right)}{1-u_1} \tag{59}$$

$$\theta'_{y,j} = \frac{\theta_{x,j} - u_1\left(\frac{1}{N}\sum_{j=1}^{N}\theta_{x,j} + u_3\right)}{1-u_1} \tag{60}$$

$$\theta'_{z,j} = \frac{\theta_{y,j} - u_1\left(\frac{1}{N}\sum_{j=1}^{N}\theta_{z,j} + u_4\right)}{1-u_1} \tag{61}$$

The new emitter is then generated as follows:

$$\theta'_{I,N+1} = u_1 \sum_{j=1}^{N} \theta_{I,j} \tag{62}$$

$$\theta'_{x,N+1} = \frac{1}{N}\sum_{j=1}^{N}\theta_{x,j} + u_2 \tag{63}$$

$$\theta'_{y,N+1} = \frac{1}{N}\sum_{j=1}^{N}\theta_{y,j} + u_3 \tag{64}$$

$$\theta'_{z,N+1} = \frac{1}{N}\sum_{j=1}^{N}\theta_{z,j} + u_4 \tag{65}$$

Applying the general acceptance rate from equation 20, the result is similar to equation 47 except the jacobian now expands to:

$$\left|\frac{\partial(\theta')}{\partial(\theta,u)}\right| = \frac{\sum_{j=1}^{N}\theta_{I,j}}{(1-u_1)^{2N+1}} \tag{66}$$

the acceptance rate then becomes:

$$A_{g-split} = \frac{P(\theta',k',D)}{P(\theta,k,D)} \frac{\sum_{j=1}^{N} \theta_{I,j}}{P(u_1,u_2,u_3,u_4)((1-u_1)^{2N+1}} \tag{67}$$

Generalized Merge The generalized merge selects an emitter and then distributes its intensity evenly over the N emitters in the corresponding cluster. As with the merge, the randomly sampled parameters can now be found deterministically:

$$u_1 = \frac{\theta'_{I,N+1}}{\sum_{j=1}^{N} \theta_j} \tag{68}$$

$$u_2 = \theta'_{x,N+1} - \frac{1}{N} \sum_{j=1}^{N} \theta_{x,j} \tag{69}$$

$$u_3 = \theta'_{y,N+1} - \frac{1}{N} \sum_{j=1}^{N} \theta_{y,j} \tag{70}$$

$$u_4 = \theta'_{z,N+1} - \frac{1}{N} \sum_{j=1}^{N} \theta_{z,j} \tag{71}$$

and the new parameters become:

$$\theta_{I,j} = \frac{\theta'_{I,j}}{1-u_1} \tag{72}$$

$$\theta_{x,j} = u_1 \theta'_{x,N+1} + (1-u_1)\theta'_{x,j} \tag{73}$$

$$\theta_{y,j} = u_1 \theta'_{x,N+1} + (1-u_1)\theta'_{y,j} \tag{74}$$

$$\theta_{z,j} = u_1 \theta'_{x,N+1} + (1-u_1)\theta'_{z,j} \tag{75}$$

$$\tag{76}$$

Inverting the acceptance rate of the generalized split yields:

$$A_{g-merge} = \frac{P(\theta,k,D)}{P(\theta',k',D)} \frac{P(u_1,u_2,u_3,u_4)((1-u_1)^{2N+1}}{\sum_{j=1}^{N} \theta_{I,j}} \tag{77}$$

Birth and Death pair

The birth and death pair can increase and decrease the number of emitters independent of the positions of the currently found emitters. Birth will generate an emitter anywhere in the allowed by the position priors, using the residual of the frame and the expected image as the probability distribution for selecting the new lateral emitter position. Death randomly chooses any of the current emitters and removes it.

Birth The first task of the birth move is finding probable locations for yet unmodelled emitters. It calculates the expected image from the current model and parameters and subtracts this from the ROI. Any negative values are fixed to zero, and the values are normalized to give a per pixel probability of an undetected emitter:

$$P(k_{new,k_p}) = \frac{n_{k_p} - \mu_{k_p}}{\sum_{k_p=1}^{N_p} n_{k_p} - \mu_{k_p}}, \qquad \{k_p : n_{k_p} - \mu_{k_p} \geq 0\} \tag{78}$$

$$\tag{79}$$

with $P(k_{new,k_p})$ the pixel probability of a yet undiscovered emitter residing in pixel p_k. After randomly selecting a pixel, a subpixel position is generated from a uniform distribution and the z position is uniformly sampled from the entire depth range. The intensity of the new emitter is sampled directly from the intensity prior distribution.

This sampling of the z position may result in a lot of ineffective proposals, however there are little effective alternatives to effectively finding the z position that do not also impact calculation times or convergence. Some tests have been run where the residual image was filtered with the 3D PSF to return a probability distribution for pairs of x, y, and z positions. It unfortunately resulted in some chains getting stuck, while only marginally improving convergence speed overall.

Fixing the birth and death selection probabilities to be equal, just as for the (generalized) split and merge moves, yields the acceptance rate:

$$A_{birth} = \frac{P(\theta', k', D)}{P(\theta, k, D)} \frac{z_{max} - z_{min}}{P(k^*_{new,k_p})} \tag{80}$$

with $P(k^*_{new,k_p})$ the probability of sampling from pixel k^*_p and z_{max}, z_{min} the depth range from which the z position is sampled. Note that since birth and death work independently from the current emitters, the determinant of the Jacobian $\partial\theta'/\partial(\theta, u)$is simply one.

Death Death is the most straightforward move of them all, simply removing one emitter from the model at random. This move helps remove unnecessary emitters that may have a low intensity or contribute little to a cluster of emitters. It also helps remove emitters spawned by birth that do not necessarily fit the data, but went through despite a low acceptance rate. The acceptance rate for death is:

$$A_{death} = \frac{P(\theta', k', D)}{P(\theta, k, D)} = \frac{P(\theta'|D)}{P(\theta|D)} \tag{81}$$

where the second simplification holds as long as the model prior is kept uniform.

Single emitter localization convergence for an astigmatic PSF

To verify if the algorithm works correctly, it is first tested on frames where only a single emitter is active. Synthetic data is generated for an emitter in focal plane of a 20 by 20 pixel ROI, assuming an effective pixel size of 100 nm. Emitters is placed in the center of the ROI, generating 50 frames, each with a random subpixel shift in x and y drawn from a uniform distribution and intensities drawn from a normal distribution:

$$\theta_x \sim \mathscr{U}(10-.5, 10+.5) \tag{82}$$
$$\theta_y \sim \mathscr{U}(10-.5, 10+.5) \tag{83}$$
$$\theta_I \sim \mathscr{N}(I_{sim}, 150) \tag{84}$$

with the position in pixels, and I_{sim} the mean intensity for the synthetic dataset.The background photon count is fixed at 20 photons. The algorithm is tested on data with $I_{sim} = 2000$ and $I_{sim} = 500$ to demonstrate convergence at varying SBR.

The hyperparameters used for localization are as follows. The RJMCMC portion ran for $20,000$ burn-in iterations, followed by another $10,000$ iterations from which the model is determined. MCMC is then run for another $5,000$ iterations, from which the final positions of the emitters are determined. The parameters for the random walk samplers within the RJMCMC and MCMC portions are detailed in Supplementary Table 2, while the move probabilities were taken from Supplementary Table 1. The depth was constrained to a range of $\begin{bmatrix}-1 & 1\end{bmatrix}$ μm.

The priors for position and background are uniform, with a custom prior used for intensity. The number of emitters was constrained to $k_{max} = 6$. The algorithm was initialized randomly with k_{max} active emitters drawn from the priors, ensuring the initial model is incorrect and allowing for meaningful conclusions from the model time series and autocorrelation.

High SBR results

For $I_{sim} = 2000$, the prior on intensity is given as follows:

$$P(I) = \begin{cases} 0, & \text{if } I > 2500 \\ \frac{1}{150\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left(\frac{I-2000}{150}\right)^2 \right\} / 1.08, & \text{if } 1600 < I < 2500 \\ 6.6 \cdot 10^{-5}, & \text{if } 500 < I < 1600 \\ \left(\frac{I}{500}\right) \cdot 6.6 \cdot 10^{-5}, & \text{if } I > 500 \end{cases} \tag{85}$$

The prior is plotted in Supplementary Figure 2.

model convergence To verify the model has converged, the time series of the models over all frames as well as the average autocorrelation of the model is studied. In Supplementary Figure 3, the model time series and autocorrelation are plotted. From the time series, we can see that from the initialization at $k_{max} = 6$ emitters, the estimated model rapidly reduces to a two or one emitter model. There is notably little mixing in the model chain. This may be due to the simplicity of localizing a single emitter, the algorithm therefore having infinitesimal probabilities of jumping to models with 3 emitters or more. Note that after the burn-in fraction, the split and merge move are turned off and there is no more change in the model found. With split and merge disabled, and G-split and G-merge failing since no clusters can be made out of one emitter, the only moves capable of changing the model are birth and death. At this point, the RJMCMC seems to have converged strongly. Evaluating the average model chain autocorrelation, it is clear that after 2000 RJMCMC iterations, the autocorrelation has gone to zero. This gives an idea of the chain length required to estimate the model distribution under these conditions. Some multiple of 2000, say 6000 RJMCMC iterations, may already be sufficient for model convergence on this problem. From the histogram in Supplementary Figure 3 d) it is clear that on average, the estimates spend nearly all their time in a model with one active emitter.

parameter convergence Again, we look to the autocorrelation of the parameters to assess convergence. In Supplementary Figure 3 b), the average autocorrelations of all parameters for their MCMC chains are plotted. Note that the MCMC chain was initalized from the last RJMCMC iteration that had the correct model, therefore the parameters can be expected to converge quickly and there is no need to discard the initial part of the chain. The autocorrelations of the lateral position and background decrease almost immediately to zero. Note that while the autocorrelation for axial position drops to zero in the same time, there are some slower transients visible in the plot. The same holds for the autocorrelation of intensity, which also takes significantly longer to reach zero. The acceptance rates for the moves are $\begin{bmatrix}0 & 0.47 \pm 0.02 & 0.28 \pm 0.01\end{bmatrix}$ for the group move, single emitter move, and background move, respectively. The group move cannot find a cluster and is therefore never executed, meanwhile the single emitter move has a good acceptance rate ensuring proper mixing. The background move has low acceptance, though it should still be enough for proper mixing.

Calculating the lateral and axial RMSE results in 0.05 pixels and 120 nm errors, respectively, giving us an accurate estimate of the emitter position.

It can be concluded that the initial position found from RJMCMC was sufficiently close to the optimum and the chosen hyperparameters and chain length were appropriate, though the sampling of intensity still shows correlation and σ_I may need to be tuned more.

Low SBR results

For $I_{sim} = 500$, the prior on intensity is given as follows:

$$P(I) = \begin{cases} 0, & \text{if } I > 950 \\ \frac{1}{150\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left(\frac{I-500}{150}\right)^2 \right\}, & \text{if } I < 950 \end{cases} \tag{86}$$

The prior is plotted in 4. This dataset samples intensities from $\mathcal{N}(500, 150)$ photons, with a background of 20 photons.

model convergence In Supplementary Figure 5 a) and c), the model time series and autocorrelation are plotted. Again, from the incorrect initalization the estimated model rapidly reduces to a two or one emitter model. There is less mixing and faster convergence compared to the simulations with $I_{sim} = 2000$. Looking at the average autocorrelation, it reaches zero with a lag of just 250. Again, this indicates that the $30,000$ RJMCMC iterations are more than sufficient for determining the model. The histogram in Supplementary Figure 5 d) again shows that the chain spends virtually all its time in a single emitter model.

parameter convergence In Supplementary Figure 5 b), the average parameter autocorrelation for the 50 frames is shown. The autocorrelations of the position chains take longer to reach zero compared to the high SBR case. This time, the axial chain is actually slower than the intensity chain, and now even the lateral chain displays some slow transient behavior. The background converges first, followed by intensity and lateral position, and finally axial position. Under these conditions, convergence is limited by the sampling of the z position instead of the intensity. The acceptance rates for the moves are $\begin{bmatrix} 0 & 0.73 \pm 0.04 & 0.270 \pm 0.009 \end{bmatrix}$. The single emitter move has a high acceptance rate and the corresponding random walk hyperparameters may be increased for better mixing and less correlation between samples.

Calculating the lateral and axial RMSE results in 2 pixels and 360 nm errors, respectively, which is a lot less accurate especially in lateral direction compared to the high SBR case. This is due to some of the outliers in the sampled intensity, for instance frames 18 and 45 having an intensity less than 100 photons (79 and 83, respectively). Not considering these two frames, intensities still range from 200 to 700 photons (covering about 85% of the interval), and the lateral and axial RMSE become 0.14 pixels and 340 nm each.

This test has shown that though the chain length is sufficient, tuning of the single emitter move may lead to faster convergence.

Finally, it can be concluded that the hyperparameters provided here generally result in convergence, though it is best practice to adjust them to the imaging conditions of your system, ideally making use of autocorrelation plots to identify sub-optimal behaviour.

Single emitter localization convergence for a tetrapod PSF

The same test is run for a tetrapod PSF, under similar conditions, simulating 50 frames of in-focus emitters in the center of the ROI with some subpixel shift, for varying SBR. The size of the ROI was set to 28 by 28 pixels, to account for the wider PSF. The same move probabilities were used, while adjusting the hyperparameters and PSF width, which was estimated from a Gaussian fit of the width in focal plane. The hyperparameters can be found in Supplementary Table 3.

High SBR results

For $I_{sim} = 2000$, using the intensity prior shown in Supplementary Figure 2, the results are as follows.

model convergence Studying the autocorrelation in Supplementary Figure 6 a), it appears that model convergence is slightly slower with the tetrapod PSF. This may be related to use of a fixed clustering radius, after all the tetrapod PSF greatly changes in width over depth. There is a lot more mixing in the chain of Supplementary Figure 6 c) this time, with two or three emitter models existing up to around 3000 iterations into the run. Beyond 5000 iterations, the RJMCMC run seems to have converged strongly. Evaluating the average model chain autocorrelation suggests that a minimum of 2600 RJMCMC iterations is needed. Some multiple of 2600, say 7800 RJMCMC iterations, should be sufficient for model convergence on this problem. From the histogram in Supplementary Figure 6 d) it is clear that on average, the estimates spend nearly all their time in a model with one active emitter.

parameter convergence Supplementary Figure 6 b) plots the average parameter autocorrelation for the 50 frames. The lateral, axial, and background autocorrelation converge rapidly. The intensity chain takes the longest to reach zero after a lag of 500 and still shows some more slow transient behavior. The acceptance rates for the moves are $\begin{bmatrix}0 & 0.18 \pm 0.02 & 0.205 \pm 0.009\end{bmatrix}$ for the group move, single emitter move, and background move, respectively. The single emitter move has a low acceptance rate, possibly due to poor intensity sampling. The background move performs similarly to the runs on astigmatic data.

Calculating the lateral and axial RMSE results in 0.09 pixels and 20 nm errors.

It can be concluded that the chosen hyperparameters and chain length can converge, though the sampling of intensity still shows correlation and σ_I may need to be tuned more.

Low SBR results

For $I_{sim} = 500$, the prior in Supplementary Figure 4 was used. The results are as follows.

model convergence In Supplementary Figure 7 a) and c), the model time series and autocorrelation are plotted. Again, from the incorrect initalization the estimated model rapidly reduces to a two or one emitter model. More mixing seems to be happening in the first $10,000$ iterations compared to the simulations with $I_{sim} = 2000$. The lower SBR gives more uncertainty when determining the model. The average model autocorrelation reaches zero at a lag of about 200, as expected from the increase in mixing. This again indicates that the $30,000$ iterations used were more than sufficient. The histogram in 7 d) shows that the chains generally stay in a single emitter model.

parameter convergence In Supplementary Figure 7 b), the average parameter autocorrelation for the 50 frames is shown. This time, the lateral and axial position chains are slowest to converge, around a lag of 1500, while the background and intensity chains still reach zero in about 400 steps. Under these conditions, convergence is limited by the emitter position sampling instead of the intensity. The acceptance rates for the moves are $\begin{bmatrix}0 & 0.55 \pm 0.08 & 0.198 \pm 0.009\end{bmatrix}$. Both the single emitter and background move show appropriate acceptance rates, the single emitter move even on the higher side. The high acceptance rates in combination with the strong correlation in parameter jumps suggests σ_x, σ_y, and σ_z may be increased to reduce sample correlation and get faster convergence.

The lateral and axial RMSE are in 1.6 pixels and 200 nm errors, respectively. Again, removing some outliers with intensities less than 200 returns precisions of 0.3 pixels and 70 nm, which is appropriate for these conditions.

Again, these tests have shown that extensive tuning of the single emitter move may lead to faster convergence in both scenarios.

Finally, the hyperparameters used for imaging with the tetrapod PSF will converge to an accurate estimate, though not as rapidly as astigmatic imaging with its respective hyperparameters.

Single emitter precision and accuracy for an astigmatic PSF

To further verify the accuracy and precision of the algorithm, another set of synthetic data was generated using the same PSFs from the convergence test. Again, single emitters with random subpixel shifts were simulated with a background of 20 photons and intensities ranging from 600 to 3000 photons, in steps of 100. 50 emitters were simulated at each intensity. Move probabilities were the same as in the convergence tests, while the hyperparameters can be found in Supplementary Table 4.

The intensity prior is now set to be uniform, fixing the range of the intensity prior from 100 to 4000 photons. This makes all the priors uniform, effectively only serving as loose constraints. It is now possible to compare our results to the Cramer-Rao lower bound (CRLB), as the algorithm is virtually unbiased. Supplementary Figure 8 plots the CRLB against the localization precision found directly on individual frames, calculated from the MCMC chain under a Gaussian assumption. This indicates that for the astigmatic PSF, the histograms from individual frames give an accurate representation of the achievable precision. Supplementary Figure 9 plots the resampled root mean squared error (RMSE) for the same results. The RMSE roughly matches the CRLB from Supplementary Figure 8, showing that the algorithm can find a precise and accurate estimate under these conditions.

Single emitter precision and accuracy for a tetrapod PSF

The same testing was done for a tetrapod PSF, using the same hyperparameters as in 4 but with a PSF width of 2.1 pixels. At emitter intensities less than 1000 photons, some frames failed to converge, showcasing the difficulty of using a complex PSF at a low SBR. This may be due to the flat intensity prior yielding poor initializations and difficulty escaping local minima. The accuracy and precision are thus plotted over a range of 1000 to 3000 photons. Supplementary Figure 10 shows the CRLB and localization precision found as above. Note that the found precision for intensity breaks the CRLB, especially at higher intensity, despite the other parameter precisions tracking the CRLB well. The precision found when assuming a Gaussian distribution is thus not representative of the information present in the frame, at least not for the emitter intensity. Supplementary Figure 11 again shows the resampled RMSE for the same data. This time, all estimates behave as expected, broadly tracking the CRLB.

Multimodality in two emitter astigmatic imaging

During testing of two emitter separability, under some conditions four peaks were found in the reconstruction of the emitter position. This occurred despite the chains in those situations converging to the correct MAP number of emitters. This phenomenon was observed at smaller distances between emitters, using the astigmatic PSF under relative angles of 0 and 90 degrees with respect to the x axis, and for the tetrapod PSF under angles of 45 degrees. To understand if this was due to pseudoconvergence or if this was representative of the underlying posterior, a multimodal frame was analyzed, shown in Supplementary Figure 12. K-means clustering was used to find four potential localizations, shown together with the ground truth in Supplementary Figure 12 a). Then, by picking fixing one of these emitters and moving the other around the ROI at the same depth and intensity, the log likelihood was mapped, shown in Supplementary Figure 12 c). Fixing the rightmost emitter finds the global optimum to be a pair of emitters near the ground truth, while fixing the bottom emitter finds an alternate mode of the solution, a pair perpendicular to the ground truth. Note that this local optimum has virtually the same likelihood. To investigate if either mode could be representative of the data, a χ-squared test was done on a 95% confidence interval. The χ^2 value is calculated as follows:

$$\chi^2 = \sum_{i=1}^{N_p} \frac{(D_i - \mu_i)^2}{\mu_i} \tag{87}$$

With N_p the number of pixels, D_i the measurement of the i^{th} pixel, and μ_i the expected value. As the frames are subject to Poisson noise, the χ^2 threshold for a 95% confidence interval can be found with:

$$\chi^2_{threshold} = N_p + 1.96\sqrt{2N_p + \sum_{i=1}^{N_p} \frac{1}{\mu_i}} \tag{88}$$

Applying it to this frame results in $\chi^2_{threshold} = 456$, with the true and alternate modes returning a χ^2 of 407 and 408, respectively. Both modes thus are well within the 95% confidence interval and are representative of the frame.

To test whether it is possible to select the true mode from these hypotheses, the log likelihood ratios of the hypotheses are used to calculate the probability of error. This log likelihood ratio can be approximated by a Gaussian distribution:

$$LLR_{q,r} = \ln\left(\frac{P(D|H_q)}{P(D|H_r)}\right) \tag{89}$$

$$\mu_{q,r,s} = \sum_i^{N_p} \mu_i^s \ln\left(\frac{\mu_i^q}{\mu_i^r}\right) - \mu_i^q + \mu_i^r \tag{90}$$

$$\sigma_{q,r,s} = \sum_i^{N_p} \mu_i^s \ln\left(\frac{\mu_i^q}{\mu_i^r}\right)^2 \tag{91}$$

with $LLR_{q,r}$ the log likelihood ratio of hypothesis q over hypothesis r, $\mu_{q,r,s}$ the mean of $LLR_{q,r}$ assuming hypothesis s is true, and $\sigma_{q,r,s}$ the corresponding width. For the two hypotheses H_0 and H_1, $P(LLR_{0,1}|H_0)$ and $P(LLR_{0,1}|H_1)$ can now be calculated, as shown in Supplementary Figure 12 b). To retrieve the probability of error, we now use the probability of making a correct selection:

$$P_c = P(LLR_{0,1} > 0|H_0) - P(LLR_{0,1} < 0|H_1) \tag{92}$$

$$P_e = (1 - P_c)/2 \tag{93}$$

with P_c the probability of being correct and P_e the probability of error. Applying this to the two modes from Supplementary Figure 12 results in a probability of error of 0.498. It is therefore not possible to select the right mode solely given the data.

Testing the multimodal reconstructions found under different angles for the astigmatic and tetrapod PSF yielded similar results, where both pairs of emitter locations pass the chi-squared test and have greater than 49.5% probability of error.

In conclusion, while both modes are representative of the data, there is no distinction to be made based on their likelihood. This indicates that pseudoconvergence is not the case and the posterior distribution in this scenario is multimodal. As both PSFs show this multimodality under different conditions, this is likely related to PSF degeneracy.

Chapter 5

Conclusion

This chapter summarizes the thesis and draws the final conclusion. The first section provides a brief summary of the field and the work done. The second section discusses the capabilities and limitations of reversible jump Markov chain Monte Carlo (RJMCMC) localization in 3D. The final section presents possible future uses for the algorithm, collecting a series of super-resolution techniques that could benefit greatly from being decoded with RJMCMC.

5-1 Summary

The ever increasing demand for resolution in biomedical imaging has led researchers to develop novel microscopy techniques. Researchers were previously limited by the diffraction limit, a physical limit imposed by the wavelength of the captured light and the numerical aperture of the lenses used. The conception [12] and development [9, 13] of Single molecule localization microscopy (SMLM) has allowed microscopes to break the diffraction limit by a factor of 20, allowing researchers to study biological processes that were previously impossible to visualize.

SMLM works in conjunction with a fluorescence microscopy setup. Illuminating a tissue sample labeled with fluorescent molecules, the blinking of these so-called emitters allows SMLM techniques to separate an object into point sources. Collecting a movie of these emitters, their true position is estimated using the impulse response function, or point spread function (PSF) for optical systems. When enough frames are collected, an estimate for every point of the object is found and the object is reconstructed by collecting these estimates. As the cameras used in the microscopes add Poisson noise to the incoming image, a common estimator used is the Maximum Likelihood Estimation (MLE), allowing algorithms to find the most likely position while reaching the theoretical minimum uncertainty [14].

Though SMLM techniques perform excellently when emitters are well separated, they start to fail when emitter images begin to overlap. As these algorithms assume only a single emitter is present in the region of interest (ROI), the estimates become biased with overlap. Densely populated ROIs also make it hard for the algorithm to correctly identify the number of active emitters present. Unfortunately, it is not always possible to achieve sparse activation

of emitters. When imaging samples with high labeling density or thick samples used in 3D imaging, overlap of emitters becomes inevitable. Imaging dense frames is also desirable as it improves information density per frame, allowing for shorter acquisition times.

Multiple emitter fitting techniques were developed to circumvents this problem. They work by fitting to multiple emitter models instead of limiting the ROI to one emitter. Various multiple emitter fitting techniques were developed for 2D imaging, one of the most prominent being Bayesian multiple-emitter fitting (BAMF) [1]. By using RJMCMC [28], BAMF reconstructs posterior densities for both the number of emitters present as well as their positions and intensities. This allows BAMF to estimate the model and parameters simultaneously, which has been one of the challenges of imaging densely labeled samples. As a Bayesian method, it also allows the user to incorporate prior information into the estimate, resulting in improved model accuracy. Finally, BAMF reconstructs the posterior density, as opposed to the deterministic estimators common in localization microscopy. This allows it to better represent the uncertainty of parameter estimates.

We constructed a Bayesian 3D multiple emitter fitting algorithm using RJMCMC to better separate emitters in dense 3D imaging and to more accurately represent the uncertainty when doing so. The algorithm follows the structure demonstrated in BAMF, using an RJMCMC run to find the maximum a posteriori (MAP) model followed by a Markov chain Monte Carlo (MCMC) run to condition the parameter estimates on the MAP model. New RJMCMC moves were designed to sample the 3D position and intensity of emitters. Convergence of the algorithm was verified using frames with a single emitter, analyzing the lag τ autocorrelation of parameters and comparing the precision to the Cramér-Rao lower bound (CRLB). The algorithm was then tested for its ability to separate emitters by placing two emitters in a ROI and gradually moving them closer. Performing this two emitter separability test with an astigmatic PSF resulted in an experimental model accuracy of no less than 98 % when using the priors from BAMF. However, the reconstructed posterior density of emitter position was multimodal, finding four possible positions while correctly estimating a two emitter model. Further investigation led to two hypotheses for emitter position and evaluating the probability of error proved these hypotheses were indistinguishable in terms of likelihood. This demonstrates why posterior density reconstruction is valuable in dense 3D imaging, as it reconstructs non-Gaussian distributions. We believe this multimodality is a consequence of PSF degeneracy [2], where the complexity of the 3D PSF allows an image of emitters at a given depth to be represented by a different emitter count at a different depth. By varying the priors and Signal-to-background ratio (SBR), we demonstrate that this PSF degeneracy problem is relevant for a range of imaging conditions. Doing the same two emitter test with biplane imaging showed no multimodality, separating emitters up to a distance of 0.75 σ_{PSF}. This reaffirms that multimodality is tied to the 3D PSF used, while also demonstrating that the algorithm can be used to analyze 3D PSF degeneracy.

5-2 Discussion

While the algorithm functions well for narrow priors on emitter intensity, it starts to drop model accuracy when priors become wider than 500 photons. Though we did our two emitter tests with the intensity prior used in BAMF, which had a Gaussian peak with a width of approximately 150 photons, in practice the distribution of emitter intensities may be wider

than 500 photons. However, the algorithm is still capable of identifying PSF degeneracy regardless of intensity prior width, which is our most valuable result.

It should also be noted that the two emitter tests were not exhaustive, as not every possible angle or 3D orientation was tested. It is therefore possible that biplane imaging, which appeared to be free of PSF degeneracy, may still prove to be degenerate when emitters take on different positions or when the ROI is packed more densely. Any researcher that may want to use this algorithm to analyze the PSF of their setup needs to carefully consider which tests are important for the tissue samples they want to image later on.

Though the algorithm demonstrated convergence on frames with a single and multiple emitters, it is not recommended to use the algorithm for localization. This is due to the sequential nature of MCMC, resulting in a slow and impractical algorithm. Though the constructed algorithm analyzes frames in parallel, reconstructing long movies of large frames would take a significant amount of time. The system memory also becomes a constraint, as every iteration of the algorithm has to be stored. This again limits the number of frames and their size. Improving the code with smarter memory usage and periodically writing to storage is necessary to reconstruct large amounts of data.

5-3 Outlook

As demonstrated in chapter 4, the algorithm is best used to analyze and compare 3D PSFs. Some possible use cases will be presented here.

For astigmatic imaging, the emitter depth is solely encoded in the PSF shape. Combined with the dilation of the astigmatic PSF when moving out of focal plane, this resulted in multimodal posterior densities. If an imaging technique encodes the emitter depth in not just the PSF shape, but also some other factor, it may be possible to constrain the PSF degeneracy and image without multimodality. One such technique could be total internal reflection fluorescence (TIRF) [30, 31], where total internal reflection is used to illuminate the tissue sample with evanescent waves. This creates a patterned illumination in the sample that decays exponentially over depth, encoding emitter depth in emitter intensity. If the algorithm is then combined with a depth-dependent emitter intensity prior, it should be possible to separate emitters while limiting multimodality.

Following this train of thought, modulation-enhanced localization microscopy (meLM) [3] in any form should help suppress PSF degeneracy. The sinusoidal patterns in SIMFLUX [32] can be combined with an intensity prior dependent on lateral position to decode directly with RJMCMC. Similarly, Rose-z [33] generates sinusoidal patterns in the axial direction and may be decoded with a depth dependent intensity prior. Though the increased complexity of the techniques may require adjustments in the RJMCMC moves, we believe that combining these techniques with a 3D RJMCMC localization algorithm to analyze the resulting posterior distributions will be a topic of great interest.

Appendix A

Chi square test

When the null hypothesis is correct, the χ^2 distribution follows from:

$$\chi^2 = \sum_{i=1}^{N_p} \frac{(D_i - \mu_i)^2}{\mu_i} = \sum_{i=1}^{N_p} \frac{(D_i - E[D_i])^2}{E[D_i]} \tag{A-1}$$

with N_p the number of pixels, D_i the measurement and μ_i the expected value for the i^{th} pixel. For a measurement with Poisson noise, we know that $E[(D_i - E[D_i])^2] = E[D_i]$, so the expected value of χ^2 then becomes:

$$E[\chi^2] = \sum_{i=1}^{N_p} \frac{E[(D_i]}{E[(D_i]} = N_p \tag{A-2}$$

and the variance then follows from:

$$var[\chi^2] = E\left[\left(\sum_{i=1}^{N_p}\frac{(D_i - E[D_i])^2}{E[D_i]} - E\left[\sum_{i=1}^{N_p}\frac{(D_i - E[D_i])^2}{E[D_i]}\right]\right)^2\right] =$$
$$E\left[\left(\sum_{i=1}^{N_p}\frac{(D_i - E[D_i])^2}{E[D_i]} - \sum_{i=1}^{N_p}E\left[\frac{(D_i - E[D_i])^2}{E[D_i]}\right]\right)^2\right] =$$
$$E\left[\left(\sum_{i=1}^{N_p}\frac{(D_i - E[D_i])^2}{E[D_i]} - \sum_{i=1}^{N_p}\frac{E\left[(D_i - E[D_i])^2\right]}{E[D_i]}\right)^2\right] =$$
$$E\left[\sum_{i=1}^{N_p}\left(\frac{(D_i - E[D_i])^2}{E[D_i]} - 1\right)^2\right] =$$
$$E\left[\sum_{i=1}^{N_p}\left(\frac{(D_i - E[D_i])^2}{E[D_i]} - 1\right)^2\right] =$$
$$\sum_{i=1}^{N_p} 1 - 2\frac{E\left[(D_i - E[D_i])\right]}{E[D_i]} + \frac{E\left[(D_i - E[D_i])^4\right]}{E[D_i]^2} =$$

$$\sum_{i=1}^{N_p} 2 + \frac{1}{E\left[D_i\right]} \tag{A-3}$$

The expected value and variance can then be used to calculate the critical value of the 95% confidence interval:

$$\chi^2_{threshold} = E[\chi^2] + 1.96\sqrt{var[\chi^2]} = N_p + 1.96\sqrt{2N_p + \sum_{i=1} N_p \frac{1}{\mu_i}} \tag{A-4}$$

Appendix B

Probability of error

For comparing the hypotheses found in the multimodal reconstructions, we calculate the probability of error as follows. If we want to select from M hypotheses H_i, we have the associated probability of error for picking either H_i when H_j is true, or vise versa:

$$P_e = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} P(H_i|H_j)P(H_j) \tag{B-1}$$

Assuming the priors are equal, the probability of error expands to:

$$P_e = \frac{1}{M} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} P(H_i|H_j)P(H_j) \tag{B-2}$$

To simplify the calculation of P_e, we subtract the probability of being correct from 1, dividing by the number of hypotheses:

$$P_e = (1 - P_c)/M \tag{B-3}$$

The log likelihood ratio can be used to calculate P_c. It can be approximated with a Gaussian distribution:

$$LLR_{i,j}(x) = \ln\left(\frac{P(x|H_i)}{P(x|H_j)}\right) \tag{B-4}$$

$$LLR_{i,j}(x) \sim \mathcal{N}(\mu_{i,j,l}, \sigma_{i,j,l}) \tag{B-5}$$

$$\mu_{i,j,l} = \sum_{k=1}^{N_p} \mu_k^l \log\left(\frac{\mu_k^i}{\mu_k^j}\right) - \mu_k^i + \mu_k^j \tag{B-6}$$

$$\sigma_{i,j,l} = \sum_{k}^{N_p} \mu_k^l \log\left(\frac{\mu_k^i}{\mu_k^j}\right)^2 \tag{B-7}$$

with $LLR_{i,j}(x)$ the log likelihood ratio of hypothesis i to j, $\mu_{i,j,l}$ and $\sigma_{i,j,l}$ the mean and standard deviation of $LLR_{i,j}(x)$ assuming hypothesis l is true, and μ_k^i the expected value of pixel k under hypothesis i. Now, we have $LLR_{i,j}(x)$ distributions where we select hypotheses i over j where $LLR_{i,j}(x) > 0$ and vice versa. For two possible hypotheses, P_c can now be calculated:

$$P_c = P(LLR_{0,1} > 0|H_0) - P(LLR_{0,1} < 0|H_1) \tag{B-8}$$

Finally, the probability of error when choosing between hypotheses H_0 and H_1 becomes:

$$P_e = \left[1 - \left\{\phi\left(\frac{\mu_{0,1,0}}{\sigma_{0,1,0}}\right) - \phi\left(-\frac{\mu_{0,1,1}}{\sigma_{0,1,1}}\right)\right\}\right]/2 \tag{B-9}$$

Bibliography

[1] M. Fazel, M. J. Wester, H. Mazloom-Farsibaf, M. B. Meddens, A. S. Eklund, T. Schlichthaerle, F. Schueder, R. Jungmann, and K. A. Lidke, "Bayesian multiple emitter fitting using reversible jump markov chain monte carlo," *Scientific reports*, vol. 9, no. 1, pp. 1–10, 2019.

[2] H. P. Babcock and X. Zhuang, "Analyzing single molecule localization microscopy data using cubic splines," *Scientific reports*, vol. 7, no. 1, pp. 1–9, 2017.

[3] L. Reymond, T. Huser, V. Ruprecht, and S. Wieser, "Modulation-enhanced localization microscopy," *Journal of Physics: Photonics*, vol. 2, no. 4, p. 041001, 2020.

[4] C. Smith, M. Huisman, M. Siemons, D. Grünwald, and S. Stallinga, "Simultaneous measurement of emission color and 3d position of single molecules," *Optics Express*, vol. 24, no. 5, pp. 4996–5013, 2016.

[5] J. Lakowicz, *Principles of Fluorescence Spectroscopy*. Springer US, 2007.

[6] L. Holtzer, T. Meckel, and T. Schmidt, "Nanometric three-dimensional tracking of individual quantum dots in cells," *Applied Physics Letters*, vol. 90, no. 5, p. 053902, 2007.

[7] Y. Shechtman, S. J. Sahl, A. S. Backer, and W. E. Moerner, "Optimal point spread function design for 3d imaging," *Physical review letters*, vol. 113, no. 13, p. 133902, 2014.

[8] Y. Shechtman, L. E. Weiss, A. S. Backer, S. J. Sahl, and W. Moerner, "Precise three-dimensional scan-free multiple-particle tracking over large axial ranges with tetrapod point spread functions," *Nano letters*, vol. 15, no. 6, pp. 4194–4199, 2015.

[9] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm)," *Nature methods*, vol. 3, no. 10, pp. 793–796, 2006.

[10] M. G. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy," *Journal of microscopy*, vol. 198, no. 2, pp. 82–87, 2000.

[11] C. B. Müller and J. Enderlein, "Image scanning microscopy," *Physical review letters*, vol. 104, no. 19, p. 198101, 2010.

[12] E. Betzig, "Proposed method for molecular optical imaging," *Optics letters*, vol. 20, no. 3, pp. 237–239, 1995.

[13] 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*, vol. 313, no. 5793, pp. 1642–1645, 2006.

[14] C. S. Smith, N. Joseph, B. Rieger, and K. A. Lidke, "Fast, single-molecule localization that achieves theoretically minimum uncertainty," *Nature methods*, vol. 7, no. 5, p. 373, 2010.

[15] J. Min, C. Vonesch, H. Kirshner, L. Carlini, N. Olivier, S. Holden, S. Manley, J. C. Ye, and M. Unser, "Falcon: fast and unbiased reconstruction of high-density super-resolution microscopy data," *Scientific reports*, vol. 4, no. 1, pp. 1–9, 2014.

[16] R. Sun, E. Archer, and L. Paninski, "Scalable variational inference for super resolution microscopy," *bioRxiv*, p. 081703, 2016.

[17] T. Dertinger, R. Colyer, G. Iyer, S. Weiss, and J. Enderlein, "Fast, background-free, 3d super-resolution optical fluctuation imaging (sofi)," *Proceedings of the National Academy of Sciences*, vol. 106, no. 52, pp. 22287–22292, 2009.

[18] N. Gustafsson, S. Culley, G. Ashdown, D. M. Owen, P. M. Pereira, and R. Henriques, "Fast live-cell conventional fluorophore nanoscopy with imagej through super-resolution radial fluctuations," *Nature communications*, vol. 7, no. 1, pp. 1–9, 2016.

[19] E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-storm: super-resolution single-molecule microscopy by deep learning," *Optica*, vol. 5, no. 4, pp. 458–464, 2018.

[20] E. Nehme, D. Freedman, R. Gordon, B. Ferdman, L. E. Weiss, O. Alalouf, T. Naor, R. Orange, T. Michaeli, and Y. Shechtman, "Deepstorm3d: dense 3d localization microscopy and psf design by deep learning," *Nature methods*, vol. 17, no. 7, pp. 734–740, 2020.

[21] A. Speiser, L.-R. Müller, P. Hoess, U. Matti, C. J. Obara, W. R. Legant, A. Kreshuk, J. H. Macke, J. Ries, and S. C. Turaga, "Deep learning enables fast and dense single-molecule localization with high accuracy," *Nature methods*, vol. 18, no. 9, pp. 1082–1090, 2021.

[22] F. Aguet, D. Van De Ville, and M. Unser, "A maximum-likelihood formalism for sub-resolution axial localization of fluorescent nanoparticles," *Optics Express*, vol. 13, no. 26, pp. 10503–10522, 2005.

[23] N. Han and Z. Song, "Bayesian multiple measurement vector problem with spatial structured sparsity patterns," *Digital Signal Processing*, vol. 75, pp. 184–201, 2018.

[24] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[25] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, *Handbook of markov chain monte carlo*. CRC press, 2011.

[26] W. K. Hastings, "Monte carlo sampling methods using markov chains and their applications," 1970.

[27] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The journal of chemical physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[28] P. J. Green, "Reversible jump markov chain monte carlo computation and bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.

[29] S. Richardson and P. J. Green, "On bayesian analysis of mixtures with an unknown number of components (with discussion)," *Journal of the Royal Statistical Society: series B (statistical methodology)*, vol. 59, no. 4, pp. 731–792, 1997.

[30] D. Axelrod, "Total internal reflection fluorescence microscopy," *Methods in cell biology*, vol. 30, pp. 245–270, 1989.

[31] N. Bourg, C. Mayet, G. Dupuis, T. Barroca, P. Bon, S. Lécart, E. Fort, and S. Lévêque-Fort, "Direct optical nanoscopy with axially localized detection," *Nature Photonics*, vol. 9, no. 9, pp. 587–593, 2015.

[32] J. Cnossen, T. Hinsdale, R. Thorsen, F. Schueder, R. Jungmann, C. S. Smith, B. Rieger, and S. Stallinga, "Localization microscopy at doubled precision with patterned illumination," *BioRxiv*, p. 554337, 2019.

[33] L. Gu, Y. Li, S. Zhang, M. Zhou, Y. Xue, W. Li, T. Xu, and W. Ji, "Molecular-scale axial localization by repetitive optical selective exposure," *Nature Methods*, vol. 18, no. 4, pp. 369–373, 2021.

Glossary

List of Acronyms

FITC	fluorescein isothiocyanate
SLM	spatial light modulator
TIRF	total internal reflection fluorescence
PSF	point spread function
OTF	optical transfer function
FWHM	full width half maximum
EMCCD	electron multiplying charge-coupled device
MLE	Maximum Likelihood Estimation
ROI	region of interest
CRLB	Cramér-Rao lower bound
MAP	Maximum a Posteriori
MCMC	Markov chain Monte Carlo
VI	Variational inference
RJMCMC	reversible jump Markov chain Monte Carlo
SMLM	Single molecule localization microscopy
PALM	Photo-activated localization microscopy
STORM	Stochastic optical reconstruction microscopy
CS-STORM	compressed sensing STORM
BAMF	Bayesian multiple-emitter fitting
MAPN	maximum a posteriori model of number of emitters
MAP	maximum a posteriori
FALCON	Fast Localization algorithm based on Continuous-spatial formulation
vEM	variational expectation-maximization

SOFI	super-resolution optical fluctuation imaging
SRRF	Super-resolution through radial fluctuations
SBR	Signal-to-background ratio
NA	Numerical aperture
Cy5	Cyanine-5
NA	Numerical Aperture
SLM	spatial light modulator
GMM	Gaussian mixture model
OMC	Ordinary Monte Carlo
CLT	central limit theorem
MH	Metropolis-Hastings
meLM	modulation-enhanced localization microscopy